Extremes and Graphical Models

Stefka Asenova

Thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Sciences

February, 2023

Institut de Statistique, Biostatistique et Sciences Actuarielles
(ISBA/LIDAM)
Université catholique de Louvain
Louvain-la-Neuve
Belgium

Thesis Committee:

Pr. Anouar **El Ghouch**	UCLouvain/ISBA/LIDAM, Belgium
Pr. Christian **Hafner** (President)	UCLouvain/ISBA/LIDAM, Belgium
Pr. Anna **Kiriliouk**	University of Namur, Belgium
Pr. Johan **Segers** (Advisor)	UCLouvain/ISBA/LIDAM, Belgium
Pr. Chen **Zhou**	Erasmus University Rotterdam, The Netherlands

Extremes and Graphical Models

by Stefka Asenova

ISBA/LIDAM
Université catholique de Louvain
Voie du Roman Pays, 20
1348 Louvain-la-Neuve
Belgium

Acknowledgments

The greatest thanks go to my supervisor Johan Segers. He has guided me in what regards more general directions of the research as well as what regards technical details of mathematical statements. I appreciate highly all our research meetings and discussions. They were always challenging, intense, it required all my effort and attention in order to follow what was being explained, always to the point and leaving me with lots of work until the next meeting.

During one of these meetings, the idea of identifiability of tail dependence parameters has arisen and has become a key research topic throughout the whole thesis. It is a central idea of the first paper, which has been published in the journal *Extremes* in 2021. I take the opportunity to thank our collaborator in this article, Gildas Mazo, who was at the origin of the application related to data on water levels of the Seine, France. The data fitted the purposes of the paper very well, illustrating the practical usefulness of our idea. I also thank David Lee for his availability and the great help and clarifications he gave me regarding their model.

I would like to thank also Anna Kiriliouk for introducing me to parallel computing in R by sending me literature to start with. This was extremely helpful in all my simulation works afterwards.

The second paper appeared from our willingness to generalize results known for tree graphical models to a general graphical model or at least to a graphical model with respect to a decomposable graph. I would like to thank Antonio Abruzzo for helping me to find the name of the special graphs we were considering.

The fourth and the fifth chapters of the thesis focus on linear structural equation models (SEMs). At the beginning the results we had were considered as subsections of the second paper until we realized that the material is fundamentally different, quite long and it deserves attention apart from the topic of the second paper. For one of the proofs I have turned for help to Ngoc Tran, working actively on max-linear models. She replied immediately and showed herself very collaborative and I wish to thank her now for that. I also would like to thank Eugen Pircalabelu for his involvement and our lengthy discussion on dependence and independence in SEMs. Eugen has another important contribution: he has found that complete directed acyclic graphs are known in literature as transitive tournaments, and this gave us an idea for the title of our third paper.

I owe thanks to the anonymous referees of the papers who have pointed me to problems that I haven't been aware of before. Their questions and comments allowed me to clarify the link of our research with existing literature eventually leading to new significant results.

I would like to thank the members of my doctoral committee, Anouar El Ghouch and Christian Hafner for our meetings held in order to update my PhD progress, for their feedback and suggestions.

I deeply thank Rainer von Sachs and Christian Hafner for their moral support especially in my last year of PhD.

After thanking all people who have contributed in one way or another to my research, I would like to remind the reader that my special thanks go to Johan Segers

for all his help throughout the research process.

I am also grateful for the funding of my first year of PhD, provided by ARC 12/17-045 “Stochastic Modeling of Dependence: Systems Under Stress” funded by UCLouvain.

Contents

List of acronyms

cdf	cumulative distribution function
CL	Composite Likelihood
CLE	Composite Likelihood Estimator
DAG	Directed Acyclic Graph
ECE	Extremal Coefficients Estimator
EV	Extreme Value
EVT	Extreme Value Theory
iid	independent and identically distributed
MM	Method of Moments
MME	Method of Moments Estimator
MPD	Multivariate Pareto Distribution
pdf	probability density function
SEM	Structural Equation Models
stdf	stable tail dependence function
tdc	tail dependence coefficient
ttt	tree of transitive tournaments

List of symbols

$\xrightarrow{d}$	convergence in distribution of random variables or vectors
$\xrightarrow{w}$	weak convergence of finite Borel measures
$\perp\!\!\!\perp$	independence
$\not\perp\!\!\!\perp$	dependence, i.e., lack of independence
$\lvert\cdot\rvert$	if used with a set it is the cardinality of this set
$\mathcal{G}$	a graph object
$\mathcal{T}$	a tree or a tree of transitive tournaments
δ	Dirac measure
$\mathrm{pa}(\cdot)$	set of the parents of some node set
$\mathrm{Pa}(\cdot)$	set of the parents of some node set including the set itself
$\mathrm{an}(\cdot)$	set of the ancestors of some node set
$\mathrm{An}(\cdot)$	set of the ancestors of some node set including the set itself
$\mathrm{ch}(\cdot)$	set of the children of some node set
$\mathrm{Ch}(\cdot)$	set of the children of some node set including the set itself
$\mathrm{desc}(\cdot)$	set of the descendants of some node set
$\mathrm{Desc}(\cdot)$	set of the descendants of some node set including the set itself
$\mathrm{cd}(\cdot)$	clique degree of a node
$\mathcal{N}$	the Gaussian distribution
$\mathcal{L}$	the law of a random variable or vector
Φ	the standard normal cdf
ϕ	the standard normal pdf

Introduction 1

This thesis combines concepts from the fields of extreme value theory (EVT) and probabilistic graphical models. The broad scope is to study the tail behavior of a random vector when the joint distribution represents a graphical model. Tail behavior is formalized through perspectives typical in EVT, such as convergence of the componentwise maxima, or of the peaks over a threshold, or through the one-component regular variation property of the joint distribution. A graphical model is a random vector, or its distribution, possessing a set of conditional independence relations. It is called *graphical* because a graph (a structure with set of nodes and set of edges) can be drawn to represent the conditional independence (or dependence) relations of the joint distribution.

The thesis focuses on particular graphical models: those that can be represented by a tree or a generalization of it. We do cover however both undirected and directed graphs, although limiting ourselves to directed acyclic graphs (DAGs). Also, the thesis studies two parametric classes: models parameterized by Hüsler–Reiss distributions and linear structural equation models (SEMs). Hence the scope of the thesis in studying extremes on graphical models is limited to:

- graphical models with respect to trees or generalizations of a tree;
- graphical models parameterized by Hüsler–Reiss distributions or linear SEMs.

There are two topics that dominate the thesis. The first one is related to the extreme value limits of a graphical model, such as the domain of attraction, a Pareto type limit and the behavior of a suitably scaled random vector, given that a high threshold is exceeded by one of the elements. The latter type of limit has been studied as earliest to my knowledge in Smith (1992) for Markov chains, followed by numerous papers enriching the literature on the topic and providing large flexibility in models and assumptions.

The second dominant topic of the thesis is rather new, in the sense that the problem has not yet been discussed in the literature at all. This is what we called the *parameter identifiability problem*. We study it in the context of the two particular parameterizations mentioned above. In short, the problem consists of identifying all parameters defining the joint distribution of the whole random vector from the joint distribution of a particular subset of random variables from this vector. The relevance of this problem is practical and this becomes clear in the first paper dedicated on the Seine river in France.

To sum up, the whole thesis gravitates around two main problems:

- extreme value limits of a graphical model, such as the domain of attraction, a Pareto type limit and the behavior of a suitably scaled random vector, given that a high threshold is exceeded by one of the elements,
- identifiability of all parameters defining the complete joint distribution from a particular marginal distribution.

The first problem has been studied in different contexts in the extreme value literature and our contribution is that we present other models and parameterizations whose limits seem logical on the background of this literature. The second problem has not been studied so far, so we hope that with the papers published and under review we have clearly defined the problem, together with its theoretical and practical features.

To better clarify the two problems in the focus of the thesis we need some introduction to extreme value theory and graphical models, limited to main notions and concepts. This is done in the first three sections of the introduction. Next, a section is dedicated to each of the chapters, summarizing the main results and contributions to the literature. This is done in sufficient detail and generality at the same time so that reading the preamble is supposed to suffice for evaluating the contribution and relevance of the thesis.

In the preamble, if not otherwise explained, one can assume the following notation. If at this point there are notions that are unfamiliar to the reader (s)he will find definitely somewhere in the preamble a definition. In this introduction $p(i, j)$ is a unique shortest *undirected* path between nodes i and j, which contains a finite number of *edges*, and $\overrightarrow{p}(i, j)$ is a unique shortest *directed* path between nodes i and j, which contains a finite number of *edges*.

1.1 Extreme value theory

The greatest flood in Paris, due to the river Seine, dates back to 1910 and it has been estimated (OECD, 2018) that nowadays a flood of such rank would affect over five million citizens and cause a damage between 3 and 30 billion euros. The water level of the Seine in Paris during this flood rose to 8.6 meters on Paris-Austerlitz scale according to Éditions OCDE (2014). How to estimate the return period of such an event on the basis of hundred years of available data in which there is only one flood of this magnitude? Even worse - consider the example in de Haan and Ferreira (2007) according to which the Dutch dikes should be so high that the probability of a flood in a given year is 10^{-4}. How to compute the height of the dikes if there are hundred years of data and hence not a single observation to give us an idea of the 10000-years event? EVT provides a solid basis for extrapolation.

Applications of EVT are not limited to floods and hydrology. Other natural hazards often modeled using EVT are earthquake magnitude (Pisarenko et al., 2014; Beirlant et al., 2019) and wildfire severity (French et al., 2019). Financial and insurance applications are countless (Resnick, 1997; Embrechts et al., 2013; Einmahl et al., 2018). Other studies can be found in the field of traffic safety Orsini et al. (2019), drugs side effects (Southworth, 2014), biology (Basnayake et al., 2019) and sport (Einmahl and Smeets, 2011).

When considering several risks at the same time we would often observe that values become more and more positively correlated at higher levels. This is what we see in the data on water levels along the Seine, France. Figure 1.1 shows observations on water levels of the Seine in Paris and Sens. More precisely, these are marginal uniform transformations using the empirical cumulative distribution functions (cdfs). Higher water levels in Paris tend to occur simultaneously with high water levels in Sens. Meanwhile, it looks that intermediate levels are rather uncorrelated. This phenomenon is known in EVT as extremal dependence. Many financial series exhibit extremal dependence too (Einmahl et al., 2018).

In order to focus on the extremal (or tail) dependence it is customary to transform the univariate margins. For a random variable X with cdf $F(x)$, a common choice is to take $1/(1 - F(X))$. We require $F(x)$ to be a continuous function. The transformed variable $1/(1 - F(X))$ has a unit Pareto distribution with cdf $1 - 1/x, x \geq 1$.

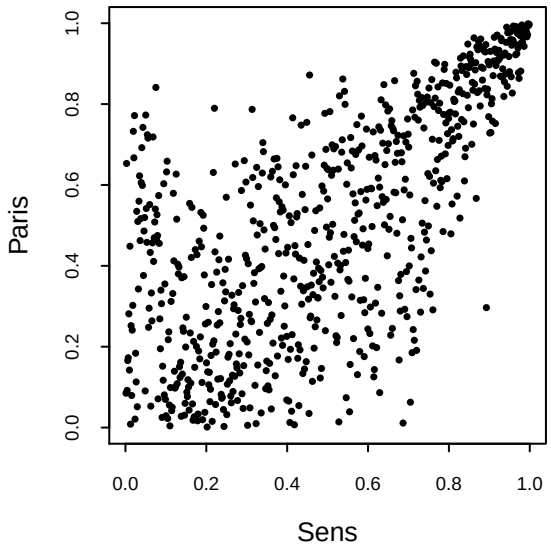


Figure 1.1: Scatterplot of water levels of the Seine in Paris and Sens, data transformed to a uniform scale.

Let $X = (X_v, v \in V)$ be a random vector for some index set V, with joint cdf $F(x), x \in [1, \infty)^V$ and univariate unit Pareto margins. The following definition can be consulted in one of the monographs on EVT such as de Haan and Ferreira (2007, Chapter 6), Resnick (1987, Chapter 5) or Beirlant et al. (2004, Chapter 7).

Definition 1 (**Domain of attraction**) If there is a cdf $G(x), x \in [0, \infty)^V \setminus \{0\}^V$ such that, for $n = 1, 2, \ldots$

$$\lim_{n\to\infty} [\mathbb{P}(X_v \leq nx_v, v \in V)]^n = \lim_{n\to\infty} [F(nx)]^n = G(x) \tag{1.1}$$

for all continuity points of F, we say that F is in the domain of attraction of G, also noted by $F \in D(G)$ which is called *simple extreme value distribution.* ■

The class of extreme value distributions coincides with the class of max-stable distributions. Both names can be used interchangeably. There are different ways to express the form of the distribution G. We will introduce the representation in terms of the so called angular measure (de Haan and Ferreira, 2007, Theorem 6.1.14), (Resnick, 1987, Proposition 5.11).

Definition 2 (**Angular measure**) Let G be an extreme value distribution with unit Fréchet margins, ith cdf $x \mapsto \exp(-1/x), x > 0$. Let $\|\cdot\|$ be some norm and $\mathcal{A}$ the space given by

$$\mathcal{A} = \{a \in [0, \infty)^V \setminus \{0\}^V : \|a\| = 1\}.$$

There exists a finite measure $\mathcal{S}$ such that

$$G(x) = \exp\left(-\int_{\mathcal{A}} \max_{v\in V}(a_v/x_v)\mathcal{S}(\mathrm{d}a)\right)$$

with $\int_{\mathcal{A}} a_v \mathcal{S}(\mathrm{d}a) = 1$ for every $v \in V$. ■

We have made use of the angular measure in Chapters 4 and 5 where we work with max- or sum-linear models.

An equivalent way of assuming $F \in D(G)$ is to assume a Pareto type limiting distribution.

Definition 3 (**Multivariate Pareto Distribution (MPD)**) The MPD is the pointwise limit of

$$\mathbb{P}\left(X_v \le t x_v, v \in V \mid \max_{v \in V} X_v \ge t\right), \qquad t \to \infty. \tag{1.2}$$

If G is a simple extreme value distribution, then the MPD is given by

$$\mathcal{P}(x) = \frac{\ln G(\min(x_v, 1), v \in V) - \ln G(x)}{\ln G(1, \dots, 1)}, \qquad x \in [0, \infty)^V \setminus \{0\}^V. \tag{1.3}$$

■

The MPD is closely related to the generalized MPD introduced in Rootzén and Tajvidi (2006). The equivalence between the existence of the MPD as a limit of (1.2) and the domain of attraction condition is established in Rootzén and Tajvidi (2006, Theorem 2.2): i.e. a distribution G exists such that (1.1) holds if and only if the limit of (1.2) exists and is given by (1.3).

The MPD is at the core of the definition of an *extremal graphical model* introduced in Engelke and Hitz (2020), a paper which has stirred much appreciation in the community working on extremes and graphical models. This thesis does not make much use of extremal graphical models, but there is one important contribution that our second paper brings to the literature: in the context of Hüsler–Reiss parameterization we find one model which in the limit is an extremal graphical model, thereby complementing one of the theorems in Engelke and Hitz (2020).

The Hüsler–Reiss distribution is considered the EVT counterpart of the Gaussian distribution. It has a somehow complicated expression and it will be introduced in details both in Chapter 2 and in Chapter 3. If $X^{(1)}, \dots, X^{(n)}$ are iid copies of a $|V|$-variate normal distribution, consider the scaled sample maxima $M_v(n) = (\max_{i=1,\dots,n} X_v^{(i)} - a_v(n))/b_v(n)$ where $a_v(n)$ and $b_v(n) > 0$ are some functions for every $v \in V$. Further, suppose that the bivariate correlations ρ_{ij} satisfy

$$\lim_{n \to \infty} (1 - \rho_{ij}(n)) \ln(n) = \delta_{ij}^2, \qquad i, j \in V. \tag{1.4}$$

The Hüsler–Reiss distribution arises as a limit of $(M_v(n), v \in V)$ as $n \to \infty$ (Hüsler and Reiss, 1989). The limit is a max-stable (extreme value) distribution with parameter matrix $\delta_{ij}, i, j \in V$ which describes the tail dependence of X. The assumption on the correlation coefficients in (1.4) is key: for constant correlations, ρ_{ij}, the limit of the scaled sample maxima is a vector with independent Gumbel variables. Hence dependence of extreme values is achieved through (1.4).

The Hüsler–Reiss distribution has seen many applications. For instance Asadi et al. (2015), Lee and Joe (2017) and Engelke and Hitz (2020) use it to estimate tail dependence on river discharges or water flows. It appears in numerous studies on spatial statistics through the Brown–Resnick process whose finite-dimensional distribution is a Hüsler–Reiss distribution (Engelke et al., 2011, 2014; Einmahl et al., 2018). A simpler analytical form of the distribution is provided in Huser and Davison (2013) and Genton et al. (2011).

1.2 Graphical models

A graphical model is a pair of a random vector $X = (X_v, v \in V)$ (or its distribution, F) and a graph $\mathcal{G} = (V, E)$, where V is the set of nodes/vertices and $E \subseteq V \times V$ is a set of edges. An edge connects two nodes, thereby having the notation (a, b) for $a, b \in V$. If two nodes are connected by an edge we say that the two nodes are *adjacent*.

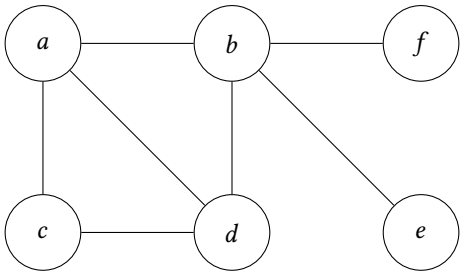


Figure 1.2: An example of a graph - node set $V = \{a, b, c, d, e, f\}$ and edge set $E = \{(a, b), (b, a), \ldots, (b, e), (e, b)\}$. Note that (a, b) corresponds to an edge directed from a to b. If the edge is undirected it means that both edges (a, b) and (b, a) are in $E \subseteq V \times V$.

The graph is an object from *graph theory* where typical problems are Euler paths, Hamiltonial paths, connectivity, coloration, etc. (Bondy et al., 1976). An arbitrary graph is illustrated in Figure 1.2.

This thesis makes use of trees, block graphs and DAGs whose skeleton (undirected version) is a block graph. A tree is characterized by unique paths between two nodes, no cycles and (minimal) separators between every two non-adjacent nodes being single nodes. While trees are popular graphs, block graphs are less so. In this thesis a block is a synonym to a complete graph, a graph where every two nodes are connected. A block graph is a graph which can be obtained from a tree, replacing every edge by a complete graph. Block graphs are generalizations of trees (Behtoei et al., 2010). The block graph is acyclic up to blocks, there exist a unique shortest path between every two nodes and the minimal separators are also singletons. To obtain a block graph in Figure 1.2 we have to add at least an edge between nodes b, c. If we direct the edges such that every block is a DAG, then we call this directed graph a *tree of transitive tournaments*, a ttt. The ttt is a DAG itself. So the three type of graphs that are used in this thesis are all variations of trees - trees themselves, block graphs and trees ot transitive tournaments.

In graphical models there is a random variable associated to every node and an edge corresponds to conditional dependence between the two variables. While graph theory dates back to the eighteenth century, graphical models are more recent. Earliest references are the work of the American physician Josiah Gibbs in 1902 (Lauritzen, 1996) and the path analysis of the biometrician Sewall Wright (Wright, 1934; Pearl, 2000). Graphical modeling has shaped with the papers Lauritzen and Wermuth (1989) and Edwards (1990). Many of the scholars commenting on these two papers (Hand et al., 1990) recognized the important contribution to the analysis of mixed (discrete and continuous) data, the significant progress in building a sound theoretical basis for modeling random variables related by a set of conditional independence relations, and the potential for applied research in areas such as biology, sociology, psychiatry where data are naturally mixed.

Graphical models can be with respect to a directed or an undirected graph. A directed edge from node a to node b means direct influence from X_a to X_b, while an undirected edge between $a, b \in V$ means that the influence can go in both directions. In what follows we assume that the graph is without loops, i.e., there are no edges that start and end at the same node.

In the context of undirected graphs there are two distinct Markov properties of conditional independence (Lauritzen, 1996).

Definition 4 (**Local Markov property**) We say that the joint distribution of X satisfies the *local* undirected Markov property with respect to a graph $\mathcal{G}$ if for every

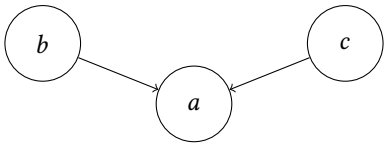


Figure 1.3: A v-structure: node with disconnected parents. It is a special feature of graphical models with respect to directed graphs. This is due to the fact that in general, that is, for almost all distributions over the three variables X_a, X_b, X_c, we have $X_b \not\perp\!\!\!\perp X_c \mid X_a$.

two non-adjacent nodes $v, u \in V$ we have

$$X_v \perp\!\!\!\perp X_u \mid X_{V\setminus\{v,u\}}. \tag{1.5}$$

■

Definition 5 (**Global Markov property**) We say that the joint distribution of X satisfies the *global* undirected Markov property with respect to a graph $\mathcal{G}$ if for every three disjoint non-empty subsets $A, B, S \subseteq V$ whenever A is separated from B by S (all paths from A to B pass through S) we have

$$X_A \perp\!\!\!\perp X_B \mid X_S. \tag{1.6}$$

■

When we talk about *Markov trees*, we assume that X is a graphical model with respect to a tree which satisfies the global Markov property. Markov trees have been used as building elements in vine copulas and pair-copula constructions (Bedford and Cooke, 2001, 2002; Aas et al., 2009). In this way multivariate dependence is modeled with bivariate copulas and often few trees suffice to approximate the true distribution (Lee and Joe, 2017). In extreme value context pair copula constructions are used in Lee and Joe (2017) and Segers (2020b) and also in Chapter 2.

If X is a graphical model with respect to a DAG (X is called also Bayesian network), then the distribution of X satisfies at least the *local* directed Markov property. In a direct graph we have $\mathrm{pa}(v) \subseteq V$, the sets of parents of node v; $\mathrm{desc}(v)$, the set of descendants of v and $\mathrm{an}(v)$, the set of ancestors of v. The capital letter in these three notation, such as $\mathrm{Pa}(v)$, will denote the set of elements in the corresponding set including node v, so $\mathrm{Pa}(v) = \mathrm{pa}(v) \cup \{v\}$.

Definition 6 (**Local directed Markov property**) We say that the joint distribution of X satisfies the local directed Markov property with respect to a DAG $\mathcal{G}$ if for every node $v \in V$

$$X_v \perp\!\!\!\perp X_{V\setminus\mathrm{Desc}(v)} \mid X_{\mathrm{pa}(v)}. \tag{1.7}$$

■

A peculiarity of graphical models with respect to directed graphs is a so called v-structure (or immorality). This represents a node with two disconnected parents, see Figure 1.3. The *global* property is meant to take into account the existence of v-structures. The special feature is due to the fact that, for *almost all* distributions over the three variables X_a, X_b, X_c, we have $X_b \not\perp\!\!\!\perp X_c \mid X_a$ (Koller and Friedman, 2009, Chapter 3). This can be compared to a graphical model with respect to the undirected version of the same graph. Given that a separates b and c, we have $X_b \perp\!\!\!\perp X_c \mid X_a$.

A SEM is defined as follows: for every node $v \in V$ we have

$$X_v = f(X_{\mathrm{pa}(v)}, \varepsilon_v),$$

where f is some function, ε_v is independent of $X_{\mathrm{pa}(v)}$ and $(\varepsilon_v, v \in V)$ are independent between each other. Gissibl and Klüppelberg (2018) introduce max-linear models

with respect to a DAG. Being in the context of extreme values, we choose the noise variables to be heavy tailed.

Definition 7 (**Max-linear model**) The random vector X follows a max-linear model on a DAG $\mathcal{G}$ when there exist $c_{ij} \geq 0$ for all $i, j \in V$, and $(Z_v, v \in V)$ independent unit Fréchet variables, such that for every $v \in V$ we have

$$X_v = \max\left(\max_{i \in \mathrm{pa}(v)} c_{iv} X_i, c_{vv} Z_v\right).$$

There exists a matrix $B = \{b_{ij}\}_{i,j \in V}$ such that we can write

$$X_v = \max_{i \in \mathrm{An}(v)} b_{vi} Z_i.$$

■

Max-linear models with respect to a DAG have been extensively studied in relation to the conditional independence relations that they possess (Klüppelberg and Lauritzen, 2019; Améndola et al., 2022; Améndola et al., 2021), because it turns out that these may differ significantly from classical independence results for Bayesian networks. The distribution of a max-linear model is an extreme value distribution with angular measure, $\mathcal{S}$, as in Definition 2 being a discrete measure.

Instead of taking the maximum, we could consider the sum, giving raise to the linear structural causal model appearing in Gnecco et al. (2021).

Definition 8 (**Sum-linear model**) The random vector $X = (X_v, v \in V)$ follows a sum-linear model on a DAG $\mathcal{G}$ when , there exist $c_{ij} \geq 0$ for all $i, j \in V, i \neq j$, and $d_{ii}, i \in V$, and $Z = (Z_v, v \in V)$ independent unit Fréchet variables, such that for every $v \in V$ we have

$$X_v = \sum_{i \in \mathrm{pa}(v)} c_{iv} X_i + d_{vv} Z_v.$$

If we gather the coefficients $c_{ij} \geq 0$ for all $i, j \in V, i \neq j$ in a $|V| \times |V|$ matrix C and $d_{ii}, i \in V$ in a diagonal matrix D, the above model can be written in matrix-vector form

$$X = CX + DZ.$$

There exists a matrix $B = \{b_{ij}\}_{i,j \in V}$ such that we can write

$$X = BZ.$$

■

The distribution of the sum-linear model with absolutely continuous independent factors is absolutely continuous, but it belongs to the domain of attraction of an extreme value distribution with discrete measure just as the max-linear model (Einmahl et al., 2012, Lemma 6.1).

The SEM with respect to a DAG is a special type of a more general linear SEM in which it is possible to have two variables influencing each other. For instance consider two variables (X_1, X_2) and the model

$$X_1 = \alpha + \beta X_2 + \varepsilon_1,$$
$$X_2 = \alpha' + \beta' X_1 + \varepsilon_2,$$

with $\varepsilon_1, \varepsilon_2$ being two independent error terms. This model is frequently used in economics where there is a feedback loop between (X_1, X_2). The corresponding graph is given in Figure 1.4. Each of the variables is once a response and once an explanatory variable. The construction of such a model can be based on an expert knowledge and in this case the causal interpretation stems from the expert belief of the relations between the variables.

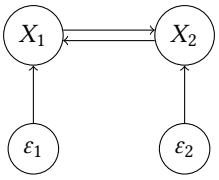


Figure 1.4: A directed graph corresponding to a model with two endogenous variables X_1, X_2 and two latent variables $\varepsilon_1, \varepsilon_2$. For a given equation, a directed edge is drawn from a variable on the right-hand side to the response variable.

Causality in SEMs is an assumption not a consequence, in the sense that if one believes in certain causal relations within a set of random variables, one way to estimate causal effects is by a set of structural equations. Another approach is based on data to discover causal relations by estimating quantities that help us identify the direction of the influence. This line of research is dominant in the literature on max-/sum-linear models, such as Gnecco et al. (2021), Tran et al. (2021b), Klüppelberg and Krali (2021), to name a few. SEMs represent one approach to causality. Others are interventional analysis, Granger causality for time series, and state space modeling (Shephard and Rambachan, 2020).

When analyzing multivariate data over time, structural equations and vector autoregression (VAR) models are often considered as alternatives. If X_t is a random vector with d elements, observed at time t, the VAR model is given by

$$X_t = BX_{t-1} + \varepsilon_t,$$

where B is a matrix of coefficients and ε_t is a vector of errors at time t. Now consider the SEM according to which every element $X_{v,t}$ for $v = 1, \ldots, d$ is given by

$$X_{v,t} = \alpha^T X_{-v,t} + \epsilon_{v,t},$$

where $X_{-v,t}$ is a vector of all elements of X_t excluding $X_{v,t}$ and $\epsilon_{v,t}$ is an independent error term. The VAR model suggests time-lagged influence between the variables, while the SEM implies contemporaneous interactions among them. Both models are largely used in neuroscience to study brain connectivity (Chen et al., 2011; Shen et al., 2019). To combine both lagged and contemporaneous effects researchers use SVAR (structural vector autoregression) models which interpret the spatio-temporal behavior observed in brain imaging data as a result of both instantaneous and time-lagged interactions between brain regions. In another study on the Italian economy, SEM and VAR are also viewed as complementary tools since the first one is shown to outperform the second in long-term forecasting and the other way around in short term forecasting (Boero, 1990). A comparison of SEM and VAR in microeconomics is given in Manera (2006).

1.3 Extremes and graphical models

Two directions in which the literature on graphical models and extremes develops fast and widely are max-linear Bayesian models and graph structure learning. Graph learning aims to estimate the edges among fixed number of nodes from data using extreme value analysis and metrics. Some references in this direction are Gissibl et al. (2018), Klüppelberg and Krali (2021), Buck and Klüppelberg (2021), Gissibl et al. (2021), Tran et al. (2021b), and Tran et al. (2021a). Graph discovery for other than max-linear models are studied in Gnecco et al. (2021), Engelke and Hitz (2020) and Engelke and Volgushev (2020).

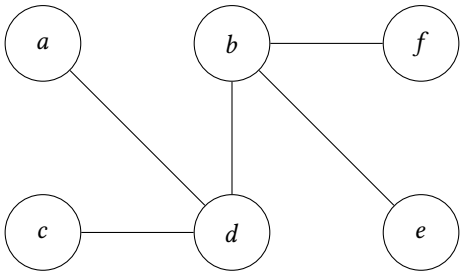


Figure 1.5: A tree on six nodes. The unique path $p(a, f)$ is the set of edges $\{(a, d), (d, b), (b, f)\}$.

A paper of significant influence to this thesis is Segers (2020b). It studies the limit of

$$(X_v/t, v \in V) \mid X_u = t, \qquad t \to \infty \tag{1.8}$$

if X satisfies the global Markov property as in Definition 5 with respect to a tree and if for every two adjacent nodes, $a, b \in V$, the corresponding bivariate distribution satisfies the condition

$$\lim_{t\to\infty} \mathcal{L}(X_b/t \mid X_a = t) = \nu_b, \tag{1.9}$$

for some probability measure ν_b. In particular, it is shown (Segers, 2020b, Theorem 1) that the limiting variables factorize into independent increments along the unique path from node u, the one at which the high threshold is exceeded, to any other node $v \in V \setminus u$.

As an example, consider the tree in Figure 1.5, and suppose the high threshold is exceeded at node a. There are five mutually independent variables $M_{ad}, M_{dc}, M_{db}, M_{be}, M_{bf}$ which determine the limiting distribution of (1.8). We have

$$(X_b/t, \ldots, X_f/t) \mid X_a = t \xrightarrow{d} (A_{ab}, \ldots, A_{af})$$

where $A_{ab} = M_{ad}M_{db}$, $A_{ac} = M_{ad}M_{dc}$, and so on for the remaining three variables.

This result has been inspired from convergence of Markov chains treated in Smith (1992), Yun (1998), Perfekt (1994), Segers (2007), and Janssen and Segers (2014). More recent studies of Markov chains allowing for complex behavior of the chain are presented in Papastathopoulos et al. (2017) and Papastathopoulos and Tawn (2019). The analytical form of the limit of (1.8) for any random vector, not necessarily possessing conditional independence relations, is given in Heffernan and Tawn (2004, Section 8).

The result of Segers (2020b, Theorem 1) has been generalized to Markov fields with respect to a block graph in Chapter 3. A posteriori, this is not unexpected as block graphs are generalizations of trees.

A key result in extremes and graphical models is presented in Papastathopoulos and Strokorb (2016) which says that an absolutely continuous extreme value (or max-stable) distribution cannot possess non-trivial conditional independence relations. By non-trivial we mean that the conditioning set cannot be an empty set. This means that if the distribution G from Definition 1 is absolutely continuous, then G cannot be a graphical model, unless the graph is either complete (all possible pairs are connected between each other) or the opposite of complete - disconnected graph (no edges).

Given that G cannot serve as a graphical model, Engelke and Hitz (2020) take the MPD $\mathcal{P}$ from Definition 3 to define an extremal graphical model as follows.

Definition 9 (Extremal graphical model, (Engelke and Hitz, 2020)) Let Y have absolutely continuous distribution $\mathcal{P}$. Define $Y^{(u)} = Y \mid Y_u > 1$. We say that

Y is an extremal graphical model with respect to $\mathcal{G} = (V, E)$ if for every $u \in V$, the random vector $Y^{(u)}$ satisfies the local Markov property from Definition 4. ■

An extremal graphical model is entirely defined in terms of the limiting distribution, $\mathcal{P}$, but no relation is made to distributions in the domain of attraction of $\mathcal{P}$ or of G. In Chapter 3, when we consider the Hüsler–Reiss distribution as a special case for our model, we take the usual perspective in EVT and provide a model which is indeed in the domain of attraction of a Hüsler–Reiss extremal graphical model. This fills the gap in Theorem 3 in Engelke and Hitz (2020) by providing the parameter matrix of the Hüsler–Reiss extremal graphical model.

Definition 10 (Hüsler–Reiss extremal graphical model, Engelke and Hitz (2020)) A Hüsler–Reiss extremal graphical model is obtained when G in the distribution $\mathcal{P}$ is a Hüsler–Reiss distribution. ■

Theorems 3 and 4 of Engelke and Hitz (2020) are dedicated to Hüsler–Reiss extremal graphical models with respect to a block graph. This is the kind of graphs that we use in Chapter 3. In Engelke and Hitz (2020) it is shown that if for every block, say C out of the set of all blocks $\mathcal{C}$, the distribution of X_C is parameterized by a Hüsler–Reiss Pareto distribution $\mathcal{P}$, then the distribution of the complete vector $X = (X_C, C \in \mathcal{C})$ is a Hüsler–Reiss extremal graphical model. Missing however in their results is the expression of the parameter matrix associated to this Hüsler–Reiss extremal graphical model, i.e., the form of $\delta_{ij}, i, j \in V$ from (1.4). In Chapter 3 we arrive at this result somehow accidentally. In our setting X_C is such that the limit in (1.8) exists and it is the one that would arise if X_C had Hüsler–Reiss distribution. Actually this limit is shown to be the log-normal distribution (Engelke et al., 2014; Segers, 2020b). Then we show that the Pareto distribution from Definition 2 of $X = (X_C, C \in \mathcal{C})$ is a Hüsler–Reiss extremal graphical model. Given that we provide the parameter matrix of the latter and according to Theorem 3 and 4 of Engelke and Hitz (2020) this matrix must be unique, we concluded that the parameter matrix found, must be the same as the one mentioned in Theorem 3 and 4 of Engelke and Hitz (2020). Recent advances in the field of Hüsler–Reiss extremal graphical models is Hentschel et al. (2022).

The studies on max-linear Bayesian networks are numerous. We have contribution to this literature too with the material from Chapter 4. We consider a DAG whose skeleton (the graph obtained by removing the directions of the edges) is a block graph. We called such a graph a *tree of transitive tournaments* (ttt). A complete graph which is a DAG is known in graph theory as transitive tournament. The ttt combines in a tree-like manner transitive tournaments and as a consequence it shares some properties with block graphs or trees. For instance if there is a directed path between two nodes, there is a unique shortest directed path. It is an acyclic graph too.

Max-linear models are quite special in what regards the conditional independence relations that they may possess. A whole new terminology and theory is created in Améndola et al. (2022) to establish conditional independence within max-linear models. The main reason for the complexity of the max-linear models in comparison with classical Bayesian networks is due to the *criticality* of paths or max-weighted paths. We refer to Gissibl and Klüppelberg (2018), Améndola et al. (2022), Gissibl et al. (2021) and Klüppelberg and Lauritzen (2019) for proper definition and examples.

The topics treated in Chapter 4 are new to the literature on max-linear models. Assume $(X_v, v \in V)$ is max-linear with respect to a ttt. We make use of Améndola et al. (2022) in order to establish a condition for X needed to satisfy the global Markov property with respect to the skeleton of the ttt. This condition is necessary (and sufficient) for the limit of

$$(X_v/X_u, v \in V) \mid X_u > t$$

to factorize into independent increments, in line with the results for Markov trees (Segers, 2020b) and Markov block graphs from Chapter 3.

Next we summarize the main findings and the main contributions of each of the chapters.

1.4 Chapter 2: Inference on Markov trees and the identifiability problem

The paper (identical to the chapter), builds on Segers (2020b), with focus on the Hüsler–Reiss parameterization and the identifiability criterion. The theory is supported by an application on the Seine river in France.

The paper starts by introducing the random vector $X = (X_v, v \in V)$ where each random variable of X is associated to a node of a tree, $\mathcal{T} = (V, E)$, and each of the edges has a weight $\theta^2_{ij} = \theta^2_{ji} > 0$ for $(i, j), (j, i) \in E$. We assume that X is in the domain of attraction of a Hüsler–Reiss distribution with parameter matrix, say $\Lambda = \{\lambda^2_{ij}\}_{i,j\in V}$. We denote this distribution by H_Λ. We assume that the element of the matrix, λ^2_{ij} for every $i, j \in V, i \neq j$, is proportional to the sum of the edge weights along the path from node i to node j, $p(i, j)$. Thus we have

$$\lambda^2_{ij} \propto \sum_{(a,b)\in p(i,j)} \theta^2_{ab}.$$

To motivate the existence of X and of H_Λ, we create the following object: Markov tree parameterized by bivariate Hüsler–Reiss copulas for every adjacent (neighboring) nodes, $i, j \in V, (i, j) \in E$, with tail dependence parameter $\theta^2_{ij} = \theta^2_{ji}$. We called this vector $Z^* = (Z^*_v, v \in V)$. We show that Z^* belongs to the domain of attraction of H_Λ. Because X and Z^* have the same domain of attraction, other asymptotic properties that hold for Z^* are valid for X too. For instance the following limit is shown to hold

$$\mathcal{L}(\ln X_v - \ln X_u, v \in V \setminus u) \mid X_u > t \xrightarrow{d} \mathcal{N}_{|V|-1}\left(\mu_u(\Lambda), \Sigma_u(\Lambda)\right), \tag{1.10}$$

where $\mathcal{N}_d$ is a d-variate Normal distribution. Recall that Λ depends on edge weights θ^2_{ij}, $(i, j) \in E$ and hence the mean vector and covariance matrix are linear combinations of the edge weights along (shared) paths:

$$\begin{aligned} \{\mu_u\}_v &\propto \sum_{(a,b)\in p(u,v)} \theta^2_{ab}, && v \in V \setminus u, \\ \{\Sigma_u\}_{ij} &\propto \sum_{(a,b)\in p(u,i)\cap p(u,j)} \theta^2_{ab}, && i, j \in V \setminus u. \end{aligned} \tag{1.11}$$

The proof of these results combines Segers (2020b, Theorem 1, Corollary 2, Example 3) applied to the Markov tree Z^* and the fact that X inherits limiting properties of Z^*.

As inference on the distribution H_Λ, we introduce three estimators inspired from classical estimation principles - method of moment estimator (MME), composite likelihood estimator (CLE) and estimator based on extremal coefficients (ECE). The first two estimators are constructed on the basis of the result in (1.10). The MME minimizes the distance between Σ_u and its sample estimate with respect to θ^2_{ij}, $(i, j) \in E$. The CLE uses the normal density. The MME borrows ideas from Engelke et al. (2014) and the one based on extremal coefficients from Einmahl et al. (2018). An extremal coefficient is a frequently used metric to measure tail dependence. For a subset $J \subseteq V$, the extremal coefficient of a set J is given by

$$\ell_J = -\ln G_J(1, \ldots, 1),$$

where G_J is the J-marginal distribution of the extreme value distribution G from Definition 1.

We apply these three estimators to real data of water levels measured on five stations on the Seine, near Paris. The schematic representation of the Seine network is shown in Figure 1.6. The nodes 2 and 5 correspond to locations where the water level has not been measured, hence the variables X_2, X_5 are unobserved (latent). This is where the *parameter identifiability problem* enters the picture.

We have to consider the distribution of the observable variables only. The set of nodes with observed variables is denoted by $U \subset V$. The subvector $X_U = (X_v, v \in U)$ is in the domain of attraction of the U-marginal distribution of H_Λ.

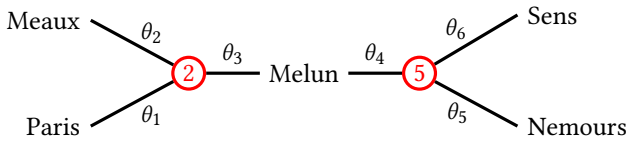


Figure 1.6: The Seine network with the tail dependence parameters associated to each edge of the tree.

According to Engelke and Hitz (2020, Example 7) the multivariate margins of the Hüsler–Reiss distribution are from the same family. We only need to take $\Lambda_U = \{\lambda^2_{ij}\}_{i,j\in U}$. Let H_{Λ_U} denote the U-marginal distribution of H_Λ. Given that we consider Λ_U the question is whether it is still possible to identify all the parameters θ_{ij}, $(i, j) \in E$ from H_{Λ_U}. The answer says that every node with latent variable should have at least three neighbors. This is a necessary and sufficient condition for the parameters' identifiability. The criterion is satisfied in the Seine network, hence in principle we are able to estimate all tail dependence parameters without having to modify the river network in order to avoid nodes with latent variables.

The application on the Seine in Section 2.5 presents several results. We consider confidence intervals on the parameters θ^2_{ij}, $(i, j) \in E$ based on a bootstrap procedure or on theoretical results from Einmahl et al. (2018). According to these, we see that all parameters θ^2_{ij}, $(i, j) \in E$ are different from zero. The goodness-of-fit of the approximating distribution in (1.10) is evaluated by comparing parametric estimates of the extremal coefficients with their non-parametric counterparts. As another check on the goodness-of-fit we present also estimates of the Pickands dependence function and bootstrapped confidence intervals as suggested in Kiriliouk et al. (2018). Finally we use several tools to discuss the extremal dependence within the Seine network, which appears rather strong, especially for flow-connected stations. Thanks to the fact that we can identify all parameters θ^2_{ij}, $(i, j) \in E$ even when X_2, X_5 are unobserved, we can compute extremal dependence between pairs of locations for which it would otherwise be impossible.

To conclude, the paper provides a method to study extremal dependence on rivers whose network is represented by a tree or it can be approximated by a tree. A main novelty of the paper is in the parameter identifiability criterion, a topic that is conceived in our paper for the first time. Another main novelty is the study of the extremal dependence of Seine. The data has been collected by us and used for the first time in this paper. To our knowledge this is the only study in the extreme value literature dedicated to flood risk of Seine. Showing $Z^* \in D(H_\Lambda)$ is also not a trivial result and it is the third main contribution to the literature.

Some novelties with minor importance are the following. The model for X is more general, it does not need to satisfy conditional independence relations. The only assumption on X is its domain of attraction of H_Λ. Because of this, however, it inherits limiting properties of Z^*, which is more specific, involving assumptions about Markovianity and a pair copula construction. The MME and the CLE, although not new in the literature, contain "local" estimation approach. When there are latent

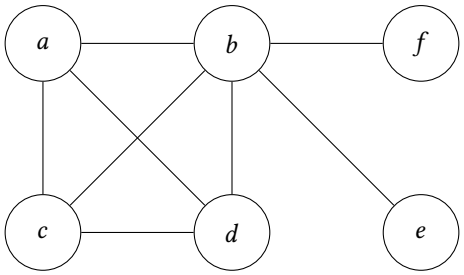


Figure 1.7: An example of a block graph with three blocks $\{a, b, c, d\}$, $\{b, e\}$, and $\{b, f\}$. Within a block all pairs must be connected.

variables, our estimators have to be modified with care to guarantee identifiability, as explained in Section 2.7.4. The idea of "local" estimation was introduced with the hope to reduce bias in the estimates. The simulations comparing the finite sample properties of the three estimators in Section 2.7.5 also represent a minor merit of the paper.

1.5 Chapter 3: Extreme value limits on Markov block graphs

The chapter is probabilistic in nature and it generalizes results from Segers (2020b) and Chapter 2.

A block graph is represented in Figure 1.7. It is considered a generalization of a tree (Le and Tuy, 2010): there is a unique shortest path between every pair of nodes and there are no cycles outside a block. A Markov block graph or a Markov random field with respect to a block graph is a random vector $X = (X_v, v \in V)$ on the node set of a block graph $\mathcal{G} = (V, E)$, which satisfies the global Markov property with respect to $\mathcal{G}$.

In Theorem 3.3.5 we generalize Segers (2020b, Theorem 1) for Markov block graphs. We assume that for each block, with set of nodes $C \subset V$, and every $u \in C$ we have

$$\lim_{t\to\infty} \mathcal{L}\{(X_v/t, v \in C) \mid X_u = t\} = \nu_{C,u}, \tag{1.12}$$

for some probability measure $\nu_{C,u}$. This assumption is the analogue to (1.9) for Markov trees. We also have an additional regularity condition which is common in literature on Markov chains (Smith, 1992; Perfekt, 1994; Resnick and Zeber, 2013). According to Theorem 3.3.5

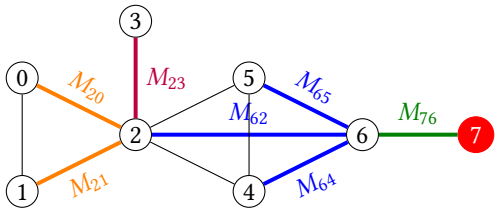
$$\{(X_v/t, v \in V \setminus u) \mid X_u = t\} \xrightarrow{d} (A_{uv}, v \in V \setminus u), \qquad t \to \infty, \tag{1.13}$$

where $A_{uv} = \prod_{(a,b)\in p(u,v)} M_{ab}$. The notation $p(u, v)$ is the *unique shortest* path between nodes u and v as a set of edges. For every $v \in V \setminus u$ the random variables $\big(M_{ab}, (a, b) \in p(u, v)\big)$ are independent between each other. For another node $w \in V \setminus u$, the shortest path $p(u, w)$ may share some of the edges with $p(u, v)$. Then the variables A_{uw} and A_{uv} share the same variables M_{ab} which belong to the shared edges. If there are two edges in $p(u, v)$ and $p(u, w)$, say (i, j) and (i', j'), which belong to the same block, then we have however M_{ij} and $M_{i'j'}$ dependent between each other.

The difference with respect to Theorem 1 in Segers (2020b) is that the factorization of the limiting variables happens along the *unique shortest* paths, because in a block graph, there may be more than one path between two nodes. We repeat for completeness the example used in Chapter 3.

Example 11 (Limit of Markov block graph) Consider the following block graph and a Markov field with respect to it, which satisfies the assumptions of

Theorem 3.3.5. We assume therefore that for every block the random vector X_C for $C \in \mathcal{C}$ satisfies (1.12) and a regularity condition mentioned earlier.



Suppose that the high threshold is exceeded at node 7. The conclusion of Theorem 3.3.5 is that as $t \to \infty$, we have

$$\big(X_v/t,\, v \in \{0,1,\dots,6\} \mid X_7 = t\big) \xrightarrow{d} \big(A_{7v},\, v \in \{0,1,\dots,6\}\big).$$

The limiting variables have the following structure:

$$\begin{aligned}
A_{7,\{0,1,2\}} &= (A_{70},\, A_{71},\, A_{72}) &&= M_{76}M_{62}(M_{20},\, M_{21},\, 1),\\
A_{7,\{2,3\}} &= (A_{72}, A_{73}) &&= M_{76}M_{62}(1,\, M_{23}),\\
A_{7,\{2,4,5,6\}} &= (A_{72},\, A_{74},\, A_{75},\, A_{76}) &&= M_{76}(M_{62}, M_{64}, M_{65}, 1),\\
A_{7,\{6\}} &= A_{76} &&= M_{76}.
\end{aligned} \tag{1.14}$$

The limit vector $(A_{7v})_v$ is similar to the one of a Markov field with respect to the tree formed by the unique shortest paths from node $u = 7$ to the other nodes. The variables M are independent from each other if they are in different colors, and dependent between each other if they are in the same color. ■

After this general result, we study the Hüsler–Reiss parameterization of the Markov block graph.

Consider the limiting probability distribution in (1.12), $\nu_{C,u}$. If the random vector X_C has a max-stable Hüsler–Reiss distribution with parameter matrix $\Delta_C = \{\delta^2_{ij}\}$, then $\nu_{C,u}$ is a multivariate log-normal distribution. The elements of Δ_C are given in (1.4) with $V = V_C$, the vertices of the block C. Thus, Δ_C is symmetric, $\delta^2_{ij} = \delta^2_{ji} > 0$, with zero diagonal. The parameters of the normal distribution associated to the log-normal distribution are determined by the edge weights in the following way

$$\begin{aligned}
\{\mu_{C,u}\}_v &\propto \delta^2_{uv}, && v \in V_C,\\
\{\Sigma_{C,u}\}_{ij} &\propto \delta^2_{ui} + \delta^2_{uj} - \delta^2_{ij}, && i,j \in V_C \setminus u.
\end{aligned} \tag{1.15}$$

However, instead of assuming that X_C has max-stable Hüsler–Reiss distribution, the subsequent analysis wouldn't change if we assume that X_C has any distribution, but it satisfies (1.12) with $\nu_{C,u}$ the log-normal distribution described above. In this way we avoid to assume that the distribution of X_C is itself max-stable. We make the same assumption for X_C for every $C \in \mathcal{C}$.

Let X be the concatenation of all $X_C, C \in \mathcal{C}$, described in the previous paragraph. We assume that X satisfies the global Markov property with respect to the block graph. According to Theorem 3.3.5 the limit of $X = (X_C, C \in \mathcal{C})$ in (1.13) involves multiplication of log-normal variables. Consider again Example 11. The random variables M_{76}, (M_{65}, M_{62}, M_{64}), $(M_{20}M_{21})$ and M_{23} are (jointly) log-normal with the parameters indicated in (1.15). If they are of different color, they are independent between each other, and if they are of the same color, they are dependent. The logarithm of $(A_{uv}, v \in V\setminus u)$ has then a multivariate normal distribution. Its parameters are described in terms of μ_C, Σ_C in (1.15), and involve the unique shortest paths from node u to any other node $v \in V \setminus u$.

We also show that $X = (X_C, C \in \mathcal{C})$ belongs to the domain of attraction of a Hüsler–Reiss distribution with parameter matrix $P(\Delta_C, C \in \mathcal{C}) = \{p_{ij}\}$ with

$$\{p_{ij}\} \propto \sum_{(a,b)\in p(i,j)} \delta_{ab}^2, \qquad i, j \in V.$$

We note this distribution H_P.

The model for X as just described is not to be confused with the extremal graphical model parameterized by Hüsler–Reiss Pareto distributions in Engelke and Hitz (2020). But there is a link between the two. The matrix $P(\Delta_C, C \in \mathcal{C})$ is the explicit solution of the matrix in Proposition 4 of Engelke and Hitz (2020).

In (1.3) when we take $G = H_P$ we would obtain a Hüsler–Reiss Pareto distribution. This distribution turns out to be an extremal graphical model as in Definition 9 with respect to the same graph $\mathcal{G}$ with respect to which the random vector X was assumed to be an ordinary graphical model.

We can conclude that the Pareto limit of a distribution parameterized blockwise through Hüsler–Reiss distributions is equal to the distribution of a vector parameterized blockwise by Hüsler–Reiss Pareto distributions.

What follows next is the parameters' identifiability criterion. Recall that when there are unobserved variables we are required to work with the $|U|$-variate distribution H_{P_U} with $P_U = \{p_{ij}\}_{i,j\in U}$ (U denotes the nodes with observed variables) to identify all non-zero parameters in Δ_C for all $C \in \mathcal{C}$. Not surprisingly the necessary and sufficient criterion is that every node with latent variable takes part in at least three blocks. Recall that the criterion for trees was that every node with latent variable has at least three neighbors.

To summarize, the main findings in Chapter 3 are related to EV limits of a general Markov block graph, and a Markov block graph parameterized through a family of Hüsler–Reiss distributions.

The topic of EV limits for Markov chains has been in the literature for quite some time (Smith, 1992; Segers, 2007; Papastathopoulos et al., 2017), and recently generalized to Markov trees (Segers, 2020b). In the context of Hüsler–Reiss parameterization, the link between an ordinary graphical model and an extremal graphical model is new and to our knowledge, there are no similar results for other parameterization. Similarly for the parameters' identifiability criterion - the problem has been introduced in our previous paper which focused on Markov trees and Chapter 3 provides a generalization to Markov block graphs.

The theory from Markov trees nicely generalizes to Markov block graphs, but the practical applicability of the latter seems more difficult to see. In Chapter 2 we have come up with at least one example - extreme water levels on a river network, which is well suited to be modeled as a random vector associated to a tree. With block graphs we could think of a block as a cluster. The variables within a block must be continuous to be suitable for extreme value analysis and must share similarities between each other. The variable that is shared by several clusters, or blocks, must have other properties that makes it similar to the variables in the other cluster too. Of course, the characteristics and similarities can be regarded as dominant, not simply present or absent. Suppose for instance, Figure 1.8 represents four different portfolios of assets. The variable at each node is the asset's return. The portfolio "Pink" contains assets $\{h, g, c\}$, that are predominantly exposed to risk(s) "Pink". However asset c also has significant part of risk(s) "Orange" which characterize(s) the assets in portfolio "Orange", $\{a, b, c, d\}$. Asset b is judged to be very sensitive to risks "Orange", "Gray" and "Yellow". However asset f is mainly characterized by risk(s) "Gray". As another example we could consider a banking system. The variable at every edge could be the inter bank lending rate of a certain bank. The cluster of $\{h, g, c\}$ could represent the

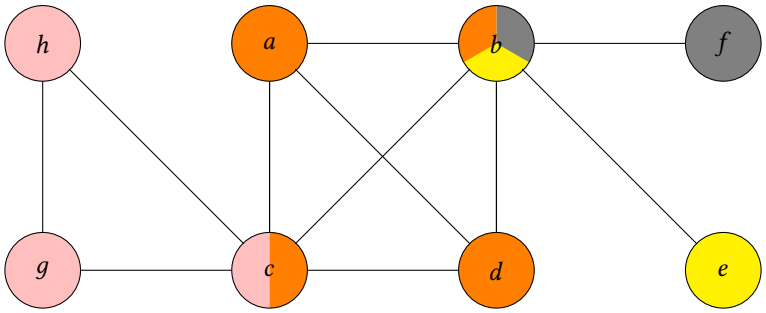


Figure 1.8: A graph representing links between four clusters - the "Pink", "Orange", "Gray" and "Yellow" clusters. Variable c is the link between the "Pink" and the "Orange" cluster. An extreme event in the orange cluster has to propagate to the pink cluster through the variable on node c.

banks of country "Pink" which do lend deposits between each other, but only bank c trades with banks from country "Orange". The banking systems of the four countries are connected between each other through the banks b and c which are active on more than one national market.

1.6 Chapter 4: Max-linear model on trees of transitive tournaments

The focus of this chapter are the max-linear Bayesian networks from Definition 7. Here and in Chapter 5 we consider a graph which we called a *tree of transitive tournaments*. It represents a DAG, whose skeleton (non-directed version) is a block graph. The idea is to study the same topics as in Chapters 2 and 3 but applied to a max-linear Bayesian network with respect to a directed acyclic block graph. The chapter is also purely probabilistic. We deal with the following main problems:

- specify the parameter space of the edge weights $\theta = (c_{ij}, (i,j) \in E)$;
- give a necessary and sufficient condition for $(X_v/X_u, v \in V \setminus u \mid X_u > t)$ to have the familiar limit involving factorization in independent multiplicative increments;
- relate this to the Markovianity of the model with respect to the undirected version of the ttt;
- study the parameter identifiability criterion.

A transitive tournament, say $\tau = (V_\tau, E_\tau)$, is a complete graph which is a DAG: so every pair of nodes $i, j \in V_\tau$ is connected by a directed edge $(i,j) \in E_\tau$ or $(j,i) \in E_\tau$ in such a way that τ is acyclic. Recall also that as soon as we talk about directed graphs we have parent node, $\mathrm{pa}(v), \mathrm{Pa}(v) = \mathrm{pa}(v) \cup \{v\}$; child nodes, more rarely used notation $\mathrm{ch}(v), \mathrm{Ch}(v) = \mathrm{ch}(v) \cup \{v\}$; ancestors, $\mathrm{an}(v), \mathrm{An}(v) = \mathrm{an}(v) \cup \{v\}$; and descendants, $\mathrm{desc}(v), \mathrm{Desc}(v) = \mathrm{desc}(v) \cup \{v\}$. In a transitive tournament on d nodes, there is one node with no parents (zero in-degree), one node with one parent, one node with two parents and so on, until one node with $d-1$ in-degree or parents. The equivalent statement is that there is one node with d children (d out-degree), one node with $d-1$ children and so on, until one node with no children at all. A ttt, $\mathcal{T} = (V, E)$, combines several transitive tournaments in a tree-like manner. This means that two transitive tournaments may have only one node in common and there cannot be a non-directed cycle involving several tournaments. A ttt is illustrated in Figure 1.9.

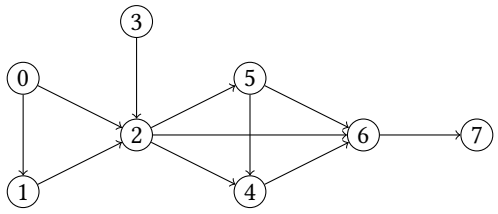


Figure 1.9: A ttt on four transitive tournaments, $\{0,1,2\},\{2,3\},\{2,4,5,6\},\{6,7\}$**. Each tournament is a DAG and the whole ttt is a DAG too. If we remove the directions, we obtain a block graph. Hence there are no undirected cycles involving several tournaments. If there is a directed path between two nodes, there is a *unique shortest directed* path. Because the skeleton is a block graph, there is a unique shortest undirected path between every pair of nodes.**

A max-linear model as in Definition 7 is parameterized by edge weights $(c_{ij} \geq 0, (i,j) \in E)$ and parameters $c_{ii} > 0, i \in V$. The assumptions we make on the parameter space are key to all the results. We have mentioned the notion of *critical* path introduced already in Gissibl and Klüppelberg (2018) and present in all subsequent literature related to max-linear models. Consider the paths from node 0 to 2 in Figure 1.9: one path is $p_1 = \{(0,2)\}$ and another is $p_2 = \{(0,1),(1,2)\}$. Consider the edge weights c_{01}, c_{02}, c_{12} and form the path products c_{02} and $c_{01}c_{12}$. The *critical* path is the one that has the highest path product, so either c_{02} or $c_{01}c_{12}$. We make the assumption that edge weights $(c_{ij}, (i,j) \in E)$ are such that all shortest paths are critical. In the example above this means that c_{01}, c_{02}, c_{12} are such that $c_{02} > c_{01}c_{12}$. In this way, we avoid that some edge weights are not represented in the distribution of X. With the assumption of criticality for the parameter space of $(c_{ij}, (i,j) \in E)$ the matrix coefficients $\{b_{ij}\}_{i,j} \in V$ can be expressed as

$$b_{ij} = c_{\overrightarrow{p}(j,i)} b_{jj}, \qquad j \in \mathrm{an}(i), \tag{1.16}$$

where

$$c_{\overrightarrow{p}(j,i)} = \prod_{(a,b)\in\overrightarrow{p}(j,i)} c_{ab}. \tag{1.17}$$

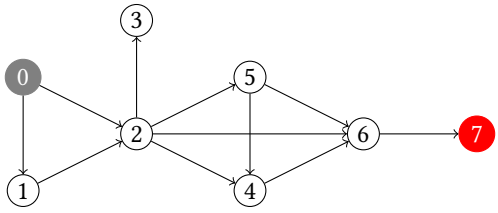
and $\overrightarrow{p}(j,i)$ denotes the unique shortest directed path between nodes j and i.

We show that in a ttt with *unique source* (node without parents in the whole ttt), the limiting vector in

$$(X_v/X_u, v \in V \setminus u \mid X_u > t) \xrightarrow{d} (A_{uv}, v \in V \setminus u), \qquad t \to \infty \tag{1.18}$$

can be represented as a factorization of independent variables along the unique shortest *undirected* path from u to any other node in $V \setminus u$. This is in analogy with already seen results related to Markov trees and Markov block graphs in the previous chapters. Also, we show that if the source is *not* unique, there is no $u \in V$ such that a factorization of $(A_{uv}, v \in V \setminus u)$ in independent variables is possible.

Example 12 (Factorization of a max-linear model) Suppose X is a max-linear model with respect to the following ttt with unique source at node 0. Let the high threshold be exceeded at node 7, so $u = 7$.



The random vector $(A_{7v}, v \in V \setminus 7)$, as a limit of $(X_v/X_7, v \in V \setminus 7 \mid X_7 > t)$ as $t \to \infty$, has the same form as the one in (1.14). The variables in the same color are dependent between each other and the variables in different color are independent from each other. In a max-linear model the increments have jointly discrete distribution. The result follows from Segers (2020b, Example 1). The important message is that for max-linear models the factorization for every $u \in V$ happens if and only if the ttt has a unique source. ■

What underlies this result is the fact that X satisfies the global Markov property with respect to the skeleton of the ttt if and only if there is a unique source. This is our second major result in the paper.

The main challenge in this paper was the identifiability criterion in the context of the max-linear model on a ttt with unique source.

In the case of a max-linear model with respect to a ttt the problem is the same as the one in Chapters 2 and 3. Namely, suppose that there are nodes in the ttt whose variables are unobserved. We have to work only with the distribution of X_U, where U is the set of nodes with observed variables. Earlier, we mentioned that the distribution of a max-linear model is an extreme value distribution with discrete angular measure (see Definition 2), say a measure $\mathcal{S}$ with atoms $\{a_v\}_{v\in V}$ and masses $\{m_v\}_{v\in V}$. The atoms and masses are defined in terms of the coefficients $\{b_{ij}\}_{i,j\in V}$ from Definition 7 and accordingly in terms of the edge weights $(c_{ij}, (i,j) \in E)$. The distribution of the subvector X_U is in the same family, say a measure $\mathcal{S}_U$, but its atoms and masses, say $\{\alpha_r\}$ and $\{\mu_r\}$ respectively, are different than $\{a_v\}_{v\in V}$ and $\{m_v\}_{v\in V}$, but still functions of $\{b_{ij}\}$ and $(c_{ij}, (i,j) \in E)$. The problem is that $\mathcal{S}_U$ might not be bijection of $(c_{ij}, (i,j) \in E)$ and hence two different sets of edge weights might generate the same $\mathcal{S}_U$. The parameter identifiability problem consists of giving a criterion such that edge weights $(c_{ij}, (i,j) \in E)$ are uniquely identified from $\mathcal{S}_U$. A pair of necessary and sufficient conditions in the case when the max-linear model is with respect to a ttt with unique source is that

(i) every node with latent variable has at least two children and
(ii) every node with latent variable is the source of some tournament.

To illustrate the criterion, consider the graph in Example 12 with unique source at node 0. The nodes that may contain a latent variable are 0 and 2. Nodes 1,3,4,6,7 have only one child, and node 5 although having two children, is not the source of any tournament.

The first contribution of the paper is the necessary and sufficient condition for the factorization of the limiting distribution in (1.18). The second one links this to Markovianity of the random vector with respect to the undirected version of the graph. The third contribution is the necessary and sufficient criterion for the parameter identifiability if we observe a subvector of the max-linear model X. There are many side results accompanying these three major statements.

The practical applications may look limited at first. We could think of rivers again – the delta of a river or a portion of a river where the stream is such that the source is unique. We could also use similar financial examples as in the previous section on block graphs. A ttt represents interactions between clusters. Given that within every tournament there is a degree ordering we can use an ordering according to some criterion, thereby imposing some hierarchy within a cluster. Literature on extreme value theory has seen many uncommon applications – such as in neurobiology, drugs testing, road security, IT network traffic, material robustness and corrosion. An appropriate application, although not straightforward to find, might exist and is a topic of additional research.

1.7 Chapter 5: Sum-linear models and their similarity with max-linear models

This chapter gathers results on sum-linear models as in Definition 8. The topics are familiar – factorization of the limiting variables into independent increments, Markovianity with respect to the undirected graph and the parameter identifiability criterion. It turns out that the two of the propositions for the sum-linear model are *identical* to these of the max-linear model. Exception makes the Markovianity of the sum-linear model, where we cannot provide the necessity part, but we conjecture it is true. What regards the factorization into independent increments and the parameter identifiability, we can use exactly the same statements of the two main propositions discussed in the previous section, replacing the max-linear model with a sum-linear model. However the proofs are different in many respects. In what follows we will outline the differences and the similarities between the two models.

The first difference is in the parameter space for the edge weights $\theta = (c_{ij}, (i,j) \in E)$. Recall that we have a ttt, $\mathcal{T} = (V, E)$, with edge weights $\theta = (c_{ij}, (i,j) \in E)$. For a sum-linear model X as in Definition 8 the notion of criticality applicable for max-linear models does not play a role.

For a sum-linear model we show that the matrix B is given by

$$b_{ij} = d_{jj} \left(\sum_{\overrightarrow{p} \in \pi(j,i)} c_{\overrightarrow{p}} \right), \qquad i \neq j, \tag{1.19}$$

where $\pi(j,i)$ is a set of all directed paths from j to i, $\overrightarrow{p}$ is an element of this set and it is composed of edges, and $c_{\overrightarrow{p}} = \prod_{(a,b) \in \overrightarrow{p}} c_{ab}$.

The second difference with respect to the max-linear model is in the coefficients $\{b_{ij}\}$: compare (1.19) with (1.16). In the max-linear model, as a consequence of the criticality assumption that we make on all shortest paths, the edge weights along the unique shortest directed path matter. In the sum-linear model, only the edge weights along all directed paths are involved. This difference induces a different method in the proof of the parameter identifiability criterion with relation to the one used for max-linear models. Some of the proofs regarding the sum-linear model are possible using classical matrix algebra, as opposed to using so called *tropical* algebra in the case of a max-linear model.

A major similarity between the two models is that they are both in the domain of attraction of a distribution of the same family (Einmahl et al., 2012, Lemma 6.1). Therefore the angular measure of this extreme value distribution for both models is of the same family too. We use this result to show that the limit in (1.18) when X is sum-linear is of the same form as when X is max-linear. In particular, this distribution is discrete and it is given by (Segers, 2020b, Example 1)

$$\sum_{i \in V} b_{ui} \delta_{(b_{vi}/b_{ui}, v \in V)}$$

where $\{b_{ij}\}_{i,j \in V}$ can be the coefficients of the max-linear or of the sum-linear model. Because these however are different, the two probability measures are not the same, but they do share the same form. In the case when the ttt is a tree, there is no difference between the two models.

Another important result states that the summation term in (1.19) can be factorized along the unique shortest directed path, which we will denote here by $\overrightarrow{p}(j,i)$. Let $\{u = v_1, \ldots, v_n = v\}$ be the set of nodes on the unique shortest directed path between

a pair of nodes $u, v \in V$. Then we have

$$\sum_{\overrightarrow{p} \in \pi(u,v)} c_{\overrightarrow{p}} = \prod_{i=1}^{n-1} \left(\sum_{\overrightarrow{p} \in \pi(v_i, v_{i+1})} c_{\overrightarrow{p}} \right). \tag{1.20}$$

Thanks to the two result above we obtain the factorization of the limit in (1.18) on a ttt with unique source. In exactly the same way as for the max-linear model, the limiting variables $(A_{uv}, v \in V \setminus u)$ from (1.18) factorize into independent variables along the unique shortest *undirected* paths, if and only if the ttt has a unique source. The proof is also similar to the proof of the analogous proposition for the max-linear model. This is due to the fact that in view of (1.20) the coefficient $b_{ij}, i, j \in V$ in both models become products of terms along the unique shortest directed path from j to i.

The sum-linear model is also Markov with respect to the undirected version of the ttt, provided the ttt has a unique source. However, since the distribution of the sum-linear model is absolutely continuous we do not need the theory of conditional independence in max-linear models, but use classical results from conditional independence in directed graphs (Lauritzen, 1996; Koller and Friedman, 2009).

Recall the parameter identifiability problem – to identify uniquely all edge weights $(c_{ij}, (i, j) \in E)$ from the distribution of a subvector X_U which is again a sum-linear model. We do not have an analytical expression of its distribution but we approach the problem by studying the angular measure of the extreme value distribution to which domain of attraction X_U belongs, say $\mathcal{S}_U$. This angular measure is discrete by Lemma 6.1 in Einmahl et al. (2012), and its masses $\{\mu_r\}$ and atoms $\{\alpha_r\}$ are functions of the coefficients $\{b_{ij}\}$ and of the edge weights $(c_{ij}, (i, j) \in E)$. However, is $\mathcal{S}_U$ a bijection of $(c_{ij}, (i, j) \in E)$? The necessary and sufficient conditions are the same as in (i) and (ii) in the section on max-linear models. In this way we may use some of the tools from the proof of the parameter identifiability criterion for the max-linear model. However, the different structure of $\{b_{ij}\}_{i,j \in V}$ of the sum-linear model requires a different approach mainly in the sufficiency part of the proof.

The main message of Chapter 5 is the striking similarity of both models and at the same time, the difference in the proofs' approaches. The sum- and the max-linear models share exactly the same properties in regard to the nature of their tail behavior, Markovianity with respect to the undirected graph and the parameter identifiability criterion. This similarity is induced by the fact that the two models share the same domain of attraction and by the factorization of the coefficients $\{b_{ij}\}$ in (1.20). Thanks to this last property, the coefficients $\{b_{ij}\}$ of both models can be factorized along the unique shortest directed paths. Differences in the proofs of these properties arise from the different parameter space for $(c_{ij}, (i, j) \in E)$. In the max-linear model, the criticality of shortest paths was necessary to assure validity of results. We have also different expressions of $\{b_{ij}\}$ – in the max-linear model only edge weights on unique shortest directed paths are involved, whereas in the sum-linear model, edge weights along all possible paths take part. Also sum-linear models may use classical matrix algebra and theory of conditional independence, as opposed to tropical algebra and new separation concepts developed for max-linear models.

1.8 Chapter 6: R package `gremes` – Estimation of Tail Dependence in Graphical Models

Packages related to extremes and graphical models are few in number. To my knowledge there is the package `graphicalExtremes` on CRAN which is based on Engelke and Hitz (2020) and which presents tools for exact simulation and inference on Hüsler–Reiss extremal graphical models. The package `gremes` is the second one with focus

on graphical models and extremes. It is mainly dedicated to the models in Chapter 2, hence Markov trees. It provides tools for estimation of the tail dependence parameters in graphical models parameterized by family of Hüsler–Reiss distributions. The only supported graphs are *trees* and *block graphs*. The estimation methods are variations of method of moments (Engelke, Malinowski, Kabluchko, and Schlather, 2014; Asenova, Mazo, and Segers, 2021), maximum likelihood (Asenova, Mazo, and Segers, 2021; Engelke and Hitz, 2020) and a method based on extremal coefficients (Einmahl, Kiriliouk, and Segers, 2018).

There is a rich documentation accompanying the package: about fifteen vignettes explain the main and additional functionalities. Estimators are theoretically motivated, to a sufficient detail so that the user rarely needs to consult the referenced articles. All notation and procedures used from the estimators are detailed. The use of each estimator is demonstrated in a separate vignette. The website of the package `www.gremes.info` presents more or less the content of the vignettes.

The package is developed in an *object-oriented* style. The classes are type `S3`. There are two main types of objects.

- An object containing the graph and the dataset is created using classes `Network`, `Tree`, `BlockGraph`, and subclasses of these.
- An object containing the graph and the edge weights is created with classes `HRMnetwork`, `HRMtree`, `HRMBG`, and subclasses of these.

The first type of objects represents the non-parametric view on the problem - all we know is the graph and the data. The second type of objects represents the Hüsler–Reiss parametric model: every clique is parameterized by a Hüsler–Reiss distribution with parameters - the edge weights within this clique. Hence all that characterizes the parametric model is the graph and the edge weights.

Consider for instance the method `extrCoeff` which is written both for classes `Tree` and `HRMtree`. If we pass an object of class `Tree` to the method `extrCoeff`, the command will return non-parametric estimates of the extremal coefficients. If the object passed is of class `HRMtree`, parametric extremal coefficients will be returned.

The main goal of the package is estimation, therefore the method `estimate` is the key functionality of the tools provided in the package. Estimation in `gremes` happens by using the method `estimate` on an object from one of the following classes:

- `MME`, `MLE`, `MLE1`, `MLE2`, `EKS`, `EKS_part`, `EngHitz`, `MMEave`, `MLEave` in which case it estimates the edge weights on a tree.
- `HRMBG` in which case it estimates the edge weights on a block graph.

Since most of the estimation methods apply to models with respect to trees we present the theoretical motivation only for these estimators. Let $T = (V, E)$ be a tree with node set V and edge set E. Consider a $|V|$-variate random vector $X = (X_v, v \in V)$ for which it holds: X satisfies the global Markov property with respect to the tree T; every bivariate distribution between two adjacent variables uses a bivariate Hüsler–Reiss copula with parameter θ_{ij} for some edge $(i, j) \in E$; univariate margins are standardized to the unit-Pareto distribution.

Based on different asymptotic results of X, we can have different estimators of the parameters θ_{ij}, $(i, j) \in E$, which we collect in a vector $\theta \in (0, \infty)^E$.

(L1) Consider the log-differences conditional on a high threshold being exceeded at a particular node

$$(\ln X_v - \ln X_u, v \in V) \mid X_u > t, \qquad t \to \infty.$$

It can be shown that the limiting distribution of the vector above is multivariate Gaussian distribution with mean vector $\mu_u(\theta)$ and covariance matrix $\Sigma_u(\theta)$ which depend on the edge weights and on the particular node u. This is a result from Chapter 2 or also Asenova, Mazo, and Segers (2021) and references therein. The method of moments type estimator and the composite likelihood estimator both aim at estimating Σ_u and accordingly θ. The estimators are implemented in methods `estimate.MME`, `estimate.MLE`, `estimate.MLE1` and `estimate.MLE2`.

(L2) Consider the limiting distribution of the scaled componentwise maxima if we dispose of a random sample of size n of X, $\{X_{v,i}\}_{i\in 1,\ldots,n; v\in V}$

$$\left(\frac{1}{n}\max_{i=1,\ldots,n} X_{v,i}, v \in V\right), \qquad n \to \infty.$$

The limit, shown in Chapter 2 or also Asenova, Mazo, and Segers (2021), is a max-stable Hüsler–Reiss copula with unit Fréchet margins and with parameter matrix Λ given by

$$\big(\Lambda(\theta)\big)_{ij} = \lambda_{ij}^2(\theta) \propto \sum_{(a,b)\in p(i,j)} \theta_{(a,b)}^2, \qquad i,j \in V,\ i \neq j, (a,b) \in E. \tag{1.21}$$

The extremal coefficient estimator, introduced in Einmahl, Kiriliouk, and Segers (2018) and used in Asenova, Mazo, and Segers (2021) is based on bi- and trivariate extremal coefficients derived from this max-stable distribution. The estimator is implemented in methods `estimate.EKS`, `estimate.EKS_part`.

(L3) Consider the scaled random vector, given that the maximum exceeds a high threshold:

$$(X_v, v \in V)/t \mid \max_{v\in V} X_v > t, \qquad t \to \infty.$$

The limit is a so called Hüsler-Reiss Pareto (Engelke and Hitz, 2020) distribution with the same matrix Λ as in (6.1). Proof of this limit is provided in Asenova and Segers (2021), see also Engelke and Hitz (2020). For Hüsler–Reiss Pareto distributions with respect to trees, Engelke and Hitz (2020) presents a cliquewise estimator which has been implemented in `gremes` and called 'Engelke and Hitz' estimator. We have adapted it to make it suitable when there are latent variables. More details are presented in Chapter 6. The estimator is implemented in method `estimate.EngHitz`.

(L4) Consider the differences with respect to the mean of the log-transformed variables

$$(\ln X_v - \overline{\ln X}, v \in V) \mid \overline{\ln X} > t, \qquad t \to \infty,$$

where $\overline{\ln X} = (1/|V|)\sum_{v\in V} \ln X_v$. The limit of this vector is also a multivariate Gaussian distribution with mean and covariance matrix, say $\bar{\Sigma}$, that contain the matrix $\Lambda(\theta)$ in (6.1). This asymptotic result is shown in an unpublished note Segers (2019) and details are not provided here. The package offers method of moments and maximum composite likelihood estimates, The estimators are implemented in methods `estimate.MMEave`, `estimate.MLEave`.

Other functionalities present in the package are a method for simulating from the exact distribution of a Markov tree, parameterized by bivariate Hüsler–Reiss copulas; tools for post-estimation analysis, such as methods calculating parametric or non-parametric stable tail dependence functions, extremal coefficients, and tail dependence coefficients.

1.9 Bibliographic notes

Journal Publications

- Asenova, S., Mazo, G. & Segers, J. Inference on extremal dependence in the domain of attraction of a structured Hüsler–Reiss distribution motivated by a Markov tree with latent variables. Extremes 24, 461–500 (2021).
- Asenova, S. & Segers, J. Extremes of Markov random fields on block graphs: max-stable limits and structured Hüsler-Reiss distributions. arXiv:2112.04847. Submitted to journal Extremes, waiting for second revision, first revision with positive recommendation.
- Asenova, S. & Segers, J. Max-linear graphical models with heavy-tailed factors on trees of transitive tournaments. Submitted to Applied Probability Journals. `https://arxiv.org/abs/2209.14938`

Posters

- Linear association models with heavy tail errors on trees of transitive tournaments. Conference 30 years Institute of Statistics, Biostatistics and Actuarial sciences (ISBA) at UCLouvain, May 2022, Louvain-la-Neuve.
- Latent variables in Hüsler-Reiss Markov trees. International conference Extreme Value Analysis (EVA2019), July 2019, Zagreb.
- Extremal dependence in Hüsler-Reiss Markov trees. Royal Statistical Society of Belgium Annual meeting, October 2018, Ovifat.
- Extremal dependence in Hüsler-Reiss Markov trees. Workshop on Rare Events, Extremes and Machine Learning, May 2018, Paris.

Miscellaneous Contributions

Asenova, S., (2022). `gremes` – Estimation of Tail Dependence in Graphical Models. R package version 0.1.0. `www.gremes.info`.

Inference on extremal dependence in the domain of attraction of a structured Hüsler–Reiss distribution motivated by a Markov tree with latent variables

2

This chapter corresponds to an article prepared jointly with Johan Segers and Gildas Mazo, carrying the same name as the chapter and published in the journal *Extremes* in 2021.

2.1 Introduction

A major topic in multivariate extreme value theory is the modeling of tail dependence between a finite number of variables. Informally, tail dependence represents the degree of association between the extreme values of these variables. Probabilistic graphical models (Lauritzen, 1996; Koller and Friedman, 2009; Wainwright et al., 2008), are distributions which embody a set of conditional independence relations and have a graph-based representation, according to which the nodes of the graph are associated to the variables and the set of edges encode the conditional independence relations. The intersection of the two fields, extreme value theory and probabilistic graphical models, gives rise to the study of the tail behavior of graphical models.

Consider a river network where the interest is in extreme water levels or water flow in relation to flood risks. Figure 2.1 illustrates part of the Seine network. The graph fixed by the seven labeled nodes and the river channels between them can be a base for building a model for extremal dependence between the water levels at these sites.

Hydrological data are often used to fit models for multivariate extremes based on graphs. Water flows of the Bavarian Danube are analyzed in Engelke and Hitz (2020). Lee and Joe (2017) study water flows of the Fraser river, British Colombia. Precipitation data in the Japanese archipelago is treated in Yu et al. (2016), where the model is based on a spatial grid viewed as an ensemble of trees. Other extreme-value models involving graphs appear in Einmahl et al. (2018) and Lee and Joe (2017), who study financial data under different models. The first paper uses max-linear models on a directed acyclic graph (DAG) (Gissibl and Klüppelberg, 2018), and the second one

a 1-factor model. Klüppelberg and Sönmez (2022) introduce an infinite max-linear model to analyze the distribution of extreme opinions in a social network.

Relatively recently the relation between extreme value distributions and conditional independence assumptions has been given theoretical relevance. The earliest is the article of Gissibl and Klüppelberg (2018) introducing max-linear models as structural equation models on a DAG, followed by the regularly varying Markov trees in Segers (2020b) and the extremal graphical models in Engelke and Hitz (2020) based on multivariate Pareto distributions. Earlier, Papastathopoulos and Strokorb (2016) showed that for a max-stable random vector with positive and continuous density, conditional independence implies unconditional independence, thereby concluding that a broad class of max-stable distributions does not exhibit an interesting Markov structure.

A key object of our paper is the multivariate Hüsler–Reiss distribution (Hüsler and Reiss, 1989) with parameter matrix having a particular structure linked to a tree as specified in Eq. (2.5). The structure is motivated by the fact that the max-domain of attraction of the said Hüsler–Reiss distribution contains certain regularly varying Markov trees. The latter property follows from results in Segers (2020b) and sets our work apart from the extremal graphical models in Engelke and Hitz (2020), who impose a non-standard conditional independence relation on the multivariate Pareto distribution associated to a max-stable distribution, but without regard for the latter's max-domain of attraction. Still, it turns out that for trees, the structured Hüsler–Reiss models in Engelke and Hitz (2020) and in our paper are the same, as explained in Section 2.7.1. Another structured Hüsler–Reiss distribution based on trees is proposed in Lee and Joe (2017). The form they propose is genuinely different from ours, however, as explained in detail in Section 2.7.2.

We consider random samples from the distribution of a random vector $\xi = (\xi_v, v \in V)$ with continuous margins whose variables are indexed by the node set $V = \{1, \ldots, d\}$ of an undirected tree with edge set E. After marginal standardization to the unit-Pareto distribution, we assume that the random vector is in the max-domain of attraction of the tree-structured Hüsler–Reiss distribution described in the previous paragraph. We emphasize that we do not assume that ξ itself satisfies any conditional independence relations with respect to the tree. The tree only comes into play via the imposed structure on the parameter matrix of the max-stable Hüsler–Reiss distribution containing the distribution of the standardized version of ξ in its max-domain of attraction.

The main result and contribution of our paper is a criterion for identifiability of all $d - 1$ parameters $\theta_e \in (0, \infty)$ for $e \in E$ of the tree-structured d-variate Hüsler–Reiss distribution in case some of the d variables are latent (unobservable). To illustrate why the problem of latent variables is relevant, consider again the Seine network on Figure 2.1. The red dots designate junctions of two river channels (conversely, in a river delta, a channel could split into several ones). No measurement stations being present there, we cannot observe the water levels at those locations. We propose to treat those water levels as latent variables. The question is then whether it is still possible to identify all $d - 1$ parameters. The answer is a surprisingly simple identifiability criterion: it is necessary and sufficient that all nodes indexing latent variables have degree at least three. The important practical implication is that, provided the criterion is met, the latent variables can be included in the model, reflecting the dependence structure more accurately than when they would have been ignored.

Given a random sample from a distribution in the max-domain of attraction of the tree-structured Hüsler–Reiss distribution, we propose three types of estimators of the edge parameters: a first one called method of moments estimator (MME) is based on the estimator proposed in Engelke et al. (2014), a second one is based on the composite likelihood function (composite likelihood estimator or CLE) and the

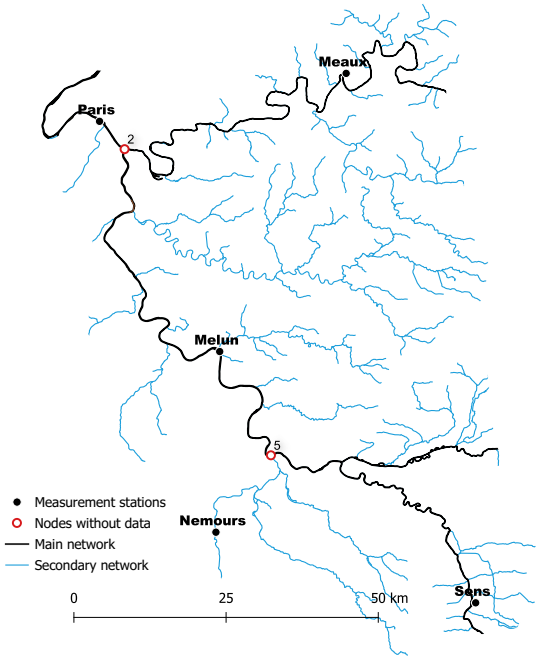


Figure 2.1: Seine network. The data is from the web-site of Copernicus Land Monitoring Service: https://land.copernicus.eu/imagery-in-situ.

third one is essentially the pairwise extremal coefficient estimator (ECE) introduced in Einmahl et al. (2018). All estimators proposed allow for the fact that some of the d variables are latent, provided the identifiability criterion is met.

We illustrate the method by a detailed analysis of data on high water levels at several locations of the Seine network. The network is represented schematically as a tree with seven nodes indexing five observable variables and two latent ones. As the identifiability criterion is met, we can estimate the six dependence parameters of the tree-structured Hüsler–Reiss distribution, each parameter corresponding to an edge in the tree. For the three proposed estimators we compute parameter estimates and confidence intervals. We assess the goodness-of-fit by comparing the model output with various non-parametric measures of tail dependence. Finally, we compare the fitted tail dependence model incorporating latent variables with a model where the latent variables are ignored.

The outline of the paper is as follows: Section 2.2 presents some general theory and describes the model to which the identifiability criterion is applied. The latter is the focus of Section 2.3. Section 2.4 introduces the three estimators, used for statistical inference and Section 2.5 is dedicated to the study of high water levels on the Seine network. Concluding remarks and perspectives for further research are discussed in Section 3.5. The supplements' section provides proofs that are not in the text, a numerical comparison between our structured Hüsler–Reiss method and the one of Lee and Joe (2017), clarification on the relationship between the different objects in our paper and the objects in Engelke and Hitz (2020), some simulation results which aim at comparing the different estimators, and details about some estimation procedures and the data preprocessing.

2.2 The model – definition and properties

2.2.1 Preliminaries

Multivariate extremes. Let $V = \{1, \dots, d\}$ for some integer $d \geq 2$. A d-variate max-stable distribution G is called simple if its margins are unit-Fréchet, that is, a random vector Z with distribution G satisfies $\mathbb{P}(Z_v \leq x) = \exp(-1/x)$ for $x \in (0, \infty)$ and $v \in V$. Let $X = (X_v, v \in V)$ be a random vector with unit-Pareto margins, i.e., $\mathbb{P}(X_v \leq x) = 1 - 1/x$ for $x \in [1, \infty)$ and $v \in V$. Let $X_i = (X_{v,i}, v \in V)$ for $i = 1, \dots, n$ be an independent random sample from the distribution of X. We say that X belongs to the max-domain of attraction of the simple max-stable distribution G, notation $X \in D(G)$, if

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{i=1,\dots,n} X_{v,i} \leq n z_v, v \in V \right) = G(z), \qquad z \in (0, \infty)^d.$$

For more background on max-stable distributions and their domains of attractions, we refer to the reader to Resnick (1987, Chapter 5) and de Haan and Ferreira (2007, Chapter 6).

Throughout the paper the stable tail dependence function (stdf) l of G or $X \in D(G)$ will appear frequently. It is defined as

$$l(x) = \lim_{t\to\infty} t \left(1 - \mathbb{P}(X_v \leq t/x_v, v \in V)\right) = -\ln G(1/x_v, v \in V), \tag{2.1}$$

for $x \in [0, \infty)^d$, with the obvious limit interpretation if $x_v = 0$ for some $v \in V$. The stdf is closely linked to the exponent function of a simple max-stable distribution in Coles and Tawn (1991, Eq. (2.4)). It is introduced and studied in Huang (1992) and Drees and Huang (1998); see also later literature in de Haan and Ferreira (2007, Chapter 6) and Beirlant et al. (2004, Chapter 8). The stdf evaluated at $x_J = (1_{\{j \in J\}}, j \in V)$ is known as an extremal coefficient, of which we make use in Sections 2.4 and 2.5.

One of the main objects in our paper is the multivariate Hüsler–Reiss distribution. This absolutely continuous max-stable distribution was introduced in Hüsler and Reiss (1989) and remains a popular parametric model in recent literature (Genton et al., 2011; Huser and Davison, 2013; Asadi et al., 2015; Engelke et al., 2014; Einmahl et al., 2018; Lee and Joe, 2017). It arises as the limiting distribution of partial maxima of a triangular array of row-wise independent and identically distributed random vectors from a multivariate normal distribution with correlation matrix $\rho(n)$ depending on the sample size n. In particular, assume that

$$\lim_{n\to\infty} \left(1 - \rho_{ij}(n)\right) \ln n = \lambda_{ij}^2 \in (0, \infty)$$

for every pair of variables $i, j \in V$ and let $\Lambda = (\lambda_{ij}^2)_{i,j\in V}$ denote this limiting matrix. Note that $\lambda_{ii}^2 = 0$ for every $i \in V$. For every subset $W \subseteq V$ and any element $u \in W$ let $\Gamma_{W,u}(\Lambda)$ be the square matrix of size $|W| - 1$ with elements

$$\left(\Gamma_{W,u}(\Lambda)\right)_{ij} = 2(\lambda_{iu}^2 + \lambda_{ju}^2 - \lambda_{ij}^2), \qquad i, j \in W \setminus u. \tag{2.2}$$

Nikoloulopoulos et al. (2009) and later Genton et al. (2011) and Huser and Davison (2013) show that the cumulative distribution function (cdf) as deduced by Hüsler and Reiss (1989) can be written as

$$H_\Lambda(z) = \exp\left\{ -\sum_{u\in V} \frac{1}{z_u} \Phi_{d-1}\left(\ln \frac{z_v}{z_u} + 2\lambda_{uv}^2, v \in V \setminus u; \Gamma_{V,u}(\Lambda) \right) \right\}, \tag{2.3}$$

where $z \in (0, \infty)^d$, and $\Phi_p(\,\cdot\,; \Sigma)$ denotes the p-variate zero mean Gaussian cdf with covariance matrix Σ. The distribution H_Λ in (2.3) is a simple max-stable distribution.

In particular, its margins are unit-Fréchet, whereas Hüsler and Reiss (1989) originally proposed the distribution in terms of Gumbel margins.

Multivariate margins of the d-variate Hüsler–Reiss distribution are Hüsler–Reiss distributions too. The corresponding parameter matrix is obtained by selecting the appropriate rows and columns in the original parameter matrix (see, e.g., Engelke and Hitz, 2020, Example 7). In particular, if $X \in D(H_\Lambda)$ and if $U \subseteq V$ is non-empty, the stdf l_U of $X_U = (X_u, u \in U)$ is

$$l_U(x) = \sum_{u \in U} x_u \, \Phi_{|U \setminus u|}\left(\ln \frac{x_u}{x_v} + 2\lambda_{uv}^2, v \in U \setminus u; \Gamma_{U,u}(\Lambda) \right), \quad x \in [0, \infty)^U. \tag{2.4}$$

Here we write $U \setminus u$ instead of $U \setminus \{u\}$. In case $x_u = 0$ for some $u \in U$, the corresponding term in the sum in (2.4) vanishes.

Trees. We will need some notions from graph theory. A graph is a pair $\mathcal{G} = (V, E)$ where $V = \{1, \ldots, d\}$ is the set of nodes or vertices and $E \subseteq \{(a, b) \in V \times V : a \neq b\}$ is the set of edges. Edges will also be denoted by $e = (a, b) \in E$. The number of vertices in a subset $U \subseteq V$ will be denoted by $|U|$, while d is reserved for $|V|$ only. A graph is undirected if $(a, b) \in E$ is equivalent to $(b, a) \in E$. A path $(u \rightsquigarrow v)$ from node u to node v is a collection $\{(u_0, u_1), (u_1, u_2), \ldots, (u_{n-1}, u_n)\}$ of distinct, directed edges such that $u_0 = u$ and $u_n = v$. An undirected tree is an acyclic undirected graph $\mathcal{T} = (V, E)$ such that for every pair of distinct nodes a and b there is a unique path $(a \rightsquigarrow b)$.

2.2.2 Model definition

Let $\mathcal{T} = (V, E)$ be an undirected tree with node set $V = \{1, \ldots, d\}$ and let $\xi = (\xi_v, v \in V)$ be a random vector with joint cdf F and continuous margins $F_v(z) = \mathbb{P}(\xi_v \leq z)$ for $z \in \mathbb{R}$ and $v \in V$. Let the random vector $X = (X_v, v \in V)$ be defined as $X_v = 1/\big(1 - F_v(\xi_v)\big)$ for every $v \in V$. Because the functions F_v for $v \in V$ are continuous, the marginal distributions of X are unit-Pareto.

We assume that X is in the max-domain of attraction of the Hüsler–Reiss distribution H_Λ in (2.3) with $\Lambda = (\lambda_{ij}^2)_{i,j \in V}$ having the following structure linked to the tree $\mathcal{T}$: there exists a vector $\theta = (\theta_e)_{e \in E}$ of positive scalars with $\theta_{ab} = \theta_{ba}$ and such that $\Lambda = \Lambda(\theta)$ where

$$\big(\Lambda(\theta)\big)_{ij} = \lambda_{ij}^2(\theta) = \frac{1}{4} \sum_{e \in (i \rightsquigarrow j)} \theta_e^2, \qquad i, j \in V, \; i \neq j. \tag{2.5}$$

The assumption can thus be written compactly as $X \in D(H_{\Lambda(\theta)})$ for some $\theta \in (0, \infty)^E$.

The motivation for the proposed structure is that $H_{\Lambda(\theta)}$ contains in its max-domain of attraction a certain graphical model with respect to $\mathcal{T}$ as explained in Section 2.2.3. Still, it is to be noted that, despite the structure of the parameter matrix, $H_{\Lambda(\theta)}$ itself does not and cannot satisfy any Markov properties with respect to the tree $\mathcal{T}$: by Papastathopoulos and Strokorb (2016), max-stable distributions with continuous joint densities cannot possess any non-trivial conditional independence properties.

In the parametrization in (2.5) the extremal dependence in ξ and in X depends on a vector $\theta = (\theta_e, e \in E)$ of $d - 1$ free parameters, indexed by the edges of the tree. The main theme in this paper concerns inference on the parameter vector θ in case some of the variables ξ_v are latent (unobservable). The first question is whether all edge parameters θ_e are still identifiable from (2.4) when $\Lambda = \Lambda(\theta)$ and when $U \subsetneq V$ contains the indices of variables that can still be observed. For the Seine network in Figure 2.1, for instance, there are $d = 7$ variables in total, of which two are latent. A necessary and sufficient criterion for parameter identifiability is given in Proposition 2.3.1 below. Provided the criterion is fulfilled, the second question is how

to estimate the parameters. Three estimation methods are proposed in Section 2.4 and illustrated in Section 2.5.

Note that the random vector ξ itself does not necessarily belong to the max-domain of attraction of some max-stable distribution. The reason is that we do not impose that the marginal distributions of ξ are in the max-domain of attraction of some univariate extreme value distributions. To focus on the tail dependence of ξ, we standardize its margins and formulate the assumption in terms of X.

2.2.3 Motivation of the structured Hüsler–Reiss model

To motivate the structured Hüsler–Reiss parameter matrix $\Lambda(\theta)$ in (2.5), we construct a graphical model Z^* that satisfies the global Markov property with respect to the undirected tree $\mathcal{T} = (V, E)$ and such that $Z^* \in D(H_{\Lambda(\theta)})$. Besides serving as a motivation, the auxiliary model Z^* plays another important role: in view of Segers (2020b, Theorem 2) we are able to project certain asymptotic properties that hold for Z^* to X.

For disjoint subsets A, B, C of V, the expression $A \perp\!\!\!\perp_{\mathcal{T}} B \mid C$ means that C separates A from B in $\mathcal{T}$, also called graphical separation, i.e., all paths from A to B pass through at least one vertex in C. Let Z^* be defined on a probability space $(\Omega, \mathcal{B}, \mathbb{P})$. Conditional independence of Z^*_A and Z^*_B given Z^*_C will be denoted by $Z^*_A \perp\!\!\!\perp_{\mathbb{P}} Z^*_B \mid Z^*_C$; here $Z^*_A = (Z^*_a, a \in A)$ and so on. If $P = \mathbb{P}(Z^* \in \cdot\,)$ is the law of Z^*, we say that the tree $\mathcal{T}$ is an independence map (I-map) of P if for any disjoint subsets A, B, C of V it holds that

$$A \perp\!\!\!\perp_{\mathcal{T}} B \mid C \implies Z^*_A \perp\!\!\!\perp_{\mathbb{P}} Z^*_B \mid Z^*_C \tag{2.6}$$

(Koller and Friedman, 2009). This assumption is equivalent to the assumption that Z^* obeys the global Markov property with respect to $\mathcal{T}$ (Lauritzen, 1996).

The law of the random vector $Z^* = (Z^*_v, v \in V)$ is defined by the following two assumptions:

(Z1) Z^* satisfies the global Markov property (2.6) with respect to the undirected tree $\mathcal{T} = (V, E)$;

(Z2) every pair of variables (Z^*_a, Z^*_b) on adjacent nodes $(a, b) = e \in E$ has a bivariate Hüsler–Reiss distribution with parameter $\theta_e \in (0, \infty)$ and unit-Fréchet margins, i.e., the special case of (2.4) with $U = \{a, b\}$ and $\lambda^2_{ab} = \theta^2_e/4$.

The law of Z^* is absolutely continuous and its joint density function factorizes in terms of the bivariate Hüsler–Reiss densities along pairs of variables on adjacent nodes through the Hammersley–Clifford theorem; see Section 2.7.5 where we describe how to sample from Z^*. Moreover, for $e = (a, b) \in E$ and if Z has distribution $H_{\Lambda(\theta)}$, the law of (Z^*_a, Z^*_b) is the same as the one of (Z_a, Z_b). However, unless $d = 2$, the law of Z^* is itself not max-stable and thus not equal to the one of Z. One way to see this is to note that by Papastathopoulos and Strokorb (2016), the law of Z cannot satisfy the global Markov property with respect to $\mathcal{T}$.

Let $(M_e, e \in E)$ be a random vector of independent lognormal random variables with $\ln M_e \sim \mathcal{N}(-\theta^2_e/2, \theta^2_e)$ for each $e \in E$. In view Theorem 1 and Corollary 1 in Segers (2020b), we have the convergence in distribution

$$\begin{aligned}(Z^*_v/Z^*_u, v \in V \setminus u) \mid Z^*_u > x &\xrightarrow{d} (\Xi_{u,v}, v \in V \setminus u) \\ &= \big(\textstyle\prod_{e \in (u \rightsquigarrow v)} M_e,\ v \in V \setminus u\big), \qquad x \to \infty,\end{aligned} \tag{2.7}$$

for every $u \in V$. For every $u \in V$ the vector $(\Xi_{u,v}, v \in V \setminus u)$ is called a tail tree. The multiplicative structure in (2.7) goes back to the theory of extremes of Markov chains

due to Smith (1992), Perfekt (1994), Yun (1998) and Segers (2007). Note that a chain can be seen as a tree with a single branch.

The vector $(\ln \Xi_{u,v}, v \in V \setminus u)$ is a linear transformation of a Gaussian random vector and is therefore itself Gaussian. Its mean vector $\mu_{V,u}(\theta)$ and its covariance matrix $\Sigma_{V,u}(\theta)$ have elements

$$\{\mu_{V,u}(\theta)\}_v = -\frac{1}{2} \sum_{e \in (u \rightsquigarrow v)} \theta_e^2, \qquad v \in V \setminus u, \tag{2.8}$$

$$\{\Sigma_{V,u}(\theta)\}_{ij} = \sum_{e \in (u \rightsquigarrow i) \cap (u \rightsquigarrow j)} \theta_e^2, \qquad i, j \in V \setminus u. \tag{2.9}$$

Hence for every $u \in V$ and as $x \to \infty$, we have the convergence in distribution

$$\begin{aligned}(\ln Z_v^* - \ln Z_u^*, v \in V \setminus u) \mid Z_u^* > x \xrightarrow{d} (\ln \Xi_{u,v}, v \in V \setminus u) \\ \sim \mathcal{N}_{|V \setminus u|}\big(\mu_{V,u}(\theta), \Sigma_{V,u}(\theta)\big),\end{aligned} \tag{2.10}$$

where $\mathcal{N}_p$ is the p-variate normal distribution. By construction, $\Sigma_{V,u}(\theta)$ is a covariance matrix and hence positive semi-definite for any $\theta \in (0, \infty)^{d-1}$; it is actually positive definite since the vector $(\ln \Xi_{u,v}, v \in V \setminus u)$ is the result of an invertible linear transformation applied to the vector $(\ln M_e, e \in E)$ of independent and non-degenerate normal random variables. The matrix $\Sigma_{V,u}(\theta)$ is moreover the same as the matrix $\Gamma_{W,u}(\Lambda)$ in (6.4) with $W = V$ and $\Lambda = \Lambda(\theta)$ in (2.5):

$$\begin{aligned}\big\{\Sigma_{V,u}(\theta)\big\}_{ij} &= \sum_{e \in (u \rightsquigarrow i) \cap (u \rightsquigarrow j)} \theta_e^2 = \frac{1}{2}\left(\sum_{e \in (u \rightsquigarrow i)} \theta_e^2 + \sum_{e \in (u \rightsquigarrow j)} \theta_e^2 - \sum_{e \in (i \rightsquigarrow j)} \theta_e^2 \right) \\ &= 2(\lambda_{iu}^2 + \lambda_{ju}^2 - \lambda_{ij}^2) = \big\{\Gamma_{V,u}\big(\Lambda(\theta)\big)\big\}_{ij}, \qquad i, j \in V \setminus u. \end{aligned} \tag{2.11}$$

In the second equality it is needed to divide by two because the parameters on shared edges are added twice. In addition, the Hüsler–Reiss parameters λ_{uv}^2 are proportional to the means:

$$2\lambda_{uv}^2 = \frac{1}{2} \sum_{e \in (u \rightsquigarrow v)} \theta_e^2 = -\{\mu_{V,u}(\theta)\}_v, \qquad v \in V \setminus u. \tag{2.12}$$

Proposition 2.2.1. *Let $\mathcal{T} = (V, E)$ be a tree. If the law of $Z^* = (Z_v^*, v \in V)$ is given by (Z1)–(Z2) above, then $Z^* \in D(H_{\Lambda(\theta)})$ with $\Lambda(\theta)$ in* (2.5).

The proof is given in Section 2.7.3 and relies on the properties of Z^* mentioned above, in particular on (2.10). By constructing a graphical model with respect to $\mathcal{T}$ in the max-domain of attraction of $H_{\Lambda(\Theta)}$, we have argued that the latter is a sensible dependence model for extremes of graphical models on trees. Moreover, it follows that any random vector $X = (X_v, v \in V)$ with unit-Pareto margins and in the max-domain of attraction of $H_{\Lambda(\theta)}$ shares property (2.10) with Z^*.

Corollary 2.2.2. *Let $\mathcal{T} = (V, E)$ be a tree and let $X = (X_v, v \in V)$ have unit-Pareto margins and belong to $D(H_{\Lambda(\theta)})$ with $\Lambda(\theta)$ as in* (2.5) *for a vector $\theta = (\theta_e, e \in E)$ of positive scalars. Then for every $u \in V$, we have as $t \to \infty$*

$$(\ln X_v - \ln X_u, v \in V \setminus u) \mid X_u > t \xrightarrow{d} \mathcal{N}_{|V \setminus u|}\big(\mu_{V,u}(\theta), \Sigma_{V,u}(\theta)\big). \tag{2.13}$$

Proof. The max-domain of attraction condition $X \in D(H_{\Lambda(\theta)})$ is known to be equivalent to convergence of the measures $t\, \mathbb{P}(X/t \in \cdot\,)$ as $t \to \infty$ to the exponent measure of $H_{\Lambda(\theta)}$ (Resnick, 1987, Proposition 5.17). Such measure convergence is in turn

equivalent to convergence in distribution of $X/X_u \mid X_u > t$ as $t \to \infty$ for every $u \in V$ to a limit that can be written in terms of the said exponent measure (Segers, 2020b, Theorem 2). But for X replaced by Z^*, the limiting conditional distribution was found to be a certain multivariate lognormal distribution in (2.7). The equivalence between (2.7) and (2.10) with Z^* replaced by X is clear by the continuous mapping theorem. □

The random vector X in Corollary 2.2.2 does not need to be a graphical model with respect to $\mathcal{T}$. The convergence in (2.13) appears in Engelke et al. (2014, Theorem 2) for a general random vector with standardized margins and in the max-domain of attraction of a Hüsler–Reiss distribution. With Corollary 2.2.2 we arrive at the same result but through the properties of the auxiliary model Z^*. The convergence in (2.13) is used to build two estimators in the next section.

In Engelke and Hitz (2020), a notion of conditional independence different from the classical one is introduced in the context of multivariate Pareto distributions. When specialized to the Pareto distribution associated to a max-stable Hüsler–Reiss distribution, it yields certain restrictions on the Hüsler–Reiss parameter matrix Λ. In case the conditional independence relations are the ones induced by a tree through graphical separation, the structure of the parameter matrix is the same as the one in (2.5). We explain the connection in Section 2.7.1. Here we just emphasize that in Engelke and Hitz (2020), no graphical model in the classical sense of the term is constructed that belongs to the max-domain of attraction of $H_{\Lambda(\theta)}$. The way we arrive at the structure of $\Lambda(\theta)$ via the graphical model Z^* in (Z1)–(Z2) is thus entirely different from their approach.

Finally, quite another tree-induced structure of the Hüsler–Reiss parameter matrix is proposed in Lee and Joe (2017). We provide a comparison in Section 2.7.2.

2.3 Latent variables and parameter identifiability

A typical application of our model arises in relation to quantities measured on river networks that have a tree-like structure. It is natural to associate a node to an existing measurement station or to locations where two river channels meet (junction) or one channel splits (split) even if there is no measurement station there. Stations are supposed to generate data for the quantity of interest, so for any node associated to a station there is a corresponding variable. In practice, junctions/splits may lack measurements, and this means that there are nodes in the tree with latent variables. Nodes with latent variables are those labelled 2 and 5 in the Seine network in Figure 2.1.

A naive approach to the presence of latent variables would be to ignore them, that is, to remove the corresponding nodes and all edges incident to them. This will yield a disconnected graph, making it necessary to add edges in some arbitrary way so as to obtain a tree again. In Figure 2.2 for instance, if node 2 is suppressed, there are three possible ways to reconnect the remaining nodes and form a tree. Each implies a different structured Hüsler–Reiss parameter matrix and thus a different dependence model.

In this paper we do not modify the original tree but take the latent variables into account. Let $\mathcal{T} = (V, E)$ be an undirected tree and consider the Hüsler–Reiss distribution (2.3) with parameter matrix $\Lambda = \Lambda(\theta)$ in (2.5). When there are nodes with latent variables, the question is whether it is still possible to identify the $d - 1$ free edge parameters θ_e from the distribution of the subvector of observable variables only. Let $U \subseteq V$ denote the set of indices of the observable variables. On the one hand, Eq. (2.13) implies

$$(\ln X_v - \ln X_u)_{v \in U \setminus u} \mid X_u > t \xrightarrow{d} \mathcal{N}_{|U \setminus u|}\big(\mu_{U,u}(\theta), \Sigma_{U,u}(\theta)\big), \qquad t \to \infty, \tag{2.14}$$

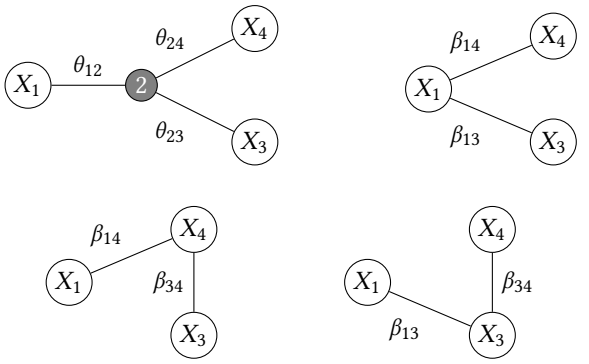


Figure 2.2: The first tree from the left has four nodes where node 2 has a latent variable. If node 2 is suppressed, there are three possible ways to reconnect the three remaining nodes into a tree again.

with $\mu_{U,u}(\theta)$ and $\Sigma_{U,u}(\theta)$ as in (2.8) and (2.9) but with V replaced by U. On the other hand, $\mu_{U,u}(\theta)$ and $\Sigma_{U,u}(\theta)$ together determine the stdf l_U of the subvector X_U in (2.4) through the identities (2.11) and (2.12). The question is thus whether the parameter vector θ is still identifiable from the $|U \setminus u|$-variate normal distributions on the right-hand side of (2.14), where u ranges over U.

Example. Let $X = (X_a, X_b, X_c)$ have unit-Pareto margins and suppose that $X \in D(H_{\Lambda(\theta)})$ where $\Lambda(\theta)$ is as in (2.5) with respect to the chain tree $\mathcal{T}$ with nodes $V = \{a, b, c\}$ and edges between a and b and between b and c. Since a parameter is linked to each (undirected) edge of the graph, the parameter vector is $\theta = (\theta_{ab}, \theta_{bc})$. Suppose the variable X_b is latent. By (2.14) we have

$$\ln X_c - \ln X_a \mid X_a > t \xrightarrow{d} \mathcal{N}\big(-(\theta_{ab}^2 + \theta_{bc}^2)/2, (\theta_{ab}^2 + \theta_{bc}^2)\big), \qquad t \to \infty.$$

It is clear that from the limiting normal distribution, we cannot identify θ_{ab} and θ_{bc}.

In this section it is shown that as long as all nodes with missing variables have degree at least three, the parameters associated to the Hüsler–Reiss distribution of the full vector are still identifiable and hence there is no need to change the tree. To this end, note that by (2.8), (2.11) and (2.12) with V replaced by $U \subseteq V$ such that $u \in U$, the mean vectors $\mu_{U,u}(\theta)$ and covariance matrices $\Sigma_{U,u}(\theta)$ are determined completely by the path sums

$$p_{ab} = \sum_{e \in (a \rightsquigarrow b)} \theta_e^2 = 4\lambda_{ab}^2, \qquad a, b \in U, \tag{2.15}$$

and that, vice versa, the values of these path sums are determined by the vectors $\mu_{U,u}(\theta)$ and the matrices $\Sigma_{U,u}(\theta)$. If we know the distribution of $X_U = (X_u, u \in U)$, we can compute the values of these sums, and if we know these sums, we can compute the stdf l_U of X_U. The question is thus whether or not the edge parameters θ_e are identifiable from the values of the path sums p_{ab} for $a, b \in U$. According to the following proposition, there is a surprisingly simple criterion to decide whether this is the case or not.

Proposition 2.3.1. *Let $\mathcal{T} = (V, E)$ be an undirected tree and let $X = (X_v, v \in V)$ have unit Pareto margins and be in the max-domain of attraction of the structured Hüsler–Reiss distribution H_Λ in* (2.3) *with parameter matrix $\Lambda = \Lambda(\theta)$ in* (2.5)*. Let $U \subseteq V$ be the set of nodes corresponding to the observable variables. The parameter vector θ is identifiable from $X_U = (X_u, u \in U)$ if and only if every node $u \in V \setminus U$ has degree at least three.*

Proof. Necessity. Assume that the elements of the edge parameter $\theta \in (0,\infty)^{d-1}$ are uniquely identifiable. Let $\bar{U} = V \setminus U \neq \emptyset$ be the set of nodes with latent variables. We need to show that every $v \in \bar{U}$ has degree $d(v)$ at least 3. We will do this by contraposition. As a tree is connected by definition, there cannot be a node of degree zero.

First, assume there is $v \in \bar{U}$ such that $d(v) = 1$. The node v must be a leaf node, and in this case there is no path $(a \rightsquigarrow b)$ with $a, b \in U$ that passes by v, and thus θ_{uv}^2, with u the unique neighbor of v, does never appear in the sum (2.15). Hence θ_{uv} is not identifiable, which is a contradiction to the assumption.

Second, assume there exists $v \in \bar{U}$ with $d(v) = 2$. Then v has exactly two neighbors, i and j, say. Every path sum p_{ab} for $a, b \in U$ will contain either the sum of the squared parameters, $\theta_{iv}^2 + \theta_{jv}^2$, or neither of these. Hence, the individual edge parameters θ_{iv} and θ_{jv} are not identifiable, yielding a contradiction. (This generalizes the example given before the statement of the proposition.)

Sufficiency. Assume that all nodes with latent variables are of degree three or more. Let $e = (u, v) \in E$. We will find a linear combination of the path sums (2.15) equal to θ_{uv}^2.

If $u, v \in U$, then the one-edge path sum $p_{uv} = \theta_{uv}^2$ already meets the condition.

Suppose that $u \in \bar{U}$. By assumption, u has at least two other neighbors besides v, say w and x. If $v \in U$, then put $\hat{v} = v$. Otherwise, start walking at v away from u until you encounter the first visible node, say $\hat{v} \in U$. There must always be such a node, since V is finite and since all leaves are observable by assumption. Similarly, let $\hat{w} \in U$ and $\hat{x} \in U$ be the first visible nodes encountered when walking away from u and starting in w and x, respectively. Note that $\hat{v}$, $\hat{w}$, and $\hat{x}$ are all different since otherwise the graph would contain a non-trivial cycle, which is not possible in the case of a tree. We can thus observe the sums

$$\begin{aligned} p_{\hat{v}\hat{w}} &= p_{\hat{v}u} + p_{u\hat{w}}\,, \\ p_{\hat{v}\hat{x}} &= p_{\hat{v}u} + p_{u\hat{x}}\,, \\ p_{\hat{w}\hat{x}} &= p_{\hat{w}u} + p_{u\hat{x}}\,. \end{aligned}$$

Since $p_{yz} = p_{zy}$ for every $y, z \in V$, the previous identities constitute three linear equations in three unknowns that can be solved explicitly, producing the values of $p_{u\hat{v}}$, $p_{u\hat{w}}$, $p_{u\hat{x}}$. In particular, summing the first two equations, subtracting the third, and dividing by two, we find

$$p_{u\hat{v}} = \tfrac{1}{2}p_{\hat{v}\hat{w}} + \tfrac{1}{2}p_{\hat{v}\hat{x}} - \tfrac{1}{2}p_{\hat{w}\hat{x}}\,.$$

If $v \in U$, then $v = \hat{v}$, and $(u \rightsquigarrow v) = \{e\}$, so that the above equation shows how to combine path sums in a linear way to extract $p_{uv} = \theta_e^2$.

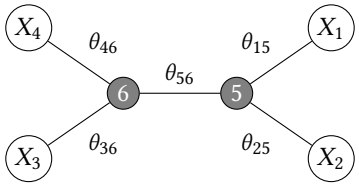
If $v \notin U$, then we can repeat the same procedure with u replaced by v. The result is a formula expressing $p_{v\hat{v}}$ as a linear combination of three visible path sums. Now since

$$\theta_e^2 = p_{u\hat{v}} - p_{v\hat{v}}\,,$$

we have found a way to extract θ_e^2 by a linear combination of at most six visible path sums. □

The proof of Proposition 2.3.1 consists in solving the equations (2.15) with p_{ab} as known and θ_e^2 as unknown. Clearly, this is a linear system of equations and the question is thus whether the coefficient matrix defining the system has full column rank. It is an open question how to write down this matrix, which contains only zeroes and ones, in terms of the tree's adjacency matrix in such a way that an algebraic criterion on the latter matrix can be formulated.

The identifiability criterion in Proposition 2.3.1 allows nodes with latent variables to be adjacent and still counting in the computation of each other's degree. Consider for instance the following tree:



The variables at the adjacent nodes 5 and 6 are latent. Both nodes have degree three and each of the five edge parameters θ_e can be solved from the path sums p_{ab} between nodes $a, b \in \{1, \ldots, 4\}$.

The previous example may give the impression that for the identifiability criterion to hold it is actually enough that all variables on leaf nodes are observable. Although the latter property is indeed necessary, it is not sufficient, as illustrated by the example before Proposition 2.3.1.

2.4 Estimation

Let $\mathcal{T} = (V, E)$ be an undirected tree with nodes $V = \{1, \ldots, d\}$ and let $(\xi_{v,i}, v \in V, i = 1, \ldots, n)$ be an independent random sample from the distribution of ξ satisfying the assumptions in Section 2.2.2. Further, let $U \subseteq V$ be the set of indices of observable variables and assume that every $u \in V \setminus U$ has degree at least three, so that, by Proposition 2.3.1, the Hüsler–Reiss edge parameters $\theta = (\theta_e, e \in E)$ in the definition of $\Lambda(\theta)$ in (2.5) are identifiable from the distribution of the subvector $\xi_U = (\xi_v, v \in U)$.

We propose three methods for estimating the parameter vector θ. The first one, called moment estimator (Section 2.4.1), builds upon the one introduced in Engelke et al. (2014). The second estimator comes from the optimization of a composite likelihood function (Section 2.4.2). The third estimator, finally, is based on bivariate extremal coefficients (Section 2.4.3) and on the method in Einmahl et al. (2018). All estimators are functions of the subvectors $(\xi_{U,i}) = (\xi_{v,i}, v \in U)$ for $i = 1, \ldots, n$ only.

An important remark for this whole section is related to the fact that X as introduced in Section 2.2.2 should have unit Pareto margins, obtained after the transformation $X_v = 1/(1 - F_v(\xi_v))$ where F_v is the marginal distribution function of ξ_v for $v \in V$. It is unrealistic to assume that the functions F_v are known, so in practice we use their empirical versions, $\hat{F}_{v,n}(x) = \left[\sum_{i=1}^{n} \mathbb{1}(\xi_{v,i} \leq x) \right] / (n+1)$. The estimates of the edge parameters will then be based upon the sample $\hat{X}_1, \ldots, \hat{X}_n$ with coordinates

$$\hat{X}_{v,i} = \frac{1}{1 - \hat{F}_{v,n}(\xi_{v,i})}, \qquad v \in U, \quad i = 1, \ldots, n,$$

considered as a random sample from the distribution of $X_U = (X_u, u \in U)$.

A variable indexed by the double subscript W, i will denote the i-th observation of variables on nodes belonging to the set $W \subseteq U$: for instance $\hat{X}_{W,i} = (\hat{X}_{v,i}, v \in W)$. Such vectors are taken to be column vectors of length $|W|$. When $W = U$ we just write $\hat{X}_i$.

2.4.1 Method of moments estimator

Engelke et al. (2014) introduce an estimator of the matrix Λ of the Hüsler–Reiss distribution, based on sample counterparts of the matrices $\Gamma_{W,u}(\Lambda)$ in (6.4). Relying on (2.11) with V replaced by $W \subseteq U$, we will apply their method to the vector of

observable variables and then add a least-squares step to extract the edge parameters θ_e.

As a starting point we take the result in (2.14) and as suggested by Engelke et al. (2014) for given $k \in \{1, \dots n\}$ we obtain the log-differences

$$\Delta_{uv,i} = \ln \hat{X}_{v,i} - \ln \hat{X}_{u,i}\,, \tag{2.16}$$

for $u, v \in U$ and for $i \in I_u = \{i = 1, \dots, n : \hat{X}_{u,i} > n/k\}$. The proposed estimators of $\mu_{U,u}$ and $\Sigma_{U,u}$ are respectively the sample mean vector

$$\hat{\mu}_{U,u} = \frac{1}{|I_u|} \sum_{i \in I_u} (\Delta_{uv,i}, v \in U \setminus u)$$

and the sample covariance matrix

$$\hat{\Sigma}_{U,u} = \frac{1}{|I_u|} \sum_{i \in I_u} (\Delta_{uv,i} - \hat{\mu}_{U,u}, v \in U \setminus u)(\Delta_{uv,i} - \hat{\mu}_{U,u}, v \in U \setminus u)^\top\,.$$

To estimate the vector of edge parameters $\theta = (\theta_e, e \in E)$, we propose the least squares estimator

$$\hat{\theta}^{\mathrm{MM}}_{n,k} = \operatorname*{arg\,min}_{\theta \in (0,\infty)^E} \sum_{u \in U} \|\hat{\Sigma}_{U,u} - \Sigma_{U,u}(\theta)\|_F^2\,. \tag{2.17}$$

where $\|\cdot\|_F$ is the Frobenius norm. In this way, we take advantage of the empirical covariance matrices $\hat{\Sigma}_{U,u}$ for each $u \in U$ and thus of each exceedance set I_u.

In (2.17), for each $u \in U$, we consider the covariance matrix of the log-differences $\Delta_{uv,i}$ for all $v \in U \setminus u$. However, if v is far away from u in the tree, then the extremal dependence between ξ_u and ξ_v may be weak and the difference $\Delta_{uv,i}$ may carry little information. Therefore, we propose a modified estimator where, for each $u \in U$, we limit the scope to a subset $W_u \subseteq U$ of observable variables indexed by nodes near u, producing the estimator

$$\hat{\theta}^{\mathrm{MM}}_{n,k} = \operatorname*{arg\,min}_{\theta \in (0,\infty)^E} \sum_{u \in U} \|\hat{\Sigma}_{W_u,u} - \Sigma_{W_u,u}(\theta)\|_F^2\,. \tag{2.18}$$

Besides being simpler to compute, the modified estimator (2.18) performed better than the one in (2.17) in Monte Carlo experiments. One possible explanation is that by excluding pairs with weak extremal dependence, the bias of the estimator diminishes.

When choosing the sets W_u, care needs to be taken that the parameter vector θ is still identifiable from the collection of covariance matrices $\Sigma_{W_u,u}(\theta)$ for $u \in U$. The set of path sums p_{ab} for $a, b \in U$ in Proposition 2.3.1 is now reduced to the set of the path sums p_{ab} for $a, b \in W_u$ and $u \in U$. Whether or not these are still sufficient to identify θ needs to be checked on a case-by-case basis. This issue is illustrated in Section 2.7.4.

2.4.2 Composite likelihood estimator

The composite likelihood estimator (CLE) is again based on the result in (2.14). This time however we maximize a composite likelihood function with respect to the parameter θ directly. The composite likelihood function consists of multiplication of likelihoods which are defined on subtrees.

As for the method of moments estimator in Section 2.4.1, we consider for each $u \in U$ a set $W_u \subseteq U$ of nodes that are close to u in the tree, taking care to include sufficiently many variables so that the edge parameters are still identifiable (Section 2.7.4). Recall

the log-differences $\Delta_{uv,i}$ in (3.19) and the exceedance set I_u right below (3.19). Let $\phi_p(\,\cdot\,;\Sigma)$ be the density function of the centered p-variate normal distribution with covariance matrix Σ. The composite likelihood estimator $\hat{\theta}^{\mathrm{CLE}}_{n,k}$ is the maximizer of the composite likelihood

$$L\big(\theta;\,\{\Delta_{uv,i}:v\in W_u\setminus u,\ i\in I_u, u\in U\}\big) = \prod_{u\in U}\prod_{i\in I_u}\phi_{|W_u\setminus u|}\big((\Delta_{uv,i})_{v\in W_u}-\mu_{W_u,u}(\theta);\Sigma_{W_u,u}(\theta)\big).$$

We aggregate the likelihoods of the different normal distributions for all $u\in U$ treating the samples of log-differences as independent, although they are not. Results from Monte Carlo simulation experiments (Section 2.7.5) show that the performance of the CLE is comparable to the one of the moment estimator and the extremal coefficient estimator.

Other estimation methods based on locally defined likelihoods are used by Engelke and Hitz (2020) and Lee and Joe (2017). The method of Engelke and Hitz (2020) estimates the parameters associated to each clique separately. For trees this means that there are $d-1$ one-variate likelihood functions to optimize, a problem which is doable even in trees with many nodes. A problem with this estimator is that it is inapplicable if there are latent variables because there will always be an adjacent pair of variables with one of them being an unobservable, and making it impossible to estimate the corresponding edge parameter. The estimator of Lee and Joe (2017) is based on pairwise likelihoods, which can be any pairs, not only adjacent pairs as in the estimator of Engelke and Hitz (2020). It is obtained by optimizing the composite likelihood which consists of multiplying the pairwise likelihoods. This estimator is applicable when there are latent variables as long as all possible pairs between the observed variables are included in the composite likelihood function. It is close in spirit to the pairwise extremal coefficients estimator considered next.

2.4.3 Pairwise extremal coefficients estimator

The pairwise extremal coefficients estimator (ECE), defined for general tail dependence models in Einmahl et al. (2018), is based on the bivariate stable tail dependence function (stdf) in (2.4). It minimizes the weighted distance between a non-parametric estimate and the fitted parametric stdf.

Let l be the stdf in (2.1) and recall that the extremal coefficient associated to a node set $J\subseteq V$ is defined as

$$l(x_J) = l_J(1,\dots,1) = \lim_{t\to\infty} t\ \mathbb{P}\left(\max_{j\in J} X_j > t\right), \tag{2.19}$$

where $x_J=(\mathbb{1}_{\{j\in J\}}, j\in V)$ and where l_J is the stdf of the subvector X_J. For the Hüsler–Reiss distribution with parameter matrix Λ and for a pair of nodes $J=\{u,v\}$, the bivariate extremal coefficient is just $l_J(1,1)=2\Phi(\lambda_{uv})$, with Φ the standard normal cdf. In case $\Lambda=\Lambda(\theta)$ in (2.5), the pairwise extremal coefficient depends on the path sum $p_{uv}=\sum_{e\in(u\rightsquigarrow v)}\theta_e^2$ via

$$l_J(1,1;\theta) = 2\Phi(\sqrt{p_{uv}}/2), \qquad J=\{u,v\}. \tag{2.20}$$

The non-parametric estimator of the stdf dates back to Drees and Huang (1998) and yields the following estimator for the extremal coefficient $l_J(1,\dots,1)$ for $J\subseteq V$:

$$\hat{l}_{J;n,k}(1,\dots,1) = \frac{1}{k}\sum_{i=1}^{n}\mathbb{1}\left(\max_{j\in J} n\hat{F}_{j,n}(\xi_{j,i}) > n+1/2-k\right). \tag{2.21}$$

Let $\mathcal{Q} \subseteq \{J \subseteq U : |J| = 2\}$ be a collection of pairs of nodes associated to observable variables and put $q = |\mathcal{Q}|$, ensuring that $q \geq |E| = d - 1$, the number of free edge parameters. The pairwise extremal coefficients estimator (ECE) of θ is

$$\hat{\theta}^{\text{ECE}}_{n,k} = \underset{\theta \in (0,\infty)^E}{\arg\min} \sum_{J \in \mathcal{Q}} \left(\hat{l}_{J;n,k}(1,1) - l_J(1,1;\theta) \right)^2 . \tag{2.22}$$

If $\mathcal{Q}$ is the collection of all possible pairs of nodes in U, then the pairwise extremal coefficients (2.20) give us access to all path sums p_{ab} for $a, b \in U$, and Proposition 2.3.1 guarantees we can identify θ. If, however, $\mathcal{Q}$ is a smaller set of pairs, then the identifiability of θ from the resulting path sums needs to be checked on the case at hand.

2.5 High water levels on the Seine network

We have chosen to present an application that allows us to demonstrate the identifiability criterion outlined in Section 2.3. Data were collected from http://www.hydro.eaufrance.fr, a web-site of the french Ministry of Ecology, Energy and Sustainable Development, and span the period from January 1987 to April 2019 with gaps for some of the measurement stations. The data represent water levels, in cm, at five locations on the Seine river: Paris, Meaux, Melun, Nemours and Sens. The map on Figure 2.1 shows part of the actual Seine network. The schematic representation of the graphical model used in the estimation is shown in Figure 2.3. The tree has $d = 7$ nodes, two of which are associated to latent variables. Since both these nodes have degree equal to three, Proposition 2.3.1 guarantees we can still identify all six edge parameters $\theta_1, \ldots, \theta_6$. For more information on the data set, some summary statistics and details on data preprocessing, we refer to Section 2.7.6.

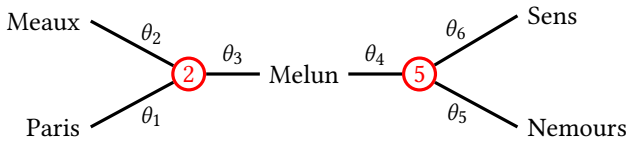


Figure 2.3: The Seine network with the tail dependence parameters associated to each edge of the tree.

2.5.1 Estimates and confidence intervals

We used all three estimators in Section 2.4 to obtain estimates of the six parameters of extremal dependence. For the pairwise extremal coefficient estimator (ECE) it is possible to calculate standard errors thanks to the asymptotic distribution derived in Einmahl et al. (2018, Theorem 2.2). Computational details for the standard errors follow in Section 2.7.7. The distributions of the MME and CLE are not known so we computed bootstrapped confidence intervals, known as basic bootstrap confidence intervals (Davison and Hinkley, 1997, Chapter 2), by resampling from the data.

The EC estimates and their 95% confidence intervals are displayed in Figure 2.4 for two of the parameters, namely θ_1 and θ_4. The confidence intervals using the MME and CLE are narrower as can be seen from Figure 2.5. The plots for $\theta_2, \theta_5, \theta_6$ are similar to the one for θ_1: the 95% confidence intervals never include zero, suggesting that the extremal dependence between the corresponding variables is not perfect and hence that the edges cannot be collapsed. In Section 2.3, we alluded to the possibility of circumventing the issue of latent variables by suppressing nodes and redrawing

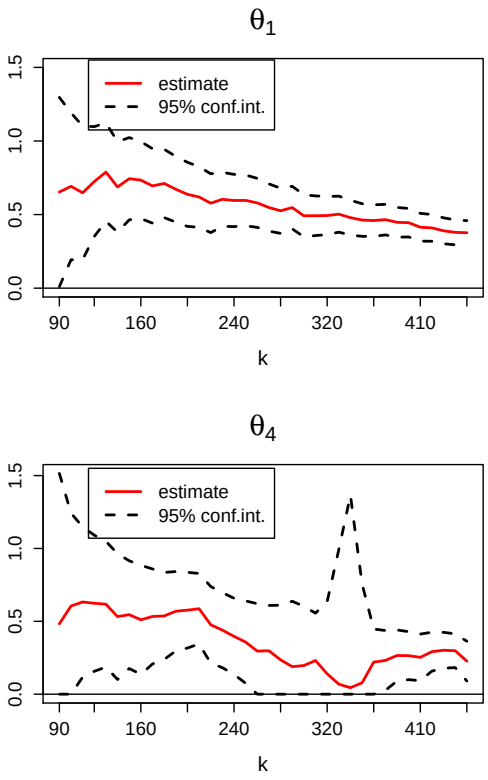


Figure 2.4: Point estimates and confidence intervals for the pairwise ECE.

edges. The fact that the confidence intervals do not include zero indicate that doing so would have produced a misleading picture of extremal dependence.

The plot of θ_3, similarly to the plot of θ_4, does contain a segment over k where the lower confidence bound reaches zero: for θ_4 this is approximately $k \in [260, 360]$, while for θ_3 it is $k \in [90, 180]$. Although the confidence intervals for θ_3 and θ_4 indicate some instability of the estimated parameters, we believe that collapsing the edges is not advisable, especially in networks with many more unobservable variables. Moreover, the river distance, which is one of the important factors in tail dependence (Asadi et al., 2015), is rather long between node 2 and Melun and between Melun and node 5, so that there is no physical motivation for collapsing the corresponding edges.

For a point estimate per parameter we need to average out over a range of k. The chosen range per estimator and per parameter need not be the same. As a rule we select a range around the beginning where the estimates start stabilizing around a certain level, omitting the most volatile part for relatively small k. Most of the time we thus consider $k \in [100, 220]$. In this way we end up with the point estimates displayed for comparison in Figure 2.5.

Given the similarities between the MME and CLE, the estimates are pooled in an average of the two for each parameter.

2.5.2 Considerations on the goodness-of-fit of the model

The Hüsler–Reiss family would not be an appropriate extremal dependence model if some of the variables would exhibit asymptotic independence. As kindly suggested by a Reviewer, we compared non-parametric estimates of multivariate tail dependence

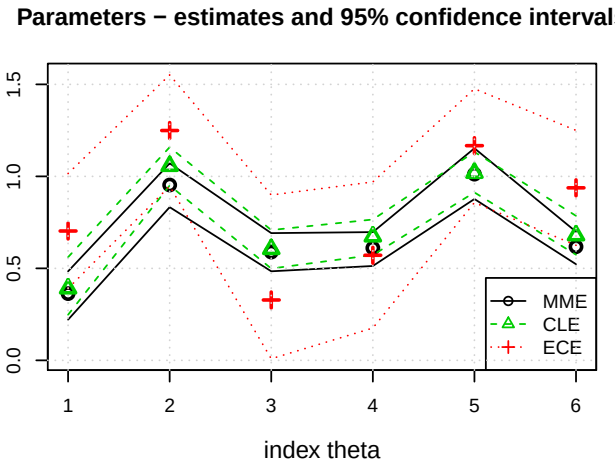


Figure 2.5: Parameters – estimates and confidence intervals. The confidence intervals of the moment and composite likelihood estimators are bootstrapped, namely $\theta \in [2\hat{\theta} - q^*_{0.975}, 2\hat{\theta} - q^*_{0.025}]$, where q^*_α is the α-quantile of the bootstrapped distribution of $\hat{\theta}$.

coefficients $\mathbb{P}(\min_{v\in W} X_v > t \mid X_u > t) = t\,\mathbb{P}(\min_{v\in W\cup u} X_v > t)$ for $W \subsetneq U$ and $u \in U \setminus W$ at finite thresholds $t = n/k$ with their postulated limits as $t \to \infty$ based on the fitted Hüsler–Reiss stdf. The results (not shown) supported the hypothesis of asymptotic dependence for nearly all subvectors $X_{W\cup u}$ of variables.

To assess how well the model from Section 2.2.2 fits the data, we compare non-parametric and model-based estimates of quantities describing extremal dependence, such as pairwise and triple-wise extremal coefficients and the Pickands dependence function. For $J \subseteq U$, recall the extremal coefficient $l_J(1,\dots,1)$ in (2.19) and its non-parametric estimate $\hat{l}_{J;n,k}(1,\dots,1)$ in (6.3.3), also called empirical extremal coefficient. The extremal coefficient $l_J(1,\dots,1)$ is always between 1 and $|J|$, corresponding to perfect extremal dependence and to extremal independence, respectively.

Figure 2.6 compares the model-based extremal coefficients obtained from (2.4) by plugging in parameter estimates with the empirical counterparts for pairs and triples $J \subseteq U$. At least visually the fit is quite good for both estimators considered, which are the average of the CLE and MME on the one hand and the ECE on the other hand. Note that the ECE in (6.3.3) is constructed explicitly to ensure that the model-based pairwise extremal coefficients fit the empirical ones as closely as possible. It is therefore only natural that the extremal coefficients based on the ECE fit the empirical ones best. A more comprehensive comparison of the finite-sample performance of MME, CLE and ECE is reported in a simulation study in Section 2.7.5.

It should be noted that it is impossible to compute empirical extremal coefficients involving latent variables. Model-based estimates of such extremal coefficients can still be computed however, thanks to the identifiability of the parameter vector θ.

As another visual check of the goodness-of-fit of the assumed model we consider the bivariate Pickands dependence function, usually denoted by $A(w)$ for $w \in [0,1]$. For the Hüsler–Reiss extreme-value distribution at the pair $J = \{u, v\}$, it is equal to

$$\begin{aligned} A_{u,v}(w;\theta) &= l_J(1-w, w; \theta) \\ &= (1-w)\,\Phi\left(\frac{\ln(\frac{1-w}{w}) + \frac{1}{2}p_{uv}}{\sqrt{p_{uv}}}\right) + w\,\Phi\left(\frac{\ln(\frac{w}{1-w}) + \frac{1}{2}p_{uv}}{\sqrt{p_{uv}}}\right), \end{aligned}$$

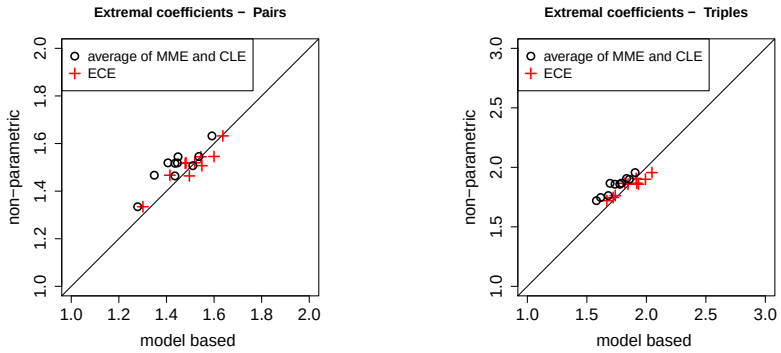


Figure 2.6: Non-parametric vs model-based extremal coefficients for pairs (left) and triples (right).

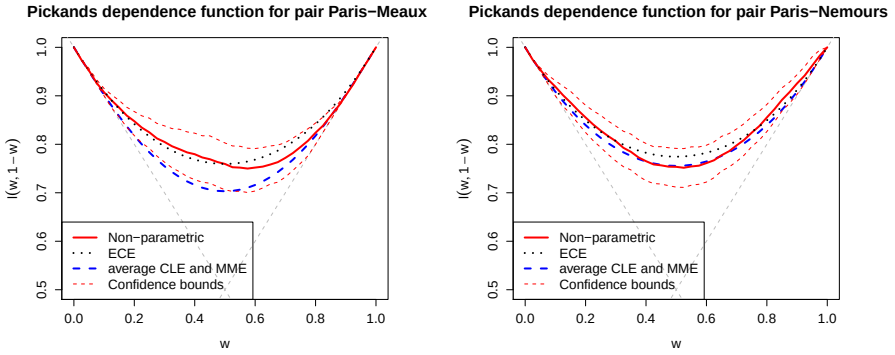


Figure 2.7: The empirical and model-based Pickands dependence function computed using the pooled CL and MM estimates and the EC estimates. The dashed gray lines show the lower limit $\max(w, 1-w)$ **of any Pickands dependence function** $A(w)$**.**

with p_{uv} as in (2.15). Hence the model-based estimator of the function $A_{u,v}(w;\theta)$ is $A_{u,v}(w;\hat{\theta}_{n,k})$ where $\hat{\theta}_{n,k}$ can be the average MME/CLE or the ECE.

The non-parametric counterpart of the Pickands dependence function is

$$\hat{A}_{u,v}(w) = \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}\left\{n\hat{F}_{u,n}(\xi_{u,i}) > n - k(1-w) + \frac{1}{2} \text{ or } n\hat{F}_{v,n}(\xi_{v,i}) > n - kw + \frac{1}{2}\right\}.$$

The model-based Pickands dependence function is compared to the empirical counterpart in Figure 2.7. The plot is complemented with non-parametric 95% confidence intervals for $A(w)$ computed by the bootstrap method introduced in Kiriliouk et al. (2018, Section 5). The general idea of the method is to approximate the distribution of $\sqrt{k}(\hat{l}_{n,k} - l)$ by the distribution of $\sqrt{k}(\hat{l}^*_{n,k} - \hat{l}^{\beta}_{n,k})$ where $\hat{l}^*_{n,k}$ is the empirical stable tail dependence function based on the ranks of a sample of size n from the empirical beta copula and $\hat{l}^{\beta}_{n,k}$ is the stdf based on the empirical beta copula using the ranks of the original sample $(\xi_{v,i}, v \in U)$ for $i = 1, \ldots, n$. A detailed description of the derived bootstrap confidence intervals is provided in Section 2.7.8.

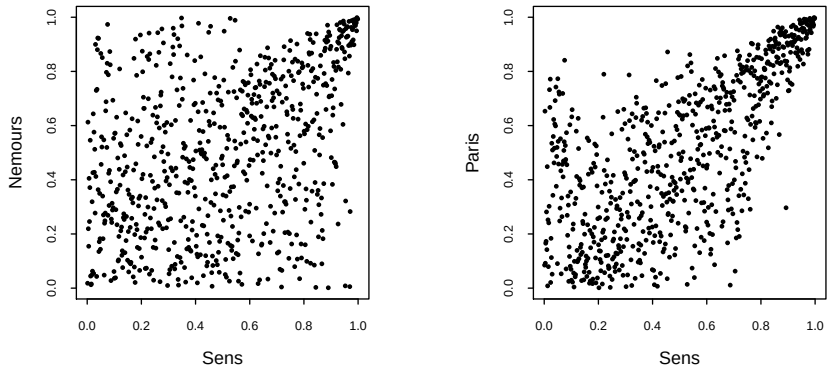


Figure 2.8: Scatterplots of uniform transformed data, $\hat{F}_{v,n}(\xi_{v,i}), v \in U, i = 1, \ldots, n$, for two pairs of locations. Sens and Nemours (left) are not flow connected while Sens and Paris (right) are flow connected. It can be seen from the Seine map in Figure 2.1 that the river and Euclidean distance from Sens to Nemours is much smaller than the one from Sens to Paris. However the tail dependence seems to be stronger for the second pair of locations.

2.5.3 Flow-connectedness and tail dependence

In a study by Asadi et al. (2015) of data from the Danube, it was found that a key factor for extremal dependence between two locations is whether or not they are flow connected. Two locations are flow connected if one of them is downstream of the other one. Flow connectedness often dominates river distance or Euclidean distance in importance: variables on distant nodes that are flow connected might have stronger tail dependence than variables on nodes that are nearby but not flow connected.

This effect is confirmed in our data too and is illustrated in Figure 2.8. The cities of Sens and Nemours are not flow connected but the Euclidean and river distance between them is smaller than the one between the flow connected cities of Sens and Paris. Still, the tail dependence seems to be stronger for the flow connected pair of locations.

Figure 2.9 illustrates the tail dependence in the Seine network through a heat map of the pairwise extremal coefficients. Pairs which are flow connected are indeed the ones with stronger tail dependence (smaller extremal coefficient). According to both estimators the strongest tail dependence is to be found between Paris and the locations at node 2, node 5 and Melun.

2.5.4 Suppressing latent variables

In Section 2.3 we alluded to the possibility of suppressing nodes with latent variables. Here we illustrate that method and compare the results with those presented so far. After removing nodes 2 and 5 from the Seine graph in Figure 2.3, there is no unique way of reconnecting the remaining five nodes into a tree. Two possible structures for the reduced Seine graph are presented in Figure 2.10. We opt for the right-hand graph and refer to the model associated to that tree as model B. Model A will refer to the one associated to the original graph in Figure 2.3.

In Figure 2.11, we compare the extremal coefficients induced by model B to the empirical ones and to those induced by model A. The extremal coefficients resulting from both models turn out to be rather close, the little black circles lying almost on

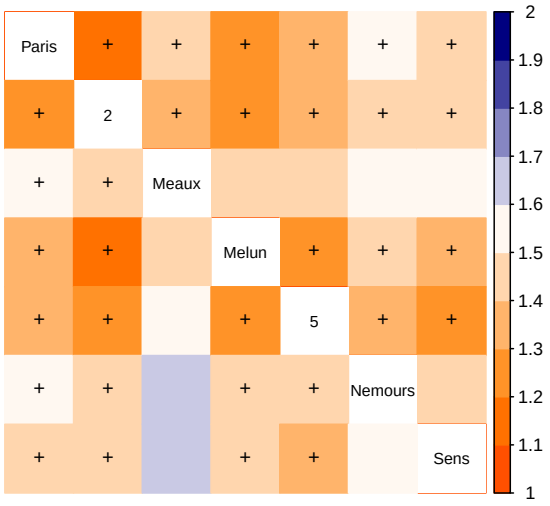


Figure 2.9: Heat map of the extremal coefficients. The upper diagonal is computed using the pooled MM and CL estimates and the lower diagonal uses the EC estimates. The crosses denote flow connected nodes.

the diagonal in all four plots. The reason may be that the original tree is small and that not many nodes have been suppressed, whereas in case of many latent variables, the impact of suppressing them may be large. Furthermore, the results may depend on the particular choice of the reduced tree: out of many possibilities two of which shown in Figure 2.10 we selected the second one. A final shortcoming of the method of suppressing nodes is that tail dependence cannot be calculated for random vectors involving latent variables.

We conclude in Figure 2.12 with a depiction of pairwise upper tail dependence in the Seine network. Shown are the complete and reduced trees with edges weighted by the tail dependence coefficients, defined for a pair $J = \{u, v\} \subset V$ by

$$2\big(1 - \ell_J(1, 1; \theta)\big) \tag{2.23}$$

in terms of the pairwise extremal coefficient in (2.20). In both trees, the strongest tail dependence occurs along the path from Sens to Paris.

2.6 Conclusion

We have presented a statistical model suitable for studying extremal dependence within a vector of random variables indexed by the nodes of a tree. The edges between the nodes are meant to indicate links between variables arising from a physical or conceptual network, although we do not impose any conditional independence relations. The main assumption is that, upon marginal standardization, the data-generating distribution is in the max-domain of attraction of a max-stable Hüsler–Reiss distribution whose parameter matrix possesses a certain structure induced by the tree: a free parameter is associated to each edge and each element of the Hüsler–Reiss parameter matrix only depends on the sum of the edge parameters along the path between the two corresponding nodes on the tree. We showed that the max-domain of attraction of this tree-structured Hüsler–Reiss distribution contains a specific distribution that, unlike the max-stable Hüsler–Reiss distribution or the associated multivariate Pareto distribution, satisfies the global Markov property with

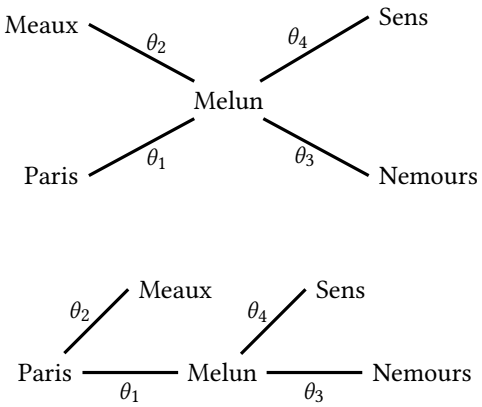


Figure 2.10: Two different versions of the graph of the Seine network in Figure 2.3 if nodes with latent variables are suppressed and new edges are drawn between the remaining nodes of the affected parts of the tree.

respect to the tree. This auxiliary model not only motivates the postulated structure, it also allowed us to find extremal dependence properties of any distribution satisfying our main assumption.

The central point and contribution of the paper is related to the identifiability of the edge parameters in case some of the variables are latent (unobservable). This situation occurs for instance in applications on river networks, when measurements on certain locations are missing. We showed that the edge parameters are uniquely identified by the distribution of the observable variables if and only if all nodes indexing latent variables are of degree at least three. Thanks to this result it is possible to quantify tail dependence even between latent variables. The characterization is due to the special structure of the variogram matrix of the Hüsler–Reiss distribution and may not be applicable to other max-stable distributions.

We fitted the model to water level data on the Seine network on a tree with seven variables, two of which were latent. As the corresponding nodes both had degree three, the six edge parameters were still identifiable and could be estimated based on data from the five observable variables. Three different estimators were proposed and implemented, based on the method of moments, on composite likelihood, and on pairwise extremal coefficients. Comparisons of non-parametric and model-based tail dependence quantities confirmed the adequacy of the fitted structured Hüsler–Reiss distribution.

For comparison we estimated a model where the two nodes with latent variables were suppressed and the edges between the affected parts of the network were redrawn in an arbitrary way. Although for the Seine data this reduction did not have a big impact on the fitted tail dependence model of the observable variables, we argued why it is still recommendable to take latent variables into account, provided there is now a sound way to do so.

An open question concerns parameter identifiability criteria in case of latent variables for Hüsler–Reiss distributions with parameter matrices structured in different ways than in this paper. Even the structure itself may be partially unknown. Another interesting direction for further research concerns extensions from the Hüsler–Reiss family to other parametric families of max-stable distributions.

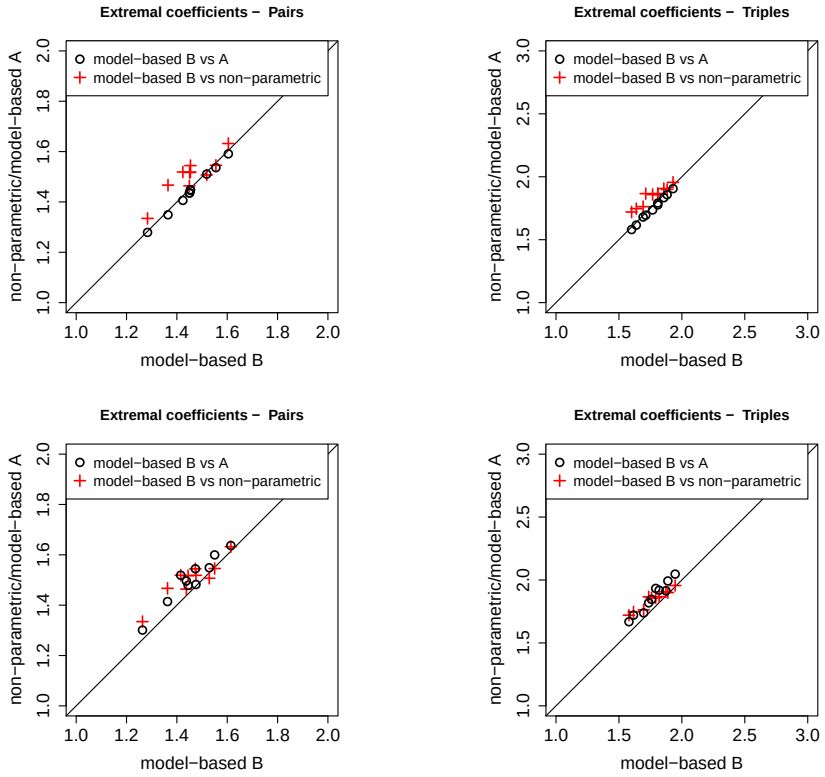


Figure 2.11: Comparison of extremal coefficients under model A (latent variables included) and model B (latent variables excluded). Top: combined MM and CL estimates; bottom: EC estimates. Left: pairs; right: triples.

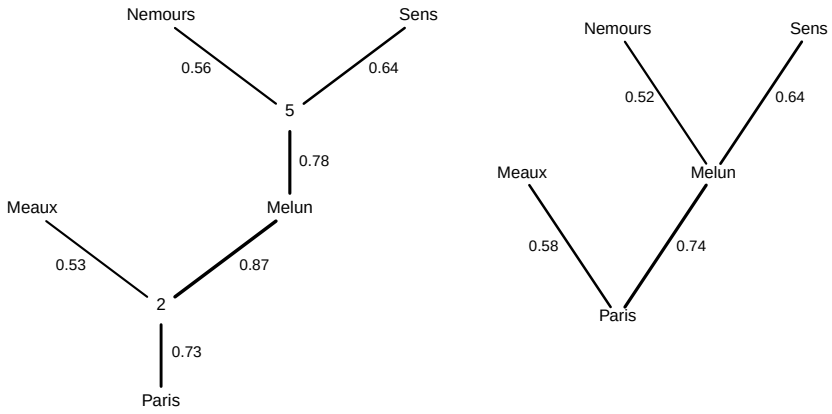


Figure 2.12: The trees of the tail dependence coefficients in equation (2.23), using the EC estimates for θ. Left: tree with nodes with latent variables; Right: reduced tree without latent variables.

2.7 Supplement

2.7.1 Relation to extremal graphical models in Engelke and Hitz (2020)

Engelke and Hitz (2020) introduce graphical models for extremes in terms of the multivariate Pareto distribution associated to a simple max-stable distribution G. We briefly review their approach and compare it with ours in case G is a Hüsler–Reiss distribution. Let $V = \{1, \ldots, d\}$ and let $X = (X_v, v \in V)$ be a random vector with unit-Pareto margins. The condition that $X \in D(G)$ is equivalent to

$$\lim_{t\to\infty} \big(\mathbb{P}(X_v \le t z_v, v \in V)\big)^t = G(z), \qquad z \in (0,\infty)^V.$$

By a direct calculation, it follows that

$$\begin{aligned} &\lim_{t\to\infty} \mathbb{P}\Big(X_v/t \le z_v, v \in V \,\Big|\, \max_{v\in V} X_v > t\Big) \\ &\qquad = \frac{\ln G\big(\min(z_v, 1), v \in V\big) - \ln G(z)}{\ln G(1, \ldots, 1)}, \end{aligned} \tag{2.24}$$

for $z \in (0,\infty)^V$, from which

$$(X_v/t, v \in V) \mid \max_{v\in V} X_v > t \xrightarrow{d} Y, \qquad t \to \infty \tag{2.25}$$

where $Y = (Y_v, v \in V)$ is a random vector whose distribution function is equal to the right-hand side in (3.6). The law of Y is a multivariate Pareto distribution, which, upon a change in location, is a special case of the multivariate generalized Pareto distributions arising in Rootzén and Tajvidi (2006) and Beirlant et al. (2004, Section 8.3) as limit distributions of multivariate peaks over thresholds.

Assuming that Y is absolutely continuous, its support is equal to the L-shaped set $\{y \in (0,\infty)^V : \max_{v\in V} Y_v > 1\}$ or a subset thereof, making conditional independence notions related to density factorizations ill-suited for Y. This is why Engelke and Hitz (2020) study conditional independence relations for the random vector Y^u defined in distribution as $Y \mid Y_u > 1$ for $u \in V$. According to Engelke and Hitz (2020, Definition 2), the law of Y is defined to be an extremal graphical model with respect to some graph $\mathcal{G}$ if for all $u \in V$, the law of Y^u satisfies the global Markov property with respect to $\mathcal{G}$. Note that Y itself is not required to satisfy the said Markov property.

The multivariate Pareto distribution derived through (3.6) from the Hüsler–Reiss distribution $G = H_\Lambda$ is referred to in Engelke and Hitz (2020) as the Hüsler–Reiss Pareto distribution. In their article, the term Hüsler–Reiss graphical models is then used for Hüsler–Reiss Pareto distributions that are extremal graphical models.

To show the relation with our approach, note that (2.25) implies that, for all $u \in V$, we have

$$(X_v/t, v \in V) \mid X_u > t \xrightarrow{d} Y^u, \qquad t \to \infty.$$

Recall the tail tree $(\Xi_{u,v}, v \in V \setminus u)$ in (2.7) and put $\Xi_{u,u} = 0$. Equations (2.10) and (2.13) in combination with Theorem 2 in Segers (2020b) and the continuous mapping theorem imply that

$$(X_v/t, v \in V) \mid X_u > t \xrightarrow{d} (\zeta \Xi_{u,v}, v \in V), \qquad t \to \infty,$$

where ζ is a unit-Pareto random variable, independent of the log-normal random vector $(\Xi_{u,v}, v \in V)$. Comparing the two previous limit relations, we find that Y^u is equal in distribution to $(\zeta \Xi_{u,v}, v \in V)$. The representation $\ln \Xi_{u,v} = \sum_{e\in(u\rightsquigarrow v)} \ln M_e$ as path sums starting from u over independent Gaussian increments $\ln M_e$ along the

edges implies that the Gaussian vector $(\ln \Xi_{u,v}, v \in V)$ satisfies the global Markov property with respect to $\mathcal{T}$. Since ζ is independent of $(\Xi_{u,v}, v \in U)$ this Markov property then also holds for $(\zeta \Xi_{u,v}, v \in V)$ and thus also for Y^u. But this means exactly that the multivariate Pareto distribution associated to the Hüsler–Reiss distribution with parameter matrix $\Lambda(\theta)$ in (2.5) is an extremal graphical model with respect to $\mathcal{T}$.

By way of comparison, the random vector Z^* constructed via properties (Z1)–(Z2) in Section 2.2.3 is not max-stable nor multivariate Pareto, but it satisfies the global Markov property with respect to $\mathcal{T}$ and it belongs to $D(H_{\Lambda(\theta)})$, motivating the chosen structure of $\Lambda(\theta)$ in (2.5). In Section 2.2.2, our assumption on ξ after transformation to X with unit-Pareto margins is that $X \in D(H_{\Lambda(\theta)})$. In this sense, we require that the extremal dependence of ξ is like the one of the graphical model Z^*. Our approach is thus different from the one in Engelke and Hitz (2020), who postulate a new definition of extremal graphical models for multivariate Pareto vectors, but without regard for the max-domain of attraction of the corresponding max-stable distributions. Still, for graphical models with respect to trees, both methods arrive at the same structure for the Hüsler–Reiss parameter matrix $\Lambda(\theta)$.

2.7.2 Comparison with the Lee–Joe structured Hüsler–Reiss model

Lee and Joe (2017) already proposed a way to bring structure to the parameter matrix $\Lambda = (\lambda_{ij}^2)_{i,j=1}^d$ of a d-variate max-stable Hüsler–Reiss distribution. Recall that Hüsler and Reiss (1989) studied the asymptotic distribution of the component-wise maxima of a triangular array of row-wise independent and identically distributed Gaussian random vectors, the n-th row having correlation matrix $\rho(n)$. Assuming $(1 - \rho_{ij}(n)) \ln(n) \to \lambda_{ij}^2$ as $n \to \infty$, they found the limit to be the distribution bearing their name. Motivated by this property, Lee and Joe (2017) propose to set $\lambda_{ij}^2 = (1 - \rho_{ij})\nu$ where $\rho = (\rho_{ij})_{i,j=1}^d$ is a structured correlation matrix and $\nu > 0$ is a free parameter. They then introduce the Hüsler–Reiss distributions that result from imposing on ρ the structure of a factor model or the one of a p-truncated vine. If $p = 1$, the latter becomes a Markov tree and we can compare their model with ours. In their case, a free correlation parameter $\alpha_e \in (-1, 1)$ is associated to each edge $e \in E$ of the tree on $V = \{1, \dots, d\}$. The correlation matrix ρ of the resulting Gaussian graphical model is

$$\rho_{ij} = \prod_{e \in (i \rightsquigarrow j)} \alpha_e, \qquad i, j \in V.$$

The Lee–Joe model for the structured Hüsler–Reiss matrix Λ_{LJ} derived from ρ is therefore

$$\lambda_{ij}^2 = (1 - \rho_{ij})\nu = \left(1 - \prod_{e \in (i \rightsquigarrow j)} \alpha_e\right)\nu, \qquad i, j \in V. \tag{2.26}$$

The model in (2.26) is to be compared with the one in our Eq. (2.5). The former has $(d - 1) + 1 = d$ free parameters, $(\alpha_e, e \in E)$ and ν, whereas the latter has only $d - 1$ free parameters $(\theta_e, e \in E)$. In Eq. (2.5), the Hüsler–Reiss parameters satisfy

$$\lambda_{ij}^2 = \sum_{e \in (i \rightsquigarrow j)} \lambda_e^2, \qquad i, j \in V,$$

where we write $\lambda_e = \lambda_{ab}$ for $e = (a, b) \in E$. In contrast, the Lee–Joe parameter matrix in Eq. (2.26) only satisfies this additivity relation asymptotically as $\nu \to \infty$. For instance, on a tree with $d = 3$ nodes and edges $(1, 2)$ and $(2, 3)$, i.e., a chain, their and our models satisfy respectively

$$\lambda_{13}^2 = \lambda_{12}^2 + \lambda_{23}^2 - \nu^{-1}\lambda_{12}^2\lambda_{23}^2 \qquad \text{for } \lambda_{ij}^2 \text{ as in Eq. (2.26)},$$

$$\lambda_{13}^2 = \lambda_{12}^2 + \lambda_{23}^2 \qquad \text{for } \lambda_{ij}^2 \text{ as in Eq. (2.5).}$$

Since the Lee–Joe parameter $\nu > 0$ takes the role of $\ln(n)$ in the Hüsler–Reiss limit relation, we can think of it as being large. In this interpretation, our parametrization becomes a limiting case of the one of Lee and Joe (2017).

Whereas the Lee–Joe parametrization is motivated from the limit result in Hüsler and Reiss (1989) for row-wise maxima of Gaussian triangular arrays, ours is motivated as the max-stable attractor of certain regularly varying Markov trees as in Segers (2020b), the vector Z^* in Section 2.2.3 serving as example. A possible advantage of our structure is that the resulting multivariate Pareto vector falls into the framework of conditional independence for such vectors is an extremal graphical model as in Engelke and Hitz (2020, Definition 2), as discussed in Section 2.7.1. In general, this is not true for the multivariate Pareto vector induced by the Lee–Joe structure. For the trivariate tree in the preceding paragraph, for instance, the criterion in Proposition 3 in Engelke and Hitz (2020) is easily checked to be verified for our matrix Λ but not for the one of Lee and Joe (2017).

For the Seine data, we compare the fitted Lee–Joe tail dependence model with ours. In order to avoid possible identifiability issues for the Lee–Joe parameters, we suppress the nodes with latent variables and use the right-hand tree in Figure 2.10 for the $d = 5$ observable ones, corresponding to the five locations in the dataset. The estimation method of Lee and Joe (2017) is based on pairwise copulas and annual maxima via composite likelihood. For year y and for variable $j \in \{1, \dots, d\}$, let $m_{y,j}$ be the maximum of all observations for that variable and that year, insofar available. These maxima are reported in Table 2.2 and their availability depends on the variable, i.e., on the location. For Melun there are only 15 such annual maxima in comparison to 33 for Nemours. For each variable j, transform these maxima to uniform margins $\hat{u}_{y,j}$ using the empirical cumulative distribution function based on all available maxima for that variable. Note that for this transformation, Lee and Joe (2017) rely on estimated generalized extreme value distributions instead. For variables $i, j \in \{1, \dots, d\}$, let $\mathcal{Y}_{ij}$ be the set of years y for which annual maxima are available for both variables. For the pair (Paris, Meaux) this is the period 1999–2019 while for the pair (Paris, Nemours) this is 1990–2019. Let $c(u, v; \lambda^2)$ denote the bivariate Hüsler–Reiss copula density with parameter λ^2. Following Lee and Joe (2017), we estimate the free parameters in Eq. (2.26) by maximizing a composite likelihood: letting $\lambda_{ij}^2(\alpha, \nu)$ denote the right-hand side in (2.26), the parameter estimates are

$$(\hat{\alpha}, \hat{\nu}) = \underset{\alpha \in (-1,1)^{d-1}, \nu \in (0,\infty)}{\arg\max} \sum_{i,j=1}^{d} \sum_{y \in \mathcal{Y}_{ij}} \ln c\big(\hat{u}_{i,y}, \hat{u}_{j,y}; \lambda_{ij}^2(\alpha, \nu)\big).$$

For the implementation, we relied on the `R` package `CopulaModel` (Krupskii, 2014).

Next, we compute bivariate extremal coefficients and compare them with the non-parametric ones on the one hand and with those obtained using our own model on the other hand. The points in Figure 2.13 being some distance away from the diagonal, the two methods indeed seem to give somewhat different results. Moreover, there is less concordance between the non-parametric estimates and the ones from the Lee–Joe model than between the non-parametric ones and those resulting from our model: compare the red crosses in Figure 2.13 with those in the left-hand plots in Figure 2.11.

2.7.3 Proof of Proposition 2.2.1

We show that the stdf l of Z^* is equal to l_U in (2.4) with $U = V$ and $\Lambda = \Lambda(\theta)$. Since the margins of Z^* are unit-Fréchet, they are tail equivalent to the unit-Pareto

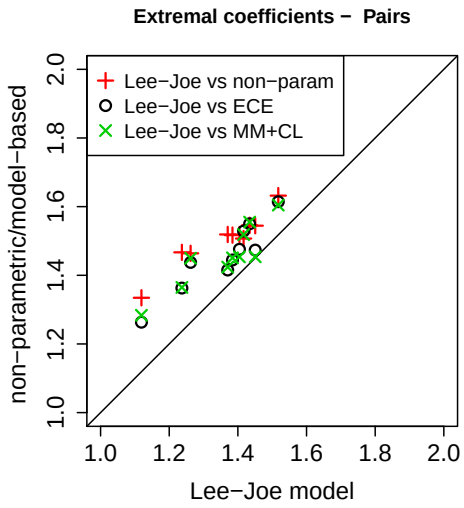


Figure 2.13: Bivariate extremal coefficients comparison: using the modelling and estimation method of Lee and Joe (2017, Section 4–5) and those proposed in this paper.

distribution, so that standardization to the unit-Pareto distribution is unnecessary. By the inclusion–exclusion principle,

$$\begin{aligned} l(x_1,\dots,x_d) &= \lim_{t\to\infty} t\ \mathbb{P}(Z_1^* > t/x_1 \text{ or } \dots \text{ or } Z_d^* > t/x_d) \\ &= \sum_{i=1}^{d}(-1)^{i-1} \sum_{\substack{W\subseteq V \\ |W|=i}} \lim_{t\to\infty} t\ \mathbb{P}(Z_v^* > t/x_v, v\in W) \end{aligned} \tag{2.27}$$

for $x \in (0,\infty)^d$. For any non-empty $W \subseteq V$ and any $u \in W$, it holds by (2.7) in combination with Theorem 2 in Segers (2020b) that

$$\begin{aligned} &\lim_{t\to\infty} t\ \mathbb{P}(Z_v^* > t/x_v, v\in W) \\ &= \lim_{t\to\infty} t\frac{1}{t/x_u}\mathbb{P}\left(\frac{Z_u^*}{t/x_u}\frac{Z_v^*}{Z_u^*} > \frac{x_u}{x_v}, v\in W\setminus u \,\middle|\, Z_u^* > \frac{t}{x_u}\right) \\ &= x_u\ \mathbb{P}(\zeta\Xi_{uv} > x_u/x_v, v\in W\setminus u), \end{aligned}$$

with ζ a unit-Pareto variable independent of $(\Xi_{uv}, v\in V\setminus u)$. Using the fact that $1/\zeta$ is a uniform variable on $[0,1]$ and setting $\Xi_{uu}=1$ we have

$$\begin{aligned} &x_u\ \mathbb{P}(\zeta\Xi_{uv} > x_u/x_v, v\in W\setminus u) \\ &= x_u\ \mathbb{P}\big(1/\zeta < \min\{(x_v/x_u)\Xi_{uv}, v\in W\setminus u\}\big) \\ &= x_u\,\mathbb{E}[\min\{1, (x_v/x_u)\Xi_{uv}, v\in W\setminus u\}] = \mathbb{E}[\min\{x_v\Xi_{uv}, v\in W\}] \\ &= \int_0^{x_u} \mathbb{P}\big(x_v\Xi_{uv} > y, v\in W\setminus u\big)\,\mathrm{d}y \\ &= \int_{-\ln x_u}^{\infty} \mathbb{P}\big(\ln\Xi_{uv} > (-\ln x_v) - z, v\in W\setminus u\big)\ \exp(-z)\,\mathrm{d}z \end{aligned}$$

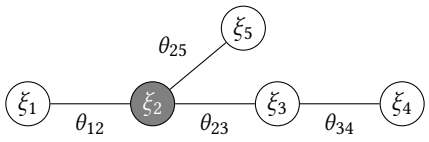
upon a change of variable $y=\exp(-z)$. Since $(\ln\Xi_{uv}, v\in V\setminus u)$ is multivariate normal with mean vector $\mu_{V,u}(\theta)$ and covariance matrix $\Sigma_{V,u}(\theta)$, we obtain from (2.27) that the stdf of Z^* is equal to the expression $-\ln H_{\Lambda(\theta)}(1/x_1,\dots,1/x_d)$, with H_Λ the

cumulative distribution function in Eqs. (3.5)–(3.6) in Hüsler and Reiss (1989), but with unit-Fréchet margins rather than Gumbel ones. By Remark 2.5 in Nikoloulopoulos et al. (2009), this stdf is equal to the one given in (2.4), as required.

2.7.4 Choice of node neighborhoods and parameter identifiability

The MM estimator in (2.18) and the CL estimator in Section 2.4.2 involve the choice of subsets $W_u \subseteq U$ for $u \in U$. These sets or neighborhoods need to be chosen in such a way that the parameter vector θ is still identifiable from the collection of covariance matrices $\Sigma_{W_u,u}(\theta)$ for $u \in U$ and thus from the path sums $p_{a,b}$ for $a, b \in W_u$ and $u \in U$. Here we illustrate this issue with an example.

Consider the following structure on five nodes where all variables are observable except for the one on node 2:



Clearly, the parameter vector $\theta = (\theta_{12}, \theta_{23}, \theta_{34}, \theta_{25})$ is identifiable from the distribution of the observable variables because the criterion of Proposition 2.3.1 is satisfied: the only node whose variable is latent has degree three.

First, consider the following subsets W_u for $u \in \{1, 3, 4, 5\}$:

$$W_1 = \{1, 5\}, \quad W_3 = \{3, 4\}, \quad W_4 = \{3, 4\}, \quad W_5 = \{1, 5\}.$$

The four 1×1 covariance matrices $\Sigma_{W_u,u}(\theta)$ that correspond to these subsets are

$$\begin{aligned} &\Sigma_{W_1,1}(\theta) = \theta_{12}^2 + \theta_{25}^2 = p_{15}, && \Sigma_{W_4,4}(\theta) = \theta_{34}^2 = p_{34}, \\ &\Sigma_{W_3,3}(\theta) = \theta_{34}^2 = p_{34}, && \Sigma_{W_5,5}(\theta) = \theta_{12}^2 + \theta_{25}^2 = p_{15}. \end{aligned}$$

We are not able to identify the parameter θ because the set of path sums $\{p_{15}, p_{34}\}$ is too small: we have only two equations and four unknowns.

Second, consider instead the following node sets

$$W_1 = \{1, 5, 3\}, \quad W_3 = \{1, 3, 4, 5\}, \quad W_4 = \{3, 4\}, \quad W_5 = \{1, 5\}.$$

The four covariance matrices $\Sigma_{W_u,u}(\theta)$ are now

$$\begin{aligned} \Sigma_{W_1,1}(\theta) &= \begin{bmatrix} \theta_{12}^2 + \theta_{25}^2 & \theta_{12}^2 \\ \theta_{12}^2 & \theta_{12}^2 + \theta_{23}^2 \end{bmatrix} = \begin{bmatrix} p_{15} & p_{12} \\ p_{12} & p_{13} \end{bmatrix}, \\ \Sigma_{W_4,4}(\theta) &= \theta_{34}^2 = p_{34}, \\ \Sigma_{W_3,3}(\theta) &= \begin{bmatrix} \theta_{12}^2 + \theta_{23}^2 & 0 & \theta_{23}^2 \\ 0 & \theta_{34}^2 & 0 \\ \theta_{23}^2 & 0 & \theta_{23}^2 + \theta_{25}^2 \end{bmatrix} = \begin{bmatrix} p_{13} & 0 & p_{23} \\ 0 & p_{34} & 0 \\ p_{23} & 0 & p_{35} \end{bmatrix}, \\ \Sigma_{W_5,5}(\theta) &= \theta_{12}^2 + \theta_{25}^2 = p_{15}. \end{aligned}$$

Clearly, the four edge parameters are identifiable from these covariance matrices.

2.7.5 Finite-sample performance of the estimators

We assess the performance of the three estimators introduced in Section 2.4 by numerical experiments involving Monte Carlo simulations.

Let $\xi' = (\xi'_v, v \in V)$ be a random vector with continuous joint probability density function and satisfying the global Markov property, (2.6), with respect to the graph in

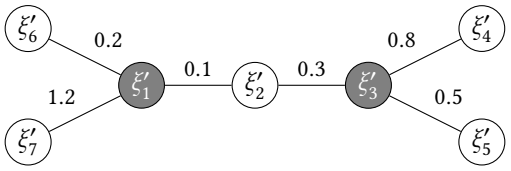


Figure 2.14: Tree used for the graphical model underlying the data-generating process in the simulation study in Section 2.7.5. The value of the parameters are $\theta_{12} = 0.1$, $\theta_{23} = 0.3$, $\theta_{34} = 0.8$, $\theta_{35} = 0.5$, $\theta_{16} = 0.2$ and $\theta_{17} = 1.2$. Variables ξ'_1 and ξ'_3 are latent.

Figure 2.14. Let $f_u(x_u)$ for any $u \in V$ be the marginal density function of the variable ξ'_u and let $x_j \mapsto f_{j|v}(x_j \mid x_v)$ be the conditional density function of ξ'_j given $\xi'_v = x_v$. For any $u \in V$ the joint density function of ξ' is

$$f(x) = f_u(x_u) \prod_{(v,j)\in E_u} f_{j|v}(x_j \mid x_v), \tag{2.28}$$

with $E_u \subseteq E$ the set of edges directed away from u, i.e., $(v, j) \in E_u$ if and only if $v = u$ or v separates u and j. The joint density f is determined by $d - 1$ bivariate densities f_{vj}. It would seem that the joint density f depends on u, but this is not so, as can be confirmed by writing out the bivariate conditional densities. We make two parametric choices: the univariate margins f_u are unit Fréchet densities, $f_j(x_j) = \exp(-1/x_j)/x_j^2$ for $x_j \in (0, \infty)$, and the bivariate margins for each pair of variables on adjacent vertices j, v are Hüsler–Reiss distributions with parameter θ_{jv}. Hence, ξ' corresponds to the vector Z^* in Section 2.2.3.

To generate an observation from the left hand-side of (6.5.1) above we use the right hand-side of that equation, proceeding iteratively, walking along paths starting from u using the conditional densities. An observation of ξ'_j given $\xi'_v = x_v$ is generated via the inverse function of the cdf $x_j \mapsto F_{j|v}(x_j \mid x_v)$, the conditional cdf of ξ'_j given $\xi'_v = x_v$. To do so, the equation $F_{j|v}(x_j \mid x_v) - p = 0$ is solved numerically as a function in x_j for fixed $p \in (0, 1)$. The choice of the Hüsler–Reiss bivariate distribution gives the following expression for $F_{j|v}(x_j \mid x_v)$:

$$\Phi\left(\frac{\theta_{jv}}{2} + \frac{1}{\theta_{jv}} \ln \frac{x_j}{x_v}\right)$$
$$\cdot \exp\left[-\frac{1}{x_v}\left\{\Phi\left(\frac{\theta_{jv}}{2} + \frac{1}{\theta_{jv}} \ln \frac{x_j}{x_v}\right) - 1\right\} - \frac{1}{x_j}\Phi\left(\frac{\theta_{jv}}{2} + \frac{1}{\theta_{jv}} \ln \frac{x_v}{x_j}\right)\right].$$

After generating all the variables $(\xi'_v)_{v\in V}$ in this way, independent standard normal noise $\varepsilon \sim \mathcal{N}_d(0, I_d)$ is added. Although the distribution of $\xi = \xi' + \varepsilon$ is not necessarily a graphical model with respect to the graph in Figure 2.14, it is still in the max-domain of attraction of a Hüsler–Reiss distribution with parametric matrix as in (2.5). Hence the vector ξ is still in the class of models under consideration in Section 2.2.2. The data on nodes 1 and 3 are discarded and not used in the estimation so as to mimic a model with two latent variables, ξ_1 and ξ_3; according to Proposition 2.3.1, the six dependence parameters are still identifiable. In this way, we generate 200 samples of size $n = 1000$. The estimators are computed with threshold tuning parameter $k \in \{25, 50, 100, 150, 200, 300\}$.

The bias, standard deviation and root mean squared errors of the three estimators are shown in Figure 2.15 and Figure 2.16 for the six parameters. The MME and CLE are computed with the sets W_u being $W_2 = \{2, 4, 5, 6, 7\}$, $W_4 = W_5 = \{2, 4, 5\}$, and $W_6 = W_7 = \{2, 6, 7\}$. As is to be expected, the absolute value of the bias is increasing with k, while the standard deviation is decreasing and the mean squared error has

a U-shape and eventually increases with k. The MME and CLE have very similar properties. For larger values of the true parameter, e.g. $\theta_{34} = 0.8$ and $\theta_{17} = 1.2$, all the three estimators perform in a comparable way. The ECE tends to have larger absolute bias and standard deviation for smaller values of the true parameters.

2.7.6 Seine case study: data preprocessing

The data represent water level in centimeters at the five locations mentioned above and were obtained from *Banque Hydro*, the web-site http://www.hydro.eaufrance.fr, of the Ministry of Ecology, Energy and Sustainable Development of France providing data on hydrological indicators across the country. The dataset encompasses the period from January 1987 to April 2019 with gaps for some of the stations.

Two major floods in Paris make part of our dataset: the one in June 2016 when the water level was measured at 6.01 m and the one at the end of January 2018 with water levels slightly less than 6 m measured in Paris too. A flood of similar magnitude to the ones in 2016 and 2018 occurred in 1982. By way of comparison, the biggest reported[1] flood in Paris is the one in 1910 when the level in Paris reached 8.6 m.

Table 2.1 shows the average and the maximum water level per station observed in the complete dataset. The maxima of Paris, Meaux, Melun and Nemours occurred either during the floods in June 2016 or the floods in January 2018, which can be seen from Table 2.2 which displays the annual maxima at the five locations and the date of occurrence.

Station	Paris	Meaux	Melun	Nemours	Sens
Period	1 Jan 1990 – 9 Apr 2019	1 Nov 1999 – 9 Apr 2019	1 Oct 2005 – 9 Apr 2019	16 Jan 1987 – 9 Apr 2019	1 Jan 1990 – 9 Apr 2019
(#obs)	(10,621)	(6,287)	(4,443)	(10,154)	(9,159)
Mean (cm)	139.11	275.85	296.61	210.07	133.46
Max (cm)	601.95	468.70	545.48	439.03	333.80

Table 2.1: Average and maximum water level per station in the whole dataset.

From Table 2.2 it can be observed that for many of the years the dates of maxima occurrence identify a period of several consecutive days during which the extreme event took place. For instance the maxima in 2007 occurred all in the period 4–8 March, which suggests that they make part of one extreme event. Similar examples are the periods 25–31 Dec 2010, 4–12 Feb 2013, 2–4 June 2016, etc. For most of the years this period spans between 3 and 7 days. We will take this into account when forming independent events from the dataset. In particular we choose a window of 7 consecutive calendar days within which we believe the extreme event have propagated through the seven locations. We have experimented with different length of that window, namely 3 and 5 days event period, but we have found that the estimation and analysis results are robust to that choice.

Figure 2.17 illustrates the water levels attained at the different locations during selected years from Table 2.2. The maxima of Sens, Nemours and Meaux seem to be relatively homogeneous compared to the maxima in Paris.

For all of the stations water level is recorded several times a day and we take the daily average to form a dataset of daily observations. Accounting for the gaps in the mentioned period (see Table 2.1 and Table 2.2) we end up with a dataset of 3408 daily observations in the period from 1 October 2005 to 8 April 2019. The dataset represents five time series each of length 3408. We consider two sources of non-stationarity: seasonality and serial correlation.

[1] According to the report of the Organisation for Economic Co-operation and Development (OECD) *Preventing the flooding of the Seine in the Paris – Ile de France region* - p.4.

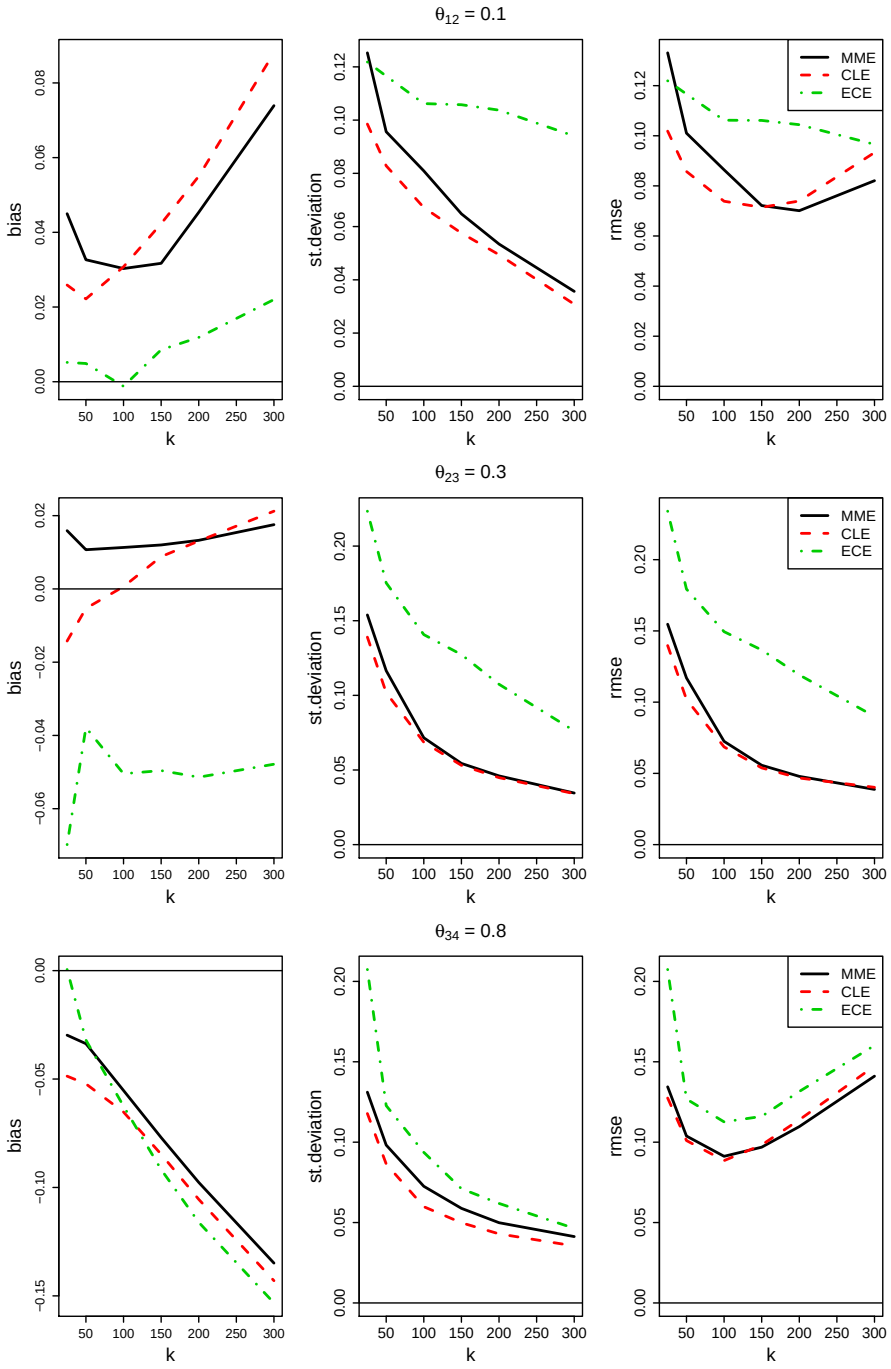


Figure 2.15: Bias (left), standard deviation (middle) and root mean squared error (right) of the method of moment estimator (MME), composite likelihood estimator (CLE) and pairwise extremal coefficient estimator (ECE) of the parameters θ_{12} (top), θ_{23} (middle), and θ_{34} (bottom) as a function of the threshold parameter k. Model and settings as described in Section 2.7.5.

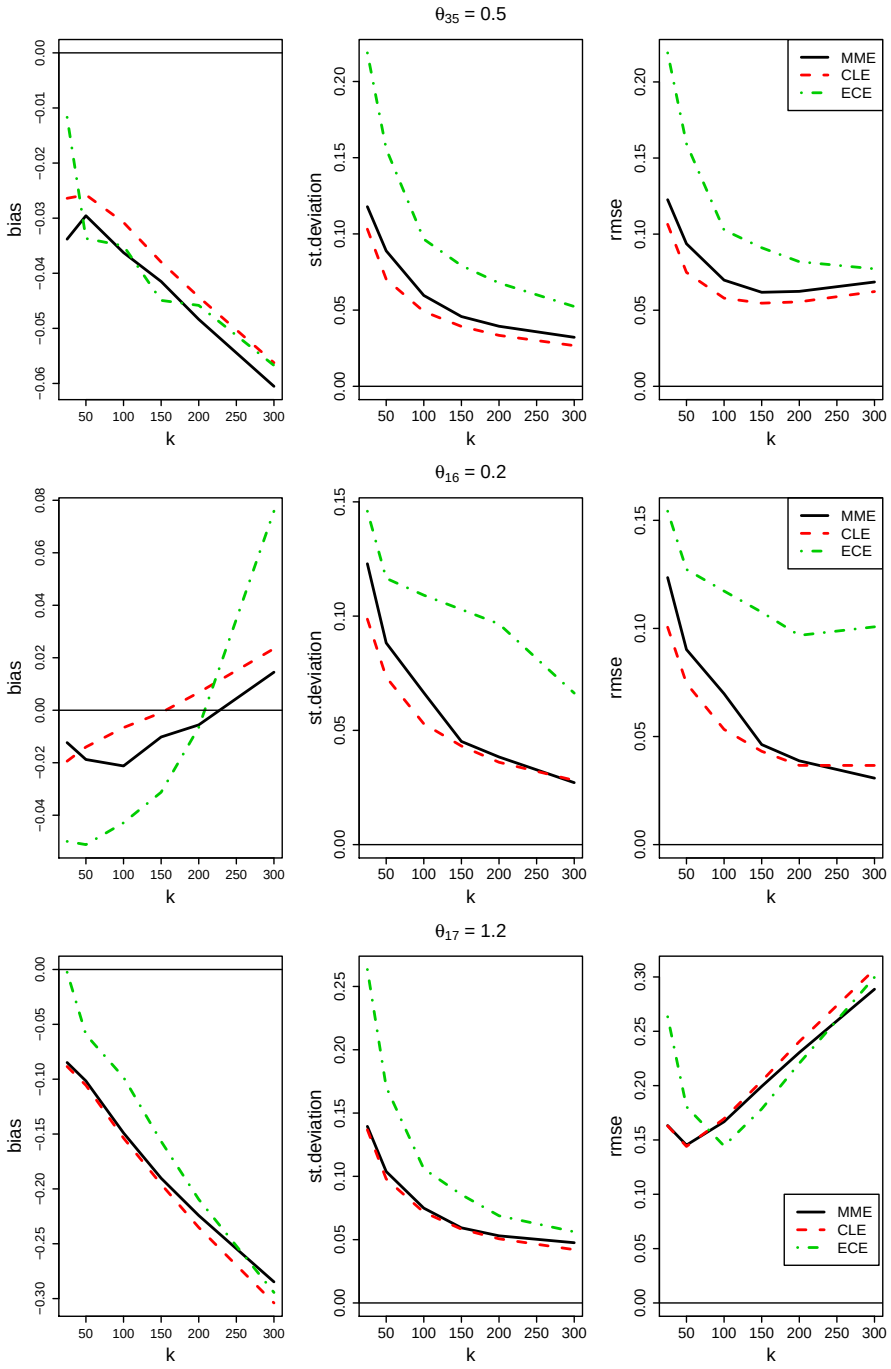


Figure 2.16: Bias (left), standard deviation (middle) and root mean squared error (right) of the method of moment estimator (MME), composite likelihood estimator (CLE) and pairwise extremal coefficient estimator (ECE) of the parameters θ_{35} (top), θ_{16} (middle), and θ_{17} (bottom) as a function of the threshold parameter k. Model and settings as described in Section 2.7.5.

Year	Paris		Meaux		Melun		Nemours		Sens	
	date	cm	date	cm	date	cm	date	cm	date	cm
1987	n/a	n/a	n/a	n/a	n/a	n/a	15/11	221	n/a	n/a
1988	n/a	n/a	n/a	n/a	n/a	n/a	13/02	247	n/a	n/a
1989	n/a	n/a	n/a	n/a	n/a	n/a	04/03	213	n/a	n/a
1990	17/02	254	n/a	n/a	n/a	n/a	03/07	217	18/02	183
1991	10/01	339	n/a	n/a	n/a	n/a	23/04	212	04/01	175
1992	06/12	293	n/a	n/a	n/a	n/a	15/01	218	06/12	170
1993	28/12	377	n/a	n/a	n/a	n/a	26/09	217	26/12	184
1994	11/01	478	n/a	n/a	n/a	n/a	19/10	253	09/01	260
1995	30/01	500	n/a	n/a	n/a	n/a	21/03	277	28/01	259
1996	04/12	324	n/a	n/a	n/a	n/a	03/12	219	04/12	194
1997	28/02	313	n/a	n/a	n/a	n/a	03/07	214	n/a	n/a
1998	02/05	358	n/a	n/a	n/a	n/a	21/12	216	n/a	n/a
1999	31/12	517	30/12	413	n/a	n/a	30/12	252	31/12	259
2000	01/01	515	02/01	407	n/a	n/a	07/06	233	01/01	239
2001	25/03	517	30/03	427	n/a	n/a	16/03	260	17/03	334
2002	03/03	410	03/03	403	n/a	n/a	01/01	272	01/01	200
2003	08/01	410	09/01	331	n/a	n/a	05/01	253	06/01	182
2004	21/01	372	21/01	383	n/a	n/a	16/01	230	20/01	205
2005	17/02	192	22/01	296	07/12	306	24/01	217	16/02	152
2006	14/03	340	08/10	333	13/03	357	11/03	219	12/03	223
2007	05/03	308	08/03	339	05/03	333	04/03	217	05/03	176
2008	29/03	301	01/01	250	23/03	342	15/04	219	23/03	167
2009	26/01	169	03/09	288	25/12	311	25/01	218	25/01	152
2010	28/12	387	31/12	355	27/12	390	25/12	230	26/12	220
2011	01/01	337	07/01	347	18/12	356	09/10	287	18/12	167
2012	09/01	330	23/12	308	09/01	353	05/01	220	08/01	186
2013	09/02	390	12/02	347	05/02	366	04/02	252	07/05	221
2014	03/03	273	13/12	295	16/02	321	02/03	226	15/02	157
2015	07/05	347	21/11	295	07/05	389	05/05	255	06/05	211
2016	03/06	602	03/06	329	03/06	545	02/06	439	04/06	235
2017	07/03	243	28/12	304	12/01	307	08/03	221	08/03	151
2018	29/01	586	02/02	469	28/01	488	24/01	264	26/01	288
2019	03/02	222	31/03	292	22/01	314	02/02	216	26/02	149

Table 2.2: Annual maxima for all stations. We highlighted some of the years where there is a clear indication that the dates of the occurrence of the maxima at the different locations form a period of several consecutive days. The maxima attained during this period across stations can thus be considered as one extreme event. The water level in centimeters is rounded to the nearest integer.

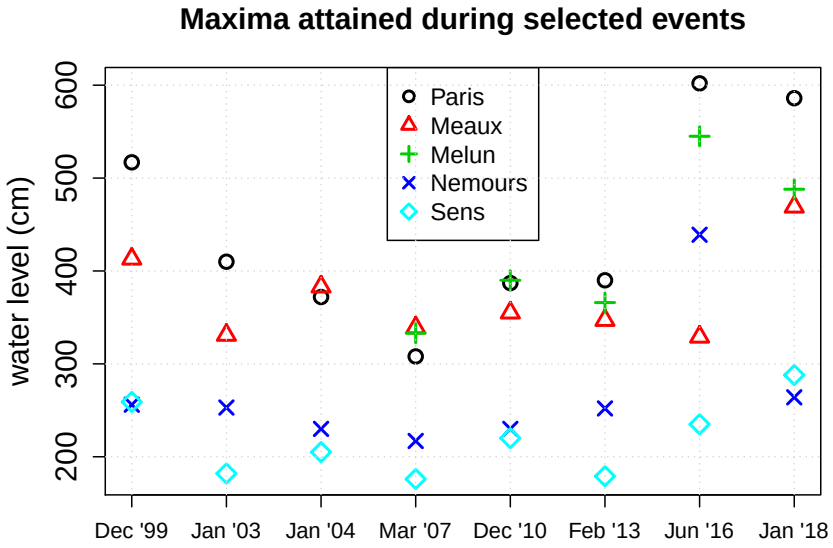


Figure 2.17: Plot of maxima attained at each location during selected events from Table 2.2.

The serial correlation can be due to closeness in time or presence of long term time trend in the observations. We first apply a declustering procedure, similar to the one in Asadi et al. (2015) in order to form a collection of supposedly independent events. As a first step each of the series is transformed to ranks and the sum of the ranks is computed for every day in the dataset. The day with the maximal rank is chosen, say d^*. A period of $2r + 1$ consecutive days, centered around d^* is considered and only the observations falling in that period are selected to form the event. Within this period the station-wise maximum is identified and the collection of the station-wise maxima forms one event. Because there is some evidence that the time an extreme event takes to propagate through the seven nodes in our model is about 3–7 days, we choose $r = 3$, hence we consider that one event lasts 7 days. In this way we obtain 717 observations of supposedly independent events. As it was mentioned the results are robust to the choice of $r = \{1, 2, 3\}$.

We test for seasonality and trends each of the series (each having 717 observations). The season factor is significant across all series and the time trend is marginally significant for some of the locations. We used a simple time series model to remove these non-stationarities. The model is based on season indicators and a linear time trend

$$X_t = \beta_0 + \beta_1 \mathbb{1}_{\text{spring}_t} + \beta_2 \mathbb{1}_{\text{summer}_t} + \beta_3 \mathbb{1}_{\text{winter}_t} + \alpha t + \epsilon_t, \tag{2.29}$$

where ϵ_t for $t = 1, 2, \ldots$ is a stationary mean zero process. After fitting the model in (2.29) to each of the five series through ordinary least squares we obtain the residuals and use those in the estimation of the extremal dependence.

2.7.7 ECE-based confidence interval for the dependence parameters

Let $\hat{\theta}_{n,k} = \hat{\theta}^{\text{ECE}}_{n,k}$ denote the pairwise extremal coefficient estimator in (6.3.3) and let θ_0 denote the true vector of parameters. By Einmahl et al. (2018, Theorem 2) with Ω

equal to the identity matrix, the ECE is asymptotically normal,

$$\sqrt{k}(\hat{\theta}_{n,k} - \theta_0) \xrightarrow{d} \mathcal{N}_{|E|}\big(0, M(\theta_0)\big), \qquad n \to \infty,$$

provided $k = k_n \to \infty$ such that $k/n \to 0$ fast enough (Einmahl et al., 2012, Theorem 4.6). The asymptotic covariance matrix takes the form

$$M(\theta_0) = (\dot{L}^\top \dot{L})^{-1} \dot{L}^\top \Sigma_L \dot{L} (\dot{L}^\top \dot{L})^{-1}.$$

The matrices $\dot{L}$ and Σ_L depend on θ_0 and are described below. For every k and every $e \in E$, an asymptotic 95% confidence interval for the edge parameter $\theta_{0,e}$ is given by

$$\theta_{0,e} \in \left[\hat{\theta}_{k,n;e} \pm 1.96 \sqrt{\{M(\hat{\theta}_{k,n})\}_{ee}/k} \right].$$

First, recall that $\mathcal{Q} \subseteq \{J \subseteq U : |J| = 2\}$ is the set of pairs on which the ECE is based and put $q = |\mathcal{Q}|$. Define the $\mathbb{R}^q$-valued map $L(\theta) = \big(l_J(1,1;\theta), J \in \mathcal{Q}\big)$ and let $\dot{L}(\theta) \in \mathbb{R}^{q \times |E|}$ be its matrix of partial derivatives. For a pair $J = \{u, v\}$ and an edge $e = (a, b)$, the partial derivative of $l_J(1,1;\theta)$ with respect to θ_e is given by

$$\frac{\partial l_J(1,1;\theta)}{\partial \theta_e} = \frac{\phi\left(\sqrt{p_{uv}}/2\right)}{\sqrt{p_{uv}}} \theta_e \mathbb{1}_{\{e \in (u \rightsquigarrow v)\}},$$

where p_{uv} is the path sum as in (2.15) and ϕ denotes the standard normal density function. The partial derivatives of $l_J(1,1;\theta)$ with respect to θ_e for every $e \in E$ form a row of the matrix $\dot{L}(\theta)$.

Second, $\Sigma_L(\theta_0)$ is the $q \times q$ covariance matrix of the asymptotic distribution of the empirical stdf,

$$\big\{\sqrt{k}\big(\hat{l}_{j;n,k}(1,1) - l_J(1,1;\theta_0)\big)\big\}_{m=1,\ldots,q} \xrightarrow{d} \mathcal{N}_q(0, \Sigma_L(\theta_0)), \qquad n \to \infty.$$

The elements of the matrix $\Sigma_L(\theta_0)$ are defined in terms of the stdf evaluated at different coordinates and of the partial derivatives of the stdf $l(x;\theta)$ with respect to the elements of x. For details we refer to Einmahl et al. (2018, Section 2.5). Here we note that the partial derivatives just mentioned are

$$\left.\frac{\partial l_J(x_u, x_v;\theta)}{\partial x_u}\right|_{(x_u,x_v)=(1,1)} = \Phi(\sqrt{p_{uv}}/2), \qquad J = \{u, v\}.$$

2.7.8 Bootstrap confidence interval for the Pickands dependence function

For assessing the goodness-of-fit of the proposed model (Section 2.5.2), we construct non-parametric 95% confidence intervals for $A(w) = l(1-w, w)$ for $w \in [0,1]$. As shown in Kiriliouk et al. (2018, Section 5) this can be achieved by resampling from the empirical beta copula. For every fixed $w \in [0,1]$ we seek with $a(w)$ and $b(w)$ such that

$$\mathbb{P}\big(a(w) \le \hat{l}_{n,k}(1-w, w) - l(1-w, w) \le b(w)\big) = 0.95,$$

where $\hat{l}_{n,k}$ is the non-parametric estimator of the stdf. For $a(w)$ and $b(w)$ satisfying the above expression, a point-wise confidence interval is given by

$$A(w) \in \left[\hat{l}_{n,k}(1-w, w) - b(w), \hat{l}_{n,k}(1-w, w) - a(w) \right]. \tag{2.30}$$

Let $(Y^*_{v,i})_{v \in U}$, for $i = 1, \ldots, n$, be a random sample from the empirical beta copula drawn according to steps A1–A4 of Kiriliouk et al. (2018, Section 5). Let the function

$\hat{l}^{\beta}_{n,k}$ be the empirical beta stdf based on the original data and let the function $\hat{l}^{*}_{n,k}$ be the non-parametric estimate of the stdf using the bootstrap sample.

We use the distribution of $\hat{l}^{*}_{n,k} - \hat{l}^{\beta}_{n,k}$ conditionally on the data as an estimate of the distribution of $\hat{l}_{n,k} - l$. Hence, we estimate $a(w)$ and $b(w)$ by $a^{*}(w)$ and $b^{*}(w)$ respectively defined implicitly by

$$\begin{aligned} 0.95 &= \mathbb{P}^{*}\left(a^{*}(w) \leq \hat{l}^{*}_{n,k}(1-w,w) - \hat{l}^{\beta}_{n,k}(1-w,w) \leq b^{*}(w)\right) \\ &= \mathbb{P}^{*}\left(a + \hat{l}^{\beta}_{n,k}(1-w,w) \leq \hat{l}^{*}_{n,k}(1-w,w) \leq b + \hat{l}^{\beta}_{n,k}(1-w,w)\right). \end{aligned}$$

We further estimate the bootstrap distribution of $\hat{l}^{*}_{n,k}$ by a Monte Carlo approximation obtained by $N = 1000$ samples of size n from the empirical beta copula. As a consequence, the lower and upper bounds for $\hat{l}^{*}_{n,k}(1-w,w)$ above are equated to the empirical 0.025- and 0.975-quantiles, respectively, yielding

$$\begin{aligned} \hat{l}^{*}_{0.025}(w,1-w) &= a^{*}(w) + \hat{l}^{\beta}_{n,k}(w,1-w), \\ \hat{l}^{*}_{0.975}(w,1-w) &= b^{*}(w) + \hat{l}^{\beta}_{n,k}(w,1-w). \end{aligned} \tag{2.31}$$

Replacing $a(w)$ and $b(w)$ in (2.30) by $a^{*}(w)$ and $b^{*}(w)$ respectively as solved from (2.31) yields the bootstrapped confidence interval for $A(w)$ shown in Figure 2.7.

Extremes of Markov random fields on block graphs: max-stable limits and structured Hüsler–Reiss distributions

3

This chapter corresponds to an article prepared with Johan Segers, carrying the same name as the chapter and currently in the second round of revision for the journal *Extremes*.

3.1 Introduction

Graphical models are statistical models for random vectors whose components are associated to the nodes of a graph, and edges serve to encode conditional independence relations. They bring structure to the web of dependence relations between the variables. In the context of extreme value analysis, they permit for instance to model the joint behavior of the whole random vector given that a specific component exceeds a high threshold. This could concern a system of intertwined financial risks, one of which is exposed to a large shock, or measurements of water heights along a river network, when a high water level is known to have occurred at a given location. Each time, the question is how such an alarming event affects conditional probabilities of similarly high values occurring elsewhere.

A Markov random field is a random vector satisfying a set of conditional independence relations with respect to a non-directed graph. For a max-stable random vector with continuous and positive density, Papastathopoulos and Strokorb (2016) showed that conditional independence implies unconditional independence. This implies that an absolutely continuous max-stable distribution can satisfy the Markov property with respect to a non-trivial graph only if variables on non-adjacent nodes are actually independent. These models clearly differ from the max-linear graphical models in Gissibl and Klüppelberg (2018) and Améndola et al. (2022) with respect to directed acyclic graphs which have max-stable, but singular, distributions. In our paper, conditional independence relations are induced by separation properties in undirected graphs, not by parent-child relations in directed ones.

Markov random fields with continuous and positive densities have made their way in extreme value analysis through the lens of multivariate Pareto distributions. Multivariate generalized Pareto distributions arise as weak limits of the normalized excesses over a threshold given the event that at least one variable exceeds a high threshold (Rootzén and Tajvidi, 2006). For a random vector $Y = (Y_1, \ldots, Y_d)$ with a multivariate Pareto distribution and a positive, continuous density, Engelke and Hitz

(2020) study conditional independence for the vectors $Y^{(k)} = (Y \mid Y_k > 1)$. They define Y as an extremal graphical model with respect to the graph for which $Y^{(k)}$ is an ordinary graphical model. In our paper we provide an example of a probabilistic graphical model in the classical sense whose graph is shared with the corresponding extremal graphical model.

We focus on Markov random fields with respect to connected block graphs, which generalize trees, because one obtains a block graph if the edges of a tree are replaced by complete subgraphs. Block graphs share some key properties with trees, such as unique shortest paths, acyclicality outside cliques and unique minimal separators. We study the limiting behavior of the normalized random field when a given variable exceeds a high threshold and we show that the limit depends on the unique shortest paths, a result familiar from Segers (2020b). As a prime example, we consider the case where the random vectors induced by the graph's cliques have limits determined by Hüsler–Reiss distributions.

Our main result, Theorem 3.3.5, is inspired by the one about Markov random fields on connected trees, or Markov trees in short, in Segers (2020b). Theorem 1 therein states that the limiting distribution of the scaled Markov tree given that a high threshold is exceeded at a particular node is a vector composed of products of independent multiplicative increments along the edges of the unique paths between the nodes. Here we show a similar result for block graphs rather than trees. This time, the products are with respect to the unique *shortest* path between pairs of nodes. The increments over the edges are independent between blocks but possibly dependent within blocks. The product structure of the limiting field originates from the asymptotic theory for Markov chains with a high initial state, going back to Smith (1992), Perfekt (1994), and Yun (1998). It is confirmed by many subsequent studies on Markov chains such as Segers (2007), Janssen and Segers (2014), Papastathopoulos et al. (2017) and Papastathopoulos and Tawn (2019).

The assumptions that underlie Theorem 3.3.5 are rather common in the literature. The first one says that the distribution of the random vector indexed by nodes within a block and scaled by the value of one of the variables, conditional on that variable taking a high value, converges to a non-degenerate distribution. The assumption is similar to the one in Heffernan and Tawn (2004) to model tail probabilities in case of asymptotic independence. Our version of the assumption implies multivariate regular variation of the clique vectors (Segers, 2020b, Theorem 2), making our model suitable for asymptotically dependent random vectors. An in-depth analysis of more refined forms of regular variation designed for modelling joint tails in case of asymptotic independence is given for instance in Heffernan and Resnick (2005, 2007), see also Resnick (2002) and Hernandez-Campos et al. (2005). The topic of asymptotic independence is investigated in Papastathopoulos et al. (2017) and Papastathopoulos and Tawn (2019) for possibly higher-order Markov chains and in Strokorb (2020) for extremal graphical models.

The second assumption in our Theorem 3.3.5 excludes processes which can become extreme again after reaching non-extreme levels. Earlier literature based on such regular behavior is Smith (1992), Perfekt (1994), Segers (2007), Janssen and Segers (2014), and Resnick and Zeber (2013). A toy model violating this principle is given in Segers (2007, Example 7.5). Markov chains exhibiting transitions from non-extreme to extreme regions are studied extensively in Papastathopoulos et al. (2017) and Papastathopoulos and Tawn (2019).

While the generalization of Markov trees to Markov fields on a larger class of graphs in Theorem 3.3.5 is the first novelty of our paper, the second novelty concerns the domain of attraction of the Hüsler–Reiss distribution in Section 3.4. For the study of extremal graphical models, the Hüsler–Reiss distribution offers many advantages akin to those of the Gaussian distribution for ordinary graphical models (Engelke and

Hitz, 2020). In Section 3.4, we study the implications of our main result for a Markov random field with respect to a block graph which has *clique-wise limits* based on Hüsler–Reiss distributions, or *Markov block graph with Hüsler–Reiss limits* in short. In Proposition 3.4.2, we show that for the said Markov block graphs with Hüsler–Reiss limits, the limiting distributions in Theorem 3.3.5 are all multivariate log-normal. In Engelke et al. (2014), such log-normal limits were found to characterize the domain of attraction of the Hüsler–Reiss max-stable distribution. Proposition 3.4.4 states that the parameter matrix of the max-stable limit has an explicit and elegant path-sum structure, as was found for Markov trees in Segers (2020b) and Asenova et al. (2021). By Proposition 3.4.5, the associated multivariate Pareto distribution is an extremal graphical model with respect to the same graph, complementing Proposition 4 in Engelke and Hitz (2020). Finally, it is proved in Proposition 3.4.6 that all parameters of the limiting structured Hüsler–Reiss max-stable distribution remain identifiable even when some of the variables are latent, as long as these variables lie on nodes belonging to least three different cliques. This generalizes a similar identifiability claim for trees in Asenova et al. (2021).

Our contribution can be cast within the classical paradigm in extreme value analysis to model maxima and high threshold excesses by distribution families validated by asymptotic theory. We find that probabilistic graphical models with respect to block graphs are in the domain of attraction of max-stable distributions that enjoy a particular structure induced by the graph and which for the Hüsler–Reiss family yield multivariate Pareto distributions that are extremal graphical models as in Engelke and Hitz (2020). Therefore, we believe that our results provide additional justification for the practical use of such models in situations where the data-generating mechanism can be described (approximately) as a graphical model on a block graph. The point is important when choosing between models. Indeed, starting from component-wise maxima of Gaussian vectors with structured correlation matrices, Lee and Joe (2017) propose a different way of incorporating graphical or factor structures into Hüsler–Reiss max-stable distributions; see Asenova, Mazo, and Segers (2021, Section A.2) for a discussion.

In our perspective the graph is known and we are interested in the tail limits of a random vector living on the nodes on that graph. This is suitable in applications such as extremes on river networks, or any application where we agree a priori on some graph. In Engelke and Volgushev (2020) the graph structure describing the dependence of extremes is the object of interest and the method proposed is nonparametric. A graph structure discovery based on Hüsler–Reiss Pareto models is proposed in Engelke and Hitz (2020).

The outline of the paper is as follows. In the preliminary Section 3.2 we introduce concepts and notation from graph theory, graphical models and extreme value analysis. The main result about the convergence of the rescaled random field, conditional on the event that a given variable exceeds a high threshold, is stated in Section 3.3. Section 3.4 concerns a Markov block graph composed of clique-wise distributions whose limits are determined by the Hüsler–Reiss family. The conclusion in Section 3.5 summarizes the main points and sketches some directions for further research. Proofs are deferred to Section 3.6.

3.2 Preliminaries

3.2.1 Graph theory and Markov random fields

A graph $\mathcal{G} = (V, E)$ is a pair consisting of a finite, non-empty vertex (node) set V and edge set $E \subseteq \{(a, b) \in V \times V : a \neq b\}$. Often, we will write $e \in E$ for a generic edge. The graph $\mathcal{G}$ is said to be non-directed if for every pair of nodes a, b we have

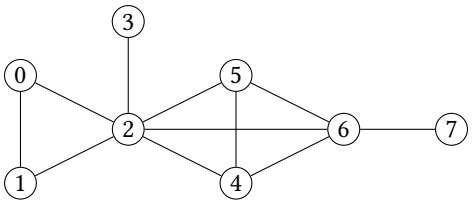


Figure 3.1: An example of a block graph. There are four blocks or (maximal) cliques, $C = \big\{\{0,1,2\},\{2,3\},\{2,4,5,6\},\{6,7\}\big\}$, as well as two minimal separators, $\mathcal{S} = \big\{\{2\},\{6\}\big\}$. The unique shortest path from 7 to 0 has edge set $(7 \rightsquigarrow 0) = \{(7,6),(6,2),(2,0)\}$.

$(a,b) \in E$ if and only if $(b,a) \in E$. A path from node a to node b is an ordered sequence of vertices $(v_1,\ldots,v_n)$ with $v_1 = a$ and $v_n = b$ such that $(v_i, v_{i+1}) \in E$ for all $i = 1,\ldots,n-1$ and in which all nodes are distinct, except possibly for the first and last nodes. A cycle is a path from a node to itself. Two distinct nodes are connected if there exists a path from one node to the other. A graph is connected if every pair of distinct nodes is connected. In this paper, we only consider connected, undirected graphs.

An induced subgraph $\mathcal{G}_A = (A, E_A)$ is formed from the vertices in a subset A of V and all edges connecting them, $E_A = \{(a,b) \in E : a,b \in A\}$. A graph is complete if every pair of distinct nodes is an edge. A set of nodes $C \subseteq V$ is said to be a clique if the induced subgraph $\mathcal{G}_C = (C, E_C)$ is complete; the latter graph will be called a clique as well. A clique is maximal if it is not properly contained in larger one. Further on we use the word 'clique' to mean maximal clique. The set of all (maximal) cliques of $\mathcal{G}$ will be denoted by $\mathcal{C}$.

A separator set $S \subseteq V$ between two other vertex subsets A and B is such that every path from a node in A to a node in B passes through at least one node in S. A separator set S is minimal when there is no proper subset of S which is a separator of A and B too.

In this paper we consider connected block graphs. A block is a maximal biconnected component, i.e., a subgraph that will remain connected after the removal of a single node. A block graph is a graph where every block is a clique; see Figure 3.1 for an example. If the edge between nodes 2 and 6 were removed, the subgraph induced by $\{2,4,5,6\}$ would still be a block, i.e., biconnected, but it would no longer be a clique, and so the graph would no longer be a block graph. Block graphs are considered natural generalizations of trees (Le and Tuy, 2010).

A path between two (distinct) nodes a and b is said to be shortest if no other path between a and b contains less nodes. In block graphs, any two nodes a and b are connected by a unique shortest path (Behtoei et al., 2010, Theorem 1), i.e., any other path connecting a and b contains strictly more nodes than the given path. If the shortest path between a and b is the ordered node set $(v_1,\ldots,v_n)$, with $v_1 = a$ and $v_n = b$, then we define $(a \rightsquigarrow b)$ as the set of edges

$$(a \rightsquigarrow b) = \{(v_1,v_2),\ldots,(v_{n-1},v_n)\}.$$

Another important property of block graphs is that cycles can only occur within cliques, i.e., a path which is not contained in a single clique has two different endpoints. Moreover, in a block graph, a minimal separator between two cliques is always a single node. Two distinct cliques have at most one node in common. The set of minimal clique separators will be denoted by $\mathcal{S}$. In the block graph in Figure 3.1, the collection of minimal clique separators is $\mathcal{S} = \{\{2\},\{6\}\}$.

Let $X = (X_v, v \in V)$ be a random vector which is indexed by the node set, V, of a graph $\mathcal{G} = (V,E)$. For non-empty $W \subseteq V$, write $X_W = (X_v, v \in W)$. We say that X is

a Markov random field with respect to $\mathcal{G}$ if it satisfies the global Markov property, that is, for all non-empty disjoint node sets $A, B, S \subset V$ we have the implication

$$S \text{ is a separator of } A \text{ and } B \text{ in } \mathcal{G} \implies X_A \perp\!\!\!\perp X_B \mid X_S,$$

where the right-hand side means that X_A and X_B are independent conditionally on X_S. In other words, conditional independence relations within X are implied by separation properties in $\mathcal{G}$. An extensive treatment of conditional independence, Markov properties and graphical models can be found in Lauritzen (1996).

Often we will use a double subscript to a random vector, e.g., $X_{u,A}$ for $A \subseteq V$, which, if not indicated otherwise, will mean that it is a vector indexed by the elements of A and in some way related to a particular node u, not necessarily in A. For scalars, the expressions $x_{u,v}$ and x_{uv} will signify the same thing and if $e = (u, v)$ is an edge they can also be written as x_e. In case of iterated subscripts, we will prefer the comma notation, x_{u_1,u_2}.

For two non-empty sets A and B, let B^A denote the set of functions $x : A \to B$. Formally, we think of x as a vector indexed by A and with elements in B, as reflected in the notation $x = (x_a, a \in A)$. We will apply this convention most often to subsets A of the node set V of a graph $\mathcal{G}$ and to subsets B of the extended real line.

3.2.2 Max-stable and multivariate Pareto distributions

Let $X = (X_v, v \in V)$ be a random vector indexed by a finite, non-empty set V and with joint cumulative distribution function F, the margins of which are continuous, i.e., have no atoms. The interest in this paper is in tail dependence properties of X. It is convenient and does not entail a large loss of generality to assume that the margins have been standardized to a common distribution, a convenient choice of which will be either the unit-Pareto distribution, $\mathbb{P}(X_v > x) = 1/x$ for $v \in V$ and $x \geq 1$, or the unit-Fréchet distribution, $\mathbb{P}(X_v \leq x) = \exp(-1/x)$ for $v \in V$ and $x > 0$. Assume that F is in the max-domain attraction of a multivariate extreme-value distribution G, i.e., for either of the two choices of the common marginal distribution we have

$$\forall z \in (0, \infty)^V, \qquad \lim_{n\to\infty} F^n(nz) = G(z), \tag{3.1}$$

a condition which will be denoted by $F \in D(G)$. Let $X^{(n)} = (X_v^{(n)}, v \in V)$ for $n = 1, 2, \ldots$ be a sequence of independent and identically distributed random vectors with common distribution F. Let $M^{(n)} = \left(M_v^{(n)}, v \in V\right)$ with $M_v^{(n)} = \max_{i=1,\ldots,n} X_v^{(i)}$ be the vector of component-wise sample maxima. Equation (3.1) then means that

$$M^{(n)}/n \xrightarrow{d} G, \qquad n \to \infty, \tag{3.2}$$

the arrow $\xrightarrow{d}$ signifying convergence in distribution. The choice of the scaling sequence n in (3.1) and (3.2) is dictated by the marginal standardization and implies that the margins of G are unit-Fréchet too, i.e., G is a simple max-stable distribution. The latter can be written as

$$G(z) = \exp\left(-\mu\left[\left\{x \in [0, \infty)^V : \exists v \in V,\, x_v > z_v\right\}\right]\right), \qquad z \in (0, \infty]^V, \tag{3.3}$$

where the exponent measure μ is a non-negative Borel measure on the punctured orthant $[0, \infty)^V \setminus \{0\}$, finite on subsets bounded away from the origin (de Haan and Resnick, 1977; Resnick, 1987). The function $\ell : [0, \infty)^d \to [0, \infty)$ defined by

$$\ell(y) := -\ln G(1/y_v, v \in V) = \mu\left[\left\{x \in [0, \infty)^V : \exists v \in V,\, x_v > 1/y_v\right\}\right] \tag{3.4}$$

More background on multivariate extreme value analysis can be found for instance in the monographs Resnick (1987), Beirlant et al. (2004) and de Haan and Ferreira (2007).

We can replace the integer n in (3.1) by the real scalar $t > 0$: the condition $F \in D(G)$ is equivalent to

$$\forall z \in (0,\infty)^V, \qquad \lim_{t\to\infty} F^t(tz) = G(z). \tag{3.5}$$

By a direct calculation starting from (3.5) it follows that

$$\begin{aligned} &\lim_{t\to\infty} \mathbb{P}\left(\forall v \in V,\, X_v/t \le z_v \;\middle|\; \max_{v\in V} X_v > t\right) = \\ &\frac{\ln G\big(\min(z_v,1), v\in V\big) - \ln G(z)}{\ln G(1,\dots,1)}, \end{aligned} \tag{3.6}$$

for $z \in (0,\infty)^V$, from which we deduce the weak convergence of conditional distributions

$$\left(t^{-1}X \;\middle|\; \max_{v\in V} X_v > t\right) \xrightarrow{d} Y, \qquad t\to\infty, \tag{3.7}$$

where $Y = (Y_v, v \in V)$ is a random vector whose cumulative distribution function is equal to the right-hand side in (3.6). The law of Y is a multivariate Pareto distribution and has support contained $[0,\infty)^V \setminus [0,1]^V$. Upon a change in location, it is a member of the family of multivariate generalized Pareto distributions. The latter arise in Rootzén and Tajvidi (2006) and Beirlant et al. (2004, Section 8.3) as limit laws of multivariate peaks over threshold; see also Rootzén et al. (2018).

3.3 Tails of Markov random fields on block graphs

Let $X = (X_v, v \in V)$ be a non-negative Markov random field with respect to the connected block graph $\mathcal{G} = (V, E)$, or Markov block graph in short. Suppose that at a given node $u \in V$ the variable X_u exceeds a high threshold, say t. This event can be expected to affect conditional probabilities of the other variables X_v too. Our main result, Theorem 3.3.5, states that, starting out from u, every other variable X_v feels the impact of the shock at u through a multiplication of increments on the edges forming the unique shortest path from u to v. The increments are independent between cliques but possibly dependent within cliques.

After discussing the set-up and the assumptions in Section 3.3.1, we state and illustrate the main result in Section 3.3.2. Consequences for multivariate regular variation and max-domains of attraction are treated in Section 3.3.3, followed by a focus in Section 3.3.4 on the special case where clique vectors are max-stable already.

3.3.1 Set-up

We will be making two assumptions on the conditional distribution of X at high levels. Assumption 3.3.1 is the main one, as it will determine the limit distribution in Theorem 3.3.5 through the construction in Definition 3.3.2 below. For a set A and an element $b \in A$, we write $A \setminus b$ rather than $A \setminus \{b\}$.

Assumption 3.3.1. *For every clique $C \in \mathcal{C}$ and every node $s \in C$ there exists a probability distribution $\nu_{C,s}$ on $[0,\infty)^{C\setminus s}$ such that, as $t \to \infty$, we have*

$$\mathcal{L}\left(\frac{X_v}{t},\, v \in C\setminus s \;\middle|\; X_s = t\right) \xrightarrow{d} \nu_{C,s}.$$

In the special case that the distribution of the clique vector X_C is max-stable, the limit $\nu_{C,s}$ can be calculated by means of Proposition 3.3.11 below.

Definition 3.3.2 (Increments). *Under Assumption 3.3.1, define, for fixed $u \in V$, the following $(|V|-1)$-dimensional non-negative random vector Z:*

(Z1) For each clique C, let s be the separator node in C between u and the nodes in C. If $u \in C$, then simply $s = u$, whereas if $u \notin C$, then s is the unique node in C such that for every $v \in C$, any path from u to v passes through s. Note that, for fixed u, the node s is a function of C, but we will suppress this dependence from the notation.

(Z2) For each C, consider the limit distribution $\nu_{C,s}$ of Assumption 3.3.1, with separator node $s \in C$ determined as in (Z1).

(Z3) Put $Z := (Z_{s,C\setminus s}, C \in \mathcal{C})$ where, for each $C \in \mathcal{C}$, the random vector $Z_{s,C\setminus s} = (Z_{sv}, v \in C \setminus s)$ has law $\nu_{C,s}$ and where these $|\mathcal{C}|$ vectors are mutually independent as C varies.

The distribution of the random vector Z in Definition 3.3.2 depends on the source node u, but this dependence is suppressed in the notation. The dimension of Z is indeed equal to $|V|-1$, since the sets $C \setminus s$ form a partition of $V \setminus u$ as C varies in $\mathcal{C}$. In fact, in the double index in Z_{sv}, every node $v \in V \setminus u$ appears exactly once; the node s is the one just before v itself on the shortest path from u to v.

Example 3.3.3. *Consider a Markov random field on the block graph in Figure 3.1. Suppose that the variable exceeding a high threshold is the one at node $u = 7$. For paths departing at u, the separator nodes associated to the four cliques are as follows:*

clique C	*separator node $s \in C$*	*node set $C \setminus s$*
$\{0,1,2\}$	2	$\{0,1\}$
$\{2,3\}$	2	$\{3\}$
$\{2,4,5,6\}$	6	$\{2,4,5\}$
$\{6,7\}$	7	$\{6\}$

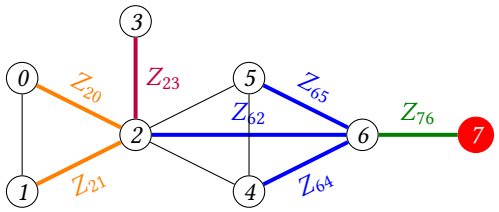
Note that the union over the sets $C \setminus s$ is equal to $\{0,1,\dots,6\} = V \setminus u$. Assumption 3.3.1 requires certain joint conditional distributions to converge weakly: as $t \to \infty$, we have

$$\left(\frac{X_0}{t}, \frac{X_1}{t} \,\middle|\, X_2 = t\right) \xrightarrow{d} \nu_{\{0,1,2\},2}, \qquad \left(\frac{X_3}{t} \,\middle|\, X_2 = t\right) \xrightarrow{d} \nu_{\{2,3\},2},$$
$$\left(\frac{X_2}{t}, \frac{X_4}{t}, \frac{X_5}{t} \,\middle|\, X_6 = t\right) \xrightarrow{d} \nu_{\{2,4,5,6\},6}, \qquad \left(\frac{X_6}{t} \,\middle|\, X_7 = t\right) \xrightarrow{d} \nu_{\{6,7\},7}.$$

The random vector Z in Definition 3.3.2, step (Z3), is a 7-dimensional random vector whose joint distribution is equal to the product of the above four distributions:

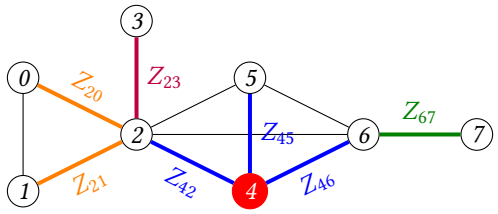
$$\begin{aligned} Z &:= (Z_{20}, Z_{21}; Z_{23}; Z_{62}, Z_{64}, Z_{65}; Z_{76}) \\ &\sim \nu_{\{0,1,2\},2} \otimes \nu_{\{2,3\},2} \otimes \nu_{\{2,4,5,6\},6} \otimes \nu_{\{6,7\},7}. \end{aligned} \tag{3.8}$$

We think of the random variable Z_{sv} as being associated to the edge $(s,v) \in E$:



By construction, the random sub-vectors (Z_{20}, Z_{21}), Z_{23}, (Z_{62}, Z_{64}, Z_{65}) *and* Z_{76} *are independent from each other and their marginal distributions are* $(Z_{20}, Z_{21}) \sim \nu_{\{0,1,2\},2}$ *and so on. Every node* $v \in \{0, 1, \ldots, 6\}$ *appears exactly once as a second index of a variable in* Z_{sv} *in* (3.8)*. For each such* v*, the first index* s *is the node right before* v *on the path from* $u = 7$ *to* v*.*

In the same block graph, we could also suppose that the variable exceeding a high threshold is the one on node $u = 4$*. The picture would then change as follows:*



Again, colour-coded sub-vectors corresponding to cliques are mutually independent. The vectors (Z_{21}, Z_{21}) *and* Z_{23} *are equal in distribution to those as when the starting node was* $u = 7$*, but the vectors* (Z_{42}, Z_{45}, Z_{46}) *and* Z_{67} *are new. In particular,* Z_{67} *is not the same as* Z_{76}*, having different indices of the conditioning variable in Assumption 3.3.1.* ◊

If the univariate margins of X_C are of Pareto-type, Assumption 3.3.1 implies that the distribution of X_C is regularly varying according to Segers (2020b, Corollary 1 and Theorem 2); see also Corollary 3.3.9 below. In case the components of X_C are asymptotically independent (Resnick, 1987), however, the limit measure $\nu_{C,s}$ is degenerate at zero. This knowledge is unhelpful for studying extremes of near independent data (Ledford and Tawn, 1996, 1997; Heffernan and Tawn, 2004) because it does not allow for modeling, inference or extrapolation. To accommodate cases of asymptotic independence, a refinement of the assumptions would be necessary. This is not pursued in the paper. A starting point would be the studies on Markov chains in Papastathopoulos et al. (2017) and Papastathopoulos and Tawn (2019).

In Assumption 3.3.1, let $\nu^v_{C,s}$ denote the univariate marginal distribution corresponding to node $v \in C \setminus s$. Recall that $\mathcal{S}$ denotes the set of minimal separator nodes between the cliques in the block graph.

Assumption 3.3.4. *Let* $\{u, \ldots, s\}$ *be the sequence of nodes of the unique shortest path between two nodes* $u \in V$ *and* $s \in \mathcal{S}$*. Let* C *be any clique which contains* s*, but no other node of* $\{u, \ldots, s\}$*. If there is an edge* (a, b) *on the path* $(u \rightsquigarrow s)$ *such that* $\nu^b_{C',a}(\{0\}) > 0$*, where* C' *is the (unique) clique containing the nodes* a *and* b*, then for any* $\eta > 0$*, we have*

$$\limsup_{\delta \downarrow 0} \limsup_{t \to \infty} \sup_{x_s \in [0,\delta]} \mathbb{P}\left(\exists v \in C \setminus s : X_v/t > \eta \mid X_s/t = x_s\right) = 0. \tag{3.9}$$

Figure 3.2 illustrates the scope of Assumption 3.3.4. For the variables in clique C', Assumption 3.3.1 implies that as $t \to \infty$,

$$\mathcal{L}\left(\frac{X_v}{t}, v \in C' \setminus a \,\middle|\, X_a = t\right) \xrightarrow{d} \nu_{C',a}.$$

The univariate margin $\nu^b_{C',a}$ related to component $b \in C' \setminus a$ could have positive mass at zero, which is the condition $\nu^b_{C',a}(\{0\}) > 0$ in Assumption 3.3.4. The condition is similar to the one in Segers (2020b, page 860) for Markov trees and the one for Markov chains in Papastathopoulos, Strokorb, Tawn, and Butler (2017, Section 4). In Figure 3.2, consider the clique C with nodes s, v_1, v_2. Relation (3.9) says that the probability that at least one rescaled variable, X_{v_1}/t or X_{v_2}/t, exceeds a small but positive value given that at the separator node, s, the value is very small already, is

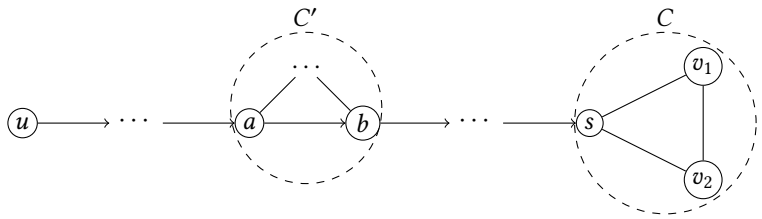


Figure 3.2: The path from u to s contains edge (a, b) which is part of clique C'. The end node s belongs to clique C which contains further nodes v_1 and v_2. Assumption 3.3.4 concerns the distribution of X_{v_j}/t given X_s/t in case the b-th marginal of $\nu_{C',a}$ has positive mass at zero, as explained below the assumption.

close to zero. In other words, in the limit all scaled variables must be smaller than any arbitrary small value, $\eta > 0$, given that the scaled variable X_s/t at the separator is known to have a small value, i.e., $x_s \in [0, \delta]$ with $\delta \downarrow 0$. Predecessors of condition (3.9) are found in earlier literature on Markov chains: Smith (1992, page 39) and Perfekt (1994, page 537). Papastathopoulos et al. (2017) and Papastathopoulos and Tawn (2019) provide limiting results for Markov chains of arbitrary order when the process is allowed to switch between non-extreme and extreme states. Example 7.5 in Segers (2007) provides a Markov chain that does not satisfy Assumption 3.3.4.

3.3.2 Main result

Theorem 3.3.5. *Let $X = (X_v, v \in V)$ be a non-negative Markov random field with respect to the connected block graph $\mathcal{G} = (V, E)$. Let Assumptions 3.3.1 and 3.3.4 be satisfied. For a given $u \in V$, let Z be the random vector in Definition 3.3.2. Then as $t \to \infty$, we have*

$$\left(\frac{X_v}{t}, v \in V \setminus u \;\middle|\; X_u = t \right) \xrightarrow{d} (A_{uv}, v \in V \setminus u) =: A_{u,V\setminus u}$$

$$\text{where} \quad A_{uv} := \prod_{e \in (u \rightsquigarrow v)} Z_e. \tag{3.10}$$

Remark 3.3.6. If the block graph is a tree, Theorem 3.3.5 reduces to Theorem 1 in Segers (2020b). In the more general case considered here, the increments Z_e can be dependent within a block, although they are still independent between blocks. Note that for a single variable A_{uv}, the increments Z_e appearing in (3.10) are independent, even for a block graph that is not a tree. The difference between a tree and a more general block graph thus manifests itself in the joint distribution of the random variables A_{uv} for $v \in V \setminus u$.

Example 3.3.7. *We continue with Example 3.3.3. Let the variable exceeding a high threshold be the one on node $u = 7$. The conclusion of Theorem 3.3.5 is that as $t \to \infty$, we have*

$$\big(X_v/t, v \in \{0, 1, \ldots, 6\} \mid X_7 = t\big) \xrightarrow{d} \big(A_{7v}, v \in \{0, 1, \ldots, 6\}\big),$$

where, in the notation of Example 3.3.3, the limiting variables have the following structure:

$$\begin{aligned}
A_{7,\{0,1,2\}} &= (A_{70}, A_{71}, A_{72}) &&= Z_{76} Z_{62} (Z_{20}, Z_{20}, 1), \\
A_{7,\{2,3\}} &= (A_{72}, A_{73}) &&= Z_{76} Z_{62} (1, Z_{23}), \\
A_{7,\{2,4,5,6\}} &= (A_{72}, A_{74}, A_{75}, A_{76}) &&= Z_{76} (Z_{62}, Z_{64}, Z_{65}, 1), \\
A_{7,\{6\}} &= A_{76} &&= Z_{76}.
\end{aligned}$$

The limit vector $(A_{7v})_v$ is similar to the one of a Markov field with respect to the tree formed by the unique shortest paths from node $u = 7$ to the other nodes. The difference is that within a block, the multiplicative increments need not be independent (Remark 3.3.6). ◊

Remark 3.3.8. A useful result that we will need further on is that the convergence in (3.10) implies a self-scaling property of the Markov random field at high levels: as $t \to \infty$, we have

$$\left(\frac{X_v}{X_u}, v \in V \setminus u \;\middle|\; X_u > t \right) \xrightarrow{d} (A_{u,v}, v \in V \setminus u). \tag{3.11}$$

The proof of (3.11) is the same as that of Corollary 1 in Segers (2020b).

3.3.3 Max-domains of attraction

In multivariate extreme value analysis, it is common to work with standardized margins, often the unit-Pareto or unit-Fréchet distribution. Both cases are covered in the next corollary, which, in the terminology of Section 3.2.2, states that X is multivariate regularly varying and in the domain of attraction of a max-stable distribution G with exponent measure μ determined by the limits A_{uv} in (3.10). Let $\mathcal{M}_0$ denote the set of Borel measures μ on $[0,\infty)^V \setminus \{0\}$ such that $\mu(B)$ is finite whenever the Borel set B is contained in a set of the form $\{x \in [0,\infty)^V : \max_{v\in V} x_v \geq \varepsilon\}$ for some $\varepsilon > 0$. The set $\mathcal{M}_0$ is equipped with the smallest topology that makes the evaluation mappings $\mu \mapsto \int f \,\mathrm{d}\mu$ continuous, where f varies over the collection of real-valued, bounded, continuous functions on $[0,\infty)^V \setminus \{0\}$ that vanish in a neighbourhood of the origin.

Corollary 3.3.9. *If, in addition to the assumptions in Theorem 3.3.5, we have $t\,\mathbb{P}(X_v > t) \to 1$ as $t \to \infty$ for all $v \in V$, then in $\mathcal{M}_0$ we have the convergence*

$$t\,\mathbb{P}(t^{-1}X \in \cdot\,) \to \mu, \qquad t \to \infty, \tag{3.12}$$

with limit measure μ on $[0,\infty)^V \setminus \{0\}$ determined by

$$\int f(x)\, \mathbb{1}\{x_u > 0\} \,\mathrm{d}\mu(x) = \mathbb{E}\left[\int_0^\infty f(zA_u) z^{-2} \,\mathrm{d}z \right] \tag{3.13}$$

for $u \in V$, Borel-measurable $f : [0,\infty)^V \setminus \{0\} \to [0,\infty]$, and $A_u = (A_{uv})_{v\in V}$ as in (3.10) *with additionally $A_{uu} = 1$. As a consequence, for $x \in (0,\infty)^V$, we have*

$$\lim_{n\to\infty} [\mathbb{P}(X_v \leq n x_v, v \in V)]^n = G(x) = \exp[-\mu(\{y : \exists v \in V, y_v > x_v\})]. \tag{3.14}$$

The indicator $\mathbb{1}\{x_u > 0\}$ in (3.13) can often be omitted, simplifying the formula for G.

Corollary 3.3.10. *In Corollary 3.3.9, if $u \in V$ is such that the increments Z_e satisfy $\mathbb{E}[Z_e] = 1$ for every $e \in (u \rightsquigarrow v)$ and every $v \in V \setminus u$, then $\mu(\{x : x_u = 0\}) = 0$ and we actually have*

$$\int f(x) \,\mathrm{d}\mu(x) = \mathbb{E}\left[\int_0^\infty f(zA_u) z^{-2} \,\mathrm{d}z \right] \tag{3.15}$$

for Borel-measurable $f : [0,\infty)^V \setminus \{0\}$. In particular, the stable tail dependence function of the max-stable limit G in (3.14) *is then*

$$\ell(x) = -\log G(1/x_v, v \in V) = \mathbb{E}\left[\max\left\{ x_v A_{u,v}, v \in V \right\} \right], \qquad x \in [0,\infty)^V. \tag{3.16}$$

For $e = (a,b)$, the condition $\mathbb{E}[Z_{ab}] = 1$ is equivalent to $\mathbb{P}(Z_{ba} > 0) = 1$.

Proof of Corollary 3.3.9. Thanks to Remark 3.3.8 above, the measure convergence (3.12) follows from Theorem 2 in Segers (2020b) . By a standard argument (Resnick, 1987, Proposition 5.17), the latter implies that $t\,\mathbb{P}(\exists v \in V, X_v > t x_v)$ converges to $\mu(\{y : \exists v \in V, y_v > x_v\})$ for $x \in (0,\infty)^V$, yielding (3.14). □

Proof of Corollary 3.3.10. The equivalence of $\mathbb{E}[Z_{ab}] = 1$ and $\mathbb{P}(Z_{ba} > 0) = 1$ is a consequence of Corollary 3 in Segers (2020b), while formula (3.15) follows from Corollary 4 in the same source and the fact that $\mathbb{E}[A_{uv}] = \prod_{e\in(u\rightsquigarrow v)} \mathbb{E}[Z_e] = 1$ for all $v \in V \setminus u$; note that the edges e on a path $(u \rightsquigarrow v)$ all belong to different blocks, implying the independence of the increments Z_e. In combination with the identity for G in (3.14), setting f to be the indicator function of the set $\{y : \max_{v\in V} x_v y_v > 1\}$ in (3.15) yields (3.16) via

$$\begin{aligned} -\log G(1/x_v, v \in V) = \int f \,\mathrm{d}\mu &= \mathbb{E}\left[\int_0^\infty \mathbb{1}\left\{\max_{v\in V} x_v A_{uv} > z^{-1}\right\} z^{-2}\,\mathrm{d}z\right] \\ &= \mathbb{E}\left[\max\{x_v A_{uv}, v \in V\}\right]. \end{aligned}$$ □

3.3.4 Special case: max-stable clique vectors

The limit distribution in Theorem 3.3.5 is determined by the graph structure and the clique-wise limit distributions $\nu_{C,s}$ in Assumption 3.3.1 via Definition 3.3.2. The next result provides those limits $\nu_{C,s}$ in the special case that X_C follows a max-stable distribution. It is a reformulation of Example 8.4 in Heffernan and Tawn (2004) in terms of the stable tail dependence function ℓ, allowing for the limit distribution ν_1 to have margins with positive mass at the origin; take for instance $d = 2$ and $\ell(x_1, x_2) = x_1 + x_2$. Since the result is not related to graphical models, we cast it in terms of a random vector $(X_1, \ldots, X_d)$.

Proposition 3.3.11. *Let $X = (X_1, \ldots, X_d)$ have a max-stable distribution G with unit-Fréchet margins and stable tail dependence function ℓ. If ℓ has a continuous first-order partial derivative $\dot{\ell}_1$ with respect to its first argument, then*

$$\left(\frac{X_j}{t},\ j \in \{2,\ldots,d\} \,\middle|\, X_1 = t\right) \xrightarrow{d} \nu_1, \qquad t \to \infty,$$

where ν_1 is a probability distribution with support contained in $[0,\infty)^d$ and determined by

$$\forall x \in (0,\infty)^{d-1}, \qquad \nu_1([0,x]) = \dot{\ell}_1(1, 1/x_2, \ldots, 1/x_d). \tag{3.17}$$

The following example will play a key role in the next section. It provides the form of the probability measure ν_1 in the special case the clique vectors X_C follow a max-stable Hüsler–Reiss distribution.

Example 3.3.12. *Suppose $X = (X_1, \ldots, X_d)$ follows a max-stable Hüsler–Reiss distribution with unit-Fréchet margins and parameter matrix $\Delta = \{\delta_{ij}^2\}$; see Section 3.4.1 below for a more detailed description. The corresponding stable tail dependence function ℓ has a continuous first-order partial derivatives, so that $X/t \mid X_1 = t$ converges weakly to some distribution ν_1 determined by ℓ via* (3.17)*. By Remark 3.3.8, the limit must be the same as the one of $(X/X_1 \mid X_1 > t)$ as $t \to \infty$. But Theorem 2 in Engelke et al. (2014) states that, as $t \to \infty$,*

$$\big(\ln X_j - \ln X_1,\ j = 2,\ldots,d \mid X_1 > t\big) \xrightarrow{d} \mathcal{N}_{d-1}\big(\mu_1(\Delta), \Psi_1(\Delta)\big),$$

for a mean vector $\mu_1 = -2(\delta_{1i}^2, i = 2,\ldots,d)$ and covariance matrix

$$(\Psi_1)_{ij} = 2(\delta_{1i}^2 + \delta_{1j}^2 - \delta_{ij}^2), \qquad i,j = 2,\ldots,d.$$

Hence $X/X_1 \mid X_1 > t$ and $X/X_1 \mid X_1 = t$ converge to a multivariate log-normal distribution with the same parameters, $\mu_1(\Delta)$ and $\Psi_1(\Delta)$. The limit is confirmed in Segers (2020b, Example 4) when X has only two elements. ◊

3.4 Cliques in the Hüsler–Reiss domain of attraction

We will apply Theorem 3.3.5 to a Markov random field X with respect to a block graph $\mathcal{G} = (V, E)$ such that for every (maximal) clique $C \in \mathcal{C}$, the sub-vector $X_C = (X_v, v \in C)$ satisfies Assumption 3.3.1 with $\nu_{C,s}$ being the one as in Example 3.3.12, i.e., a multivariate log-normal distribution with mean $\mu_{C,s}(\Delta_C)$ and covariance matrix $\Psi_{C,s}(\Delta_C)$.

In Proposition 3.4.2, we find that the limit random vector A_u in Theorem 3.3.5 is multivariate log-normal with mean vector and covariance matrix related to the graph structure. Moreover, X is in the max-domain of attraction of a max-stable Hüsler–Reiss whose parameter matrix can be derived in a simple way from the matrices Δ_C (Proposition 3.4.4). Further, the associated multivariate Pareto distribution is an extremal graphical model in the sense of Engelke and Hitz (2020) and this with respect to the same block graph $\mathcal{G}$ (Proposition 3.4.5). The elegant form of the parameter matrix makes this family a suitable candidate for modelling extremes of asymptotically dependent distributions; see also the discussion in Strokorb (2020) to Engelke and Hitz (2020) on the issue of extremal independence and disconnected graphs. Finally, we show that the parameters of the limiting max-stable Hüsler–Reiss distribution are still identifiable in case some variables are latent, and this if and only if every node with a latent variable belongs to at least three different cliques (Proposition 3.4.6).

3.4.1 Max-stable Hüsler–Reiss distribution

The max-stable Hüsler–Reiss distribution arises as the limiting distribution of normalized component-wise sample maxima of a triangular array of row-wise independent and identically distributed Gaussian random vectors with correlation matrix that depends on the sample size (Hüsler and Reiss, 1989). The Gaussian distribution is in the max-domain of attraction of the Gumbel distribution, but here we transform the margins to the unit-Fréchet distribution. Let Φ denote the standard normal cumulative distribution function. Recall from (3.4) the stdf ℓ of a general max-stable distribution G.

The stdf of the bivariate Hüsler–Reiss distribution with parameter $\delta \in (0, \infty)$ is

$$\ell_\delta(x, y) = x\,\Phi\left(\delta + \frac{\ln(x/y)}{2\delta}\right) + y\,\Phi\left(\delta + \frac{\ln(y/x)}{2\delta}\right), \qquad (x, y) \in (0, \infty)^2, \tag{3.18}$$

with obvious limits as $x \to 0$ or $y \to 0$. The boundary cases $\delta \to \infty$ and $\delta \to 0$ correspond to independence, $\ell_\infty(x, y) = x + y$, and co-monotonicity, $\ell_0(x, y) = \max(x, y)$, respectively. The limit distribution in Proposition 3.3.11 can be calculated explicitly and is equal to the one of the log-normal random variable $\exp\{2\delta(Z - \delta)\}$, with Z a standard normal random variable (Segers, 2020b, Example 4).

To introduce the multivariate Hüsler–Reiss distribution, we follow the exposition in Engelke et al. (2014). Let W be a finite set with at least two elements and let $\rho(1), \rho(2), \ldots$ be a sequence of W-variate correlation matrices, i.e., $\rho(n) = (\rho_{ij}(n))_{i,j\in W}$. Assume the limit matrix $\Delta = (\delta_{ij}^2)_{i,j\in W}$ – denoted by Λ in the cited article – exists:

$$\lim_{n\to\infty} \big(1 - \rho_{ij}(n)\big) \ln(n) = \delta_{ij}^2, \qquad i, j \in W. \tag{3.19}$$

Obviously, the matrix $\Delta \in [0, \infty)^{W\times W}$ is symmetric and has zero diagonal. Suppose further that Δ is conditionally negative definite, i.e., we have $a^\top \Delta a < 0$ for every

non-zero vector $a \in \mathbb{R}^W$ such that $\sum_{j\in W} a_j = 0$. [Note that the weak inequality $a^\top \Delta a \le 0$ automatically holds for such a and for the limit Δ in (3.19).] For $J \subseteq W$ with $|J| \ge 2$ and for $s \in J$ let $\Psi_{J,s}$ be the positive definite, $|J \setminus s|$-square symmetric matrix with elements

$$\big(\Psi_{J,s}(\Delta)\big)_{i,j} = 2(\delta_{si}^2 + \delta_{sj}^2 - \delta_{ij}^2), \qquad i, j \in J \setminus s. \tag{3.20}$$

The $|W|$-variate Hüsler–Reiss max-stable distribution with unit-Fréchet margins and parameter matrix Δ is

$$H_\Delta(x) = \exp\left\{ \sum_{j=1}^{|W|} (-1)^j \sum_{J\subseteq W : |J|=j} h_{\Delta,J}(x_J) \right\}, \qquad x \in (0,\infty)^W, \tag{3.21}$$

with $h_{\Delta,J}(x_J) = 1/x_w$ if $J = \{w\}$, while, if $|J| \ge 2$,

$$h_{\Delta,J}(x_J) = \int_{\ln(x_s)}^{\infty} \mathbb{P}\left[\forall w \in J \setminus s,\ Y_{sw} > \ln(x_w) - z + 2\delta_{sw}^2\right] e^{-z}\, \mathrm{d}z$$

where s can be any element of J and where $Y_s = (Y_{sw}, w \in J \setminus s)$ is a multivariate normal random vector with zero mean vector and covariance matrix $\Psi_{J,s}(\Delta)$ in (3.20).

A shorter expression for H_Δ is given in Nikoloulopoulos et al. (2009, Remark 2.5), later confirmed as the finite-dimensional distributions of max-stable Gaussian and Brown-Resnick processes in Genton et al. (2011) and Huser and Davison (2013) respectively:

$$H_\Delta(x) = \exp\left\{ -\sum_{s\in W} \frac{1}{x_s} \Phi_{|W|-1}\left(2\delta_{vs}^2 + \ln(x_v/x_s), v \in W \setminus s; \Psi_{W,s}(\Delta)\right) \right\},$$

for $x \in (0,\infty)^W$, and $\Phi_d(\,\cdot\,;\Sigma)$ the d-variate normal cdf with covariance matrix Σ. The stdf is thus

$$\ell_\Delta(y) = \sum_{s\in W} y_s\, \Phi_{|W|-1}\left(2\delta_{vs}^2 + \ln(y_s/y_v), v \in W \setminus s; \Psi_{W,s}(\Delta)\right).$$

for $y \in (0,\infty)^W$. If $|W| = 2$ and if the off-diagonal element of Δ is $\delta^2 \in (0,\infty)$, say, we have $\Psi_{W,s}(\Delta) = 4\delta^2 = (2\delta)^2$ and the stdf ℓ_Δ indeed simplifies to ℓ_δ in (3.18).

3.4.2 Hüsler–Reiss limits and extremal graphical models

Recall from Example 3.3.12 that for a max-stable Hüsler–Reiss vector $X = (X_1, \ldots, X_d)$, the limit of X/t given $X_1 = t$ as $t \to \infty$ is multivariate log-normal. Now we take that limit as starting point for the tails of the clique vectors of a Markov random field on a block graph.

Assumption 3.4.1 (Markov block graph with clique-wise Hüsler–Reiss limits)**.** *Let X be a Markov random field with respect to the (connected) block graph $\mathcal{G} = (V, E)$ with (maximal) cliques $\mathcal{C}$. Suppose the margins of X satisfy $t\,\mathbb{P}(X_v > t) \to 1$ as $t \to \infty$ for all $v \in V$. For every clique $C \in \mathcal{C}$, let $\Delta_C = (\delta_{ij}^2)_{i,j\in C}$ be the parameter matrix of a $|C|$-variate max-stable Hüsler–Reiss distribution, i.e., $\Delta_C \in [0,\infty)^{C\times C}$ is symmetric, conditionally negative definite, and has zero diagonal. For every $C \in \mathcal{C}$ let X_C satisfy Assumption 3.3.1 where $\nu_{C,u}$ is the limit in Example 3.3.12, i.e., a $|C\setminus u|$-variate log-normal distribution with mean vector*

$$\mu_{C,u} = -2(\delta_{ui}^2, i \in C \setminus u)$$

and covariance matrix

$$(\Psi_{C,u})_{ij} = 2(\delta_{ui}^2 + \delta_{uj}^2 - \delta_{ij}^2), \qquad i, j \in C \setminus u.$$

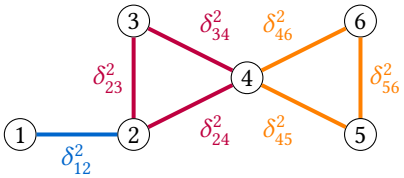


Figure 3.3: The random vector X is six-variate and it is Markov with respect to the graph which contains three cliques, $C_1 = \{1, 2\}, C_2 = \{2, 3, 4\}$ and $C_3 = \{4, 5, 6\}$. According to Example 3.3.12 for C_1 we have a limiting probability measure $\nu_{C_1,\cdot}$ which depends only on the parameter δ_{12}; for C_2 we have $\nu_{C_2,\cdot}$ which depends only on the parameters $\delta^2_{23}, \delta^2_{24}, \delta^2_{34}$ and for C_3 we have $\nu_{C_3,\cdot}$ which depends only on the parameters $\delta^2_{45}, \delta^2_{46}, \delta^2_{56}$.

Because for any $u \in C$ the random vector X_C satisfies the limit in Assumption 3.3.1 by Corollary 1 and Theorem 2 in Segers (2020b) it follows that it is in the max-stable domain of attraction of a Hüsler–Reiss distribution with parameter matrix Δ_C.

We apply Theorem 3.3.5 to study the limit of $X/t \mid X_u = t$ for some $u \in V$, i.e., the conditional distribution of the field given that it is large at a particular node. Write $\Delta = (\Delta_C, C \in \mathcal{C})$ and consider the matrix $P(\Delta) = \big(p_{ij}(\Delta)\big)_{i,j\in V}$ of path sums

$$p_{ij}(\Delta) := \sum_{e\in(i\rightsquigarrow j)} \delta^2_e, \tag{3.22}$$

where $(i \rightsquigarrow j)$ is the collection of edges on the unique shortest path from i to j and where δ^2_e is to be read off from the matrix Δ_C for the unique clique C containing the two nodes connected by e; by convention, $p_{ii}(\Delta) = 0$ for all $i \in V$, being the sum over the empty set $(i \rightsquigarrow i) = \emptyset$. Let $\mathcal{N}_r(\mu, \Sigma)$ denote the r-variate normal distribution with mean vector μ and covariance matrix Σ.

Proposition 3.4.2 (Logarithm of the limiting field)**.** *Under Assumption 3.4.1, we have, for each $u \in V$ and as $t \to \infty$,*

$$\big(\ln(X_v/t),\, v \in V \setminus u \mid X_u = t\big) \xrightarrow{d} \mathcal{N}_{|V\setminus u|}\big(\mu_u(\Delta), \Sigma_u(\Delta)\big)$$

with mean vector and covariance matrix written in terms of $p_{ij} = p_{ij}(\Delta)$ in (3.22) by

$$\big(\mu_u(\Delta)\big)_i = -2p_{ui}, \qquad i \in V \setminus u, \tag{3.23}$$
$$\big(\Sigma_u(\Delta)\big)_{i,j} = 2(p_{ui} + p_{uj} - p_{ij}), \qquad i, j \in V \setminus u, \tag{3.24}$$

and in particular $(\Sigma_u(\Delta))_{i,i} = 4p_{ui}$ for $i \in V \setminus u$. The matrix $\Sigma_u(\Delta)$ is positive definite and the matrix $P(\Delta)$ is conditionally negative definite.

Example 3.4.3. *Consider a Markov field with respect to the block graph in Figure 3.3. The graph has three cliques, to which correspond three Hüsler–Reiss limits with parameter matrices respectively*

$$\Delta_1 = \begin{bmatrix} 0 & \delta^2_{12} \\ & 0 \end{bmatrix}, \quad \Delta_2 = \begin{bmatrix} 0 & \delta^2_{23} & \delta^2_{24} \\ & 0 & \delta^2_{34} \\ & & 0 \end{bmatrix}, \quad \Delta_3 = \begin{bmatrix} 0 & \delta^2_{45} & \delta^2_{46} \\ & 0 & \delta^2_{56} \\ & & 0 \end{bmatrix}.$$

If a high threshold is exceeded at node $u = 1$, the limiting 5-variate normal distribution in Proposition 3.4.2 has means $(\mu_1(\Delta))_i$ and variances $(\Sigma_1(\Delta))_{ii}$ proportional to the path sums $p_{1i} = \sum_{e\in(1\rightsquigarrow i)} \delta^2_e$ for $i \in \{2, \ldots, 5\}$, while the off-diagonal entries of the covariance matrix are given by

$$\big(\Sigma_1(\Delta)\big)_{2,j} = 4\delta^2_{12}, \qquad j \in \{3, \ldots, 6\},$$

$$
\begin{aligned}
\big(\Sigma_1(\Delta)\big)_{3,j} &= 4\left(\delta_{12}^2 + \tfrac{1}{2}(\delta_{23}^2 + \delta_{24}^2 - \delta_{34}^2)\right), && j \in \{4,5,6\},\\
\big(\Sigma_1(\Delta)\big)_{4,j} &= 4(\delta_{12}^2 + \delta_{24}^2), && j \in \{5,6\},\\
\big(\Sigma_1(\Delta)\big)_{5,6} &= 4\left(\delta_{12}^2 + \delta_{24}^2 + \tfrac{1}{2}(\delta_{45}^2 + \delta_{46}^2 - \delta_{56}^2)\right).
\end{aligned}
$$

The dependence within blocks is visible in the covariances at entries $(3, j)$ *for* $j \in \{4, 5, 6\}$ *and the one at entry* $(5, 6)$. ◇

The exact distribution of the Markov field with clique-wise Hüsler–Reiss limits in Assumption 3.4.1 is in the max-domain of attraction of such a distribution.

Proposition 3.4.4 (Max-domain of attraction)**.** *The Markov random field* X *in Assumption 3.4.1 is in the max-domain of attraction of the Hüsler–Reiss max-stable distribution* (3.21) *with unit-Fréchet margins and parameter matrix* $P(\Delta)$ *in* (3.22)*, that is,*

$$\lim_{n\to\infty} \big(\mathbb{P}\left(\forall v \in V : X_v \le n x_v\right)\big)^n = H_{P(\Delta)}(x), \qquad x \in (0,\infty)^V.$$

Recall from Section 3.2.2 that because X in Proposition 3.4.4 belongs to the domain of attraction of the max-stable distribution $H_{P(\Delta)}$, the asymptotic distribution of the vector of high-threshold excesses is a multivariate Pareto distribution determined by $H_{P(\Delta)}$ via (3.6). The latter is the distribution of the random vector Y in the next proposition and is called a Hüsler–Reiss Pareto distribution in Engelke and Hitz (2020). The distribution of Y turns out to be an extremal graphical model in the sense of Engelke and Hitz (2020, Definitions 1 and 2). We recall this notion here. Let Y be a multivariate Pareto random vector in (3.7) and for $u \in V$, let $Y^{(u)}$ be a random vector equal in distribution to $Y \mid Y_u > 1$. Then Y is an extremal graphical model with respect to a graph $\mathcal{G} = (V, E)$ if we have conditional independence $Y_i^{(u)} \perp\!\!\!\perp Y_j^{(u)} \mid Y_{V\setminus\{u,i,j\}}^{(u)}$ for all $i, j \in V \setminus u$ such that $(i, j) \notin E$.

Proposition 3.4.5 (Attraction to extremal graphical model)**.** *The Markov block graph* X *in Assumption 3.4.1 satisfies the weak convergence relation* (3.7) *with* Y *distributed as in* (3.6) *for* $G = H_{P(\Delta)}$*, the limit in Proposition 3.4.4. This* Y *is an extremal graphical model with respect to* $\mathcal{G}$ *in the sense of Engelke and Hitz (2020, Definition 2).*

Proposition 3.4.5 leads to the elegant result that the graphical model X obtained by endowing every clique C of a block graph $\mathcal{G}$ by a limit based on the Hüsler–Reiss max-stable distribution with parameter matrix Δ_C is in the Pareto domain of attraction of a Hüsler–Reiss Pareto random vector Y which is itself an extremal graphical model with respect to the same graph $\mathcal{G}$ and with, on every clique C, a Hüsler–Reiss Pareto distribution with the same parameter matrix Δ_C. In other words, the Pareto limit of a graphical model constructed clique-wise by distributions with Hüsler–Reiss limits is an extremal graphical model constructed clique-wise by Hüsler–Reiss Pareto distributions.

Proposition 3.4.5 also sheds new light on Proposition 4 in Engelke and Hitz (2020), where the existence and uniqueness of a Hüsler–Reiss extremal graphical model was established given the Hüsler–Reiss distributions on the cliques of a block graph. In our construction, the solution is explicit and turns out to have the simple and elegant form in terms of the path sums $p_{ij}(\Delta)$ in (3.22).

3.4.3 Latent variables and parameter identifiability

In Asenova, Mazo, and Segers (2021) a criterion was presented for checking whether the parameters of the Hüsler–Reiss distribution are identifiable if for some of the nodes $v \in V$ the variables X_v are unobservable (latent). The issue was illustrated for river networks when the water level or another variable of interest is not observed at

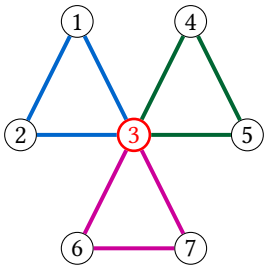


Figure 3.4: A block graph with three cliques. In the first clique with node set $C_1 = \{1, 2, 3\}$ the parameters are $\delta^2_{12}, \delta^2_{13}, \delta^2_{23}$, in the second clique $C_1 = \{3, 4, 5\}$ the parameters are $\delta^2_{34}, \delta^2_{35}, \delta^2_{45}$, and in the third clique $C_1 = \{3, 6, 7\}$ the parameters are $\delta^2_{36}, \delta^2_{37}, \delta^2_{67}$. These nine parameters determine the Hüsler–Reiss parameter matrix $P(\Delta)$ in (3.22). The nine parameters and thus the entire matrix $P(\Delta)$ is identifiable from the submatrix $P(\Delta)_U$ with $U = V \setminus \bar{U}$ for $\bar{U} = \{3\}$ because node 3 belongs to three different cliques (Proposition 3.4.6 and Example 3.4.7).

some splits or junctions. For trees, a necessary and sufficient identifiability criterion was that every node with a latent variable should have degree at least three.

For block graphs, a similar condition turns out to hold. The degree of a node $v \in V$, i.e., the number of neighbours, is now replaced by its *clique degree*, notation $\mathrm{cd}(v)$, defined as the number of cliques containing that node.

Let the setting be the same as in Proposition 3.4.4 and let $H_{P(\Delta)}(x)$ be the $|V|$-variate max-stable Hüsler–Reiss distribution with parameter matrix $P(\Delta)$ in (3.22). Let the (non-empty) set of nodes with observable variables be $U \subset V$, so that $\bar{U} = V \setminus U$ is the set of nodes with latent variables. As the max-stable Hüsler–Reiss family is stable under taking marginals (Engelke and Hitz, 2020, Example 7), the vector $X_U = (X_v, v \in U)$ is in the max-domain of attraction of the $|U|$-variate max-stable Hüsler–Reiss distribution with parameter matrix $P(\Delta)_U = (p_{ij}(\Delta))_{i,j\in U}$. If $\bar{U}$ is non-empty, U is a proper subset of V, and the question is whether we can reconstruct the whole matrix $P(\Delta)$ given only the sub-matrix $P(\Delta)_U$ and the graph $\mathcal{G}$. Note that the entries in $P(\Delta)_U$ are the path sums between nodes carrying observable variables only. The question is whether we can find the other path sums too, that is, those between nodes one or two of which carry latent variables.

Proposition 3.4.6 (Identifiability). *Given the block graph $\mathcal{G} = (V, E)$ and node sets $\bar{U} \subset V$ and $U = V \setminus \bar{U}$, the Hüsler–Reiss parameter matrix $P(\Delta)$ in (3.22) is identifiable from the restricted matrix $P(\Delta)_U = (p_{ij}(\Delta))_{i,j\in U}$ if and only if $\mathrm{cd}(v) \geq 3$ for every $v \in \bar{U}$.*

Example 3.4.7. *Consider the block graph $\mathcal{G} = (V, E)$ in Figure 3.4 with $V = \{1, \ldots, 7\}$ and $\bar{U} = \{3\}$. Therefore $U = V \setminus \{3\}$. Node $v = 3$ belongs to three different cliques and thus has clique degree $\mathrm{cd}(v) = 3$. By Proposition 3.4.6, all edge parameters δ^2_e for $e \in E$ can be identified from the path sums p_{ij} for $i, j \in U$. Indeed, for the edges $e = (a, b) \in \{(1, 2), (4, 5), (6, 7)\}$ this follows from the identity $\delta^2_e = p_{ab}$, while for the edges $\delta^2_{i,3}$ for $i \neq 3$ this follows from a calculation such as*

$$\delta^2_{13} = \frac{1}{2}\left(p_{14} + p_{16} - p_{46}\right).$$

If, however, node $v = 1$ would not belong to U, then the edge parameters δ^2_{12} and δ^2_{13} would not be identifiable from the path sums p_{ij} for $i, j \in V \setminus \{1\}$, since none of these paths contains edges $(1, 2)$ or $(1, 3)$.

3.5 Conclusion

We have studied the tails of suitably normalized random vectors which satisfy the global Markov property with respect to a block graph. Block graphs are generalizations of trees and this explains why the results presented here are closely related to the ones in Segers (2020b) and Asenova, Mazo, and Segers (2021). The common feature is the existence of a unique shortest path between each pair of nodes. This property is key for our results, although it is not sufficient in itself to explain the multiplicative random walk structure of the limiting field. The latter property is also due to the singleton nature of the minimal clique separators, a property which is no longer present for more general decomposable graphs. Still, an essential difference between tails of Markov fields with respect to trees on the one hand and more general block graphs on the other hand is that in the latter case, the increments of the random walk in the tail field are dependent within cliques. The regularity assumptions needed for the limit to hold imply multivariate regular variation and thus the model in question is suitable for asymptotically dependent variables.

We have then focused on a particular random field with respect to a block graph, namely one for which the distribution on each clique satisfies a tail condition based on a Hüsler–Reiss distribution. We have shown that the logarithm of the limiting field is a normal random vector with mean and covariance matrix that depend on the sums of the edge weights along the unique shortest paths between pairs of nodes. The same structural pattern emerges in the parameter matrix of the max-stable Hüsler–Reiss distribution to which the Markov field is attracted. The relation between the original Markov field as an ordinary graphical model on the one hand and Hüsler–Reiss extremal graphical models as in Engelke and Hitz (2020) on the other hand was highlighted. Due to the path sum structure of the parameter matrix, all edge weights remain identifiable even when variables associated to nodes with clique degree at least three are latent.

An interesting problem would be to identify a minimal requirement on a graph that leads to the multiplicative structure of the tail field of a Markov field that we found for block graphs. Another question is which structure replaces the multiplicative random walk form for more general graphs, for instance decomposable graphs, and what this means for specific parametric families. Another research direction could be the study of the tails of the Markov field under assumptions related to hidden regular variation and/or allowing for transitions from non-extreme to extreme regions.

3.6 Supplement

3.6.1 Proof of Theorem 3.3.5

The proof follows the lines of the one of Theorem 1 in Segers (2020b). To show (3.10) it is sufficient to show that for a real bounded Lipschitz function f, for any fixed $u \in V$ it holds that

$$\lim_{t\to\infty} \mathbb{E}[f(X_{V\setminus u}/t) \mid X_u = t] = \mathbb{E}[f(A_{u,V\setminus u})], \tag{3.25}$$

(van der Vaart, 1998, Lemma 2.2). Without loss of generality, we assume that $0 \leq f(x) \leq 1$ and $|f(x) - f(y)| \leq L \sum_j |x_j - y_j|$ for some constant $L > 0$.

We proceed by induction on the number of cliques, m. When there is only one clique ($m = 1$) the convergence happens by Assumption 3.3.1 with $s = u$: the distribution of $A_{u,V\setminus u}$ is equal to $\nu_{C,s}$ in the assumption, with $C = V$ and $u = s$.

Assume that there are at least two cliques, $m \geq 2$. Let the numbering of the cliques be such that the last clique, C_m, is connected to the subgraph induced by $\bigcup_{i=1}^{m-1} C_i$ only through one node, which is the minimal separator between C_m and $\bigcup_{i=1}^{m-1} C_i$. Let

$s \in \mathcal{S}$ denote this node and introduce the set

$$C_{1:m-1} = (C_1 \cup \cdots \cup C_{m-1}) \setminus u.$$

Note that $\{s\} = (C_1 \cup \cdots \cup C_{m-1}) \cap C_m$. We need to make a distinction between two cases: $s = u$ or $s \neq u$. The case $s = u$ is the easier one, since then $X_{C_m \setminus u}$ and $X_{C_{1:m-1}}$ are conditionally independent given X_u whereas the shortest paths from u to nodes in $C_m \setminus u$ just consist of single edges, avoiding $C_{1:m-1}$ altogether. So we only consider the case $s \neq u$ henceforth. In that case, paths from u to nodes in $C_m \setminus s$ pass through s and possibly other nodes in $C_{1:m-1}$.

The induction hypothesis is that as $t \to \infty$, we have

$$(X_{C_{1:m-1}}/t) \mid X_u = t \xrightarrow{d} A_{u,C_{1:m-1}}, \tag{3.26}$$

or also that for every continuous and bounded function $h : \mathbb{R}_+^{C_{1:m-1}} \to \mathbb{R}$, we have

$$\lim_{t\to\infty} \mathbb{E}\left[h(X_{C_{1:m-1}}/t) \mid X_u = t\right] = \mathbb{E}[h(A_{u,C_{1:m-1}})]. \tag{3.27}$$

To prove the convergence in (3.25) we start with the following inequality: for $\delta > 0$,

$$\begin{aligned}
&\left|\mathbb{E}[f(X_{V\setminus u}/t) \mid X_u = t] - \mathbb{E}[f(A_{u,V\setminus u})]\right| \\
&\leq \left|\mathbb{E}\left[f(X_{V\setminus u}/t)\,\mathbb{1}(X_s/t \geq \delta) \mid X_u = t\right] - \mathbb{E}\left[f(A_{u,V\setminus u})\,\mathbb{1}(A_{us} \geq \delta)\right]\right| \qquad (3.28)\\
&+ \left|\mathbb{E}\left[f(X_{V\setminus u}/t)\,\mathbb{1}(X_s/t < \delta) \mid X_u = t\right] - \mathbb{E}\left[f(A_{u,V\setminus u})\,\mathbb{1}(A_{us} < \delta)\right]\right|. \qquad (3.29)
\end{aligned}$$

We let $\delta > 0$ be a continuity point of A_{us}. Later on, we will take δ arbitrarily close to zero, which we can do, since the number of atoms of A_{us} is at most countable.

Analysis of (3.28). We first deal with (3.28). The first expectation is equal to

$$\int_{[0,\infty)^{V\setminus u}} f(x/t)\,\mathbb{1}(x_s/t \geq \delta)\;\mathbb{P}(X_{V\setminus u} \in \mathrm{d}x \mid X_u = t)\,.$$

Because of the global Markov property, $X_{C_m\setminus s}$ is conditionally independent of the variables in the set $C_{1:m-1}$ given X_s. As a consequence, the conditional distribution of $X_{C_m\setminus s}$ given $X_{C_{1:m-1}}$ is the same as the one of $X_{C_m\setminus s}$ given X_s. Hence we can write the integral as

$$\begin{aligned}
\int_{[0,\infty)^{C_{1:m-1}}} &\mathbb{E}\left[f\left(x/t, X_{C_m\setminus s}/t\right) \mid X_s = x_s\right] \\
&\mathbb{1}(x_s/t \geq \delta)\,\mathbb{P}(X_{C_{1:m-1}} \in \mathrm{d}x \mid X_u = t).
\end{aligned}$$

After the change of variables $x/t = y$, the integral becomes

$$\begin{aligned}
\int_{[0,\infty)^{C_{1:m-1}}} &\mathbb{E}\left[f\left(y, X_{C_m\setminus s}/t\right) \,\middle|\, X_s = ty_s\right] \\
&\mathbb{1}(y_s \geq \delta)\,\mathbb{P}\left(X_{C_{1:m-1}}/t \in \mathrm{d}y \mid X_u = t\right).
\end{aligned} \tag{3.30}$$

Define the functions g_t and g on $[0,\infty)^{C_{1:m-1}}$ by

$$\begin{aligned}
g_t(y) &:= \mathbb{E}\left[f\left(y, X_{C_m\setminus s}/t\right) \,\middle|\, X_s = ty_s\right] \mathbb{1}(y_s \geq \delta),\\
g(y) &:= \mathbb{E}\left[f\left(y, y_s Z_{s,C_m\setminus s}\right)\right] \mathbb{1}(y_s \geq \delta).
\end{aligned}$$

Consider points $y(t)$ and y in $[0,\infty)^{C_{1:m-1}}$ such that $\lim_{t\to\infty} y(t) = y$ and such that $y_s \neq \delta$. We need to show that

$$\lim_{t\to\infty} g_t(y(t)) = g(y). \tag{3.31}$$

If $y_s < \delta$, this is clear since $y_s(t) < \delta$ for all large t and hence the indicators will be zero. So suppose $y_s > \delta$ and thus also $y_s(t) > \delta$ for all large t, meaning that both indicators are (eventually) equal to one. By Assumption 3.3.1, we have

$$X_{C_m\setminus s}/t \mid X_s = t y_s(t) \xrightarrow{d} y_s Z_{s,C_m\setminus s}, \qquad t\to\infty.$$

Since f is continuous, also

$$f\left(y(t), X_{C_m\setminus s}/t\right) \,\Big|\, X_s = t y_s(t) \xrightarrow{d} f\left(y, y_s Z_{s,C_m\setminus s}\right), \qquad t\to\infty.$$

As the range of f is contained in $[0,1]$, the bounded convergence theorem implies that we can take expectations in the previous equation and conclude (3.31).

By the induction hypothesis (3.26) and Theorem 18.11 in van der Vaart (1998), the continuous convergence in (3.31) implies

$$g_t\left(\frac{X_{C_{1:m-1}}}{t}\right) \,\Big|\, X_u = t \xrightarrow{d} g(A_{C_{1:m-1}}), \qquad t\to\infty;$$

note that by the choice of δ, the discontinuity set of g receives zero probability in the limit. As g_t and g are bounded (since f is bounded), we can take expectations and find

$$\lim_{t\to\infty} \mathbb{E}\left[g_t\left(\frac{X_{C_{1:m-1}}}{t}\right) \,\Big|\, X_u = t\right] = \mathbb{E}[g(A_{C_{1:m-1}})]. \tag{3.32}$$

The expectation on the left-hand side of (3.32) is the integral in (3.30) while the right-hand side of (3.32) is equal to

$$\mathbb{E}[f(A_{u,C_{1:m-1}}, A_{us} Z_{s,C_m\setminus s})\,\mathbb{1}(A_{us}\geq\delta)] = \mathbb{E}[f(A_{u,V\setminus u})\,\mathbb{1}(A_{us}\geq\delta)].$$

Thus we have shown that (3.28) converges to 0 as $t\to\infty$, for any continuity point δ of A_{us}.

Analysis of (3.29). As f is a function with range $[0,1]$ we have

$$0 \leq \mathbb{E}[f(X_{V\setminus u}/t)\,\mathbb{1}(X_s/t<\delta) \mid X_u = t] \leq \mathbb{P}[X_s/t<\delta \mid X_u = t]$$

as well as

$$0 \leq \mathbb{E}[f(A_{u,V\setminus u})\,\mathbb{1}(A_{us}<\delta)] \leq \mathbb{P}[A_{us}<\delta].$$

By the triangle inequality and the two inequalities above, (3.29) is bounded from above by

$$\mathbb{P}[X_s/t<\delta \mid X_u = t] + \mathbb{P}[A_{us}<\delta]. \tag{3.33}$$

By the induction hypothesis

$$\lim_{t\to\infty} \mathbb{P}[X_s/t<\delta \mid X_u = t] = \mathbb{P}[A_{us}<\delta],$$

and (3.33) converges to $2\,\mathbb{P}[A_{us}<\delta]$, which goes to 0 as $\delta\downarrow 0$ in case $\mathbb{P}(A_{us}=0)=0$.

Suppose $\mathbb{P}(A_{us}=0)>0$. In this step we will need Assumption 3.3.4. By the induction hypothesis, we have $A_{us} = \prod_{(a,b)\in(u\rightsquigarrow s)} Z_{ab}$ and the variables Z_{ab} are independent. Hence

$$\mathbb{P}(A_{us}=0) = \mathbb{P}\Big(\min_{(a,b)\in(u\rightsquigarrow s)} Z_{ab} = 0\Big) = 1 - \prod_{(a,b)\in(u\rightsquigarrow s)} \mathbb{P}(Z_{ab}>0).$$

If for any $(a,b) \in (u \rightsquigarrow s)$ we have $\mathbb{P}(Z_{ab} = 0) > 0$ then $P(Z_{ab} > 0) < 1$ and hence $\mathbb{P}(A_{us} = 0) > 0$. Therefore the assumption applies when the marginal distribution $\nu^b_{C,a}(\{0\})$ is positive.

Then by adding and subtracting terms and using the triangle inequality, we have the following upper bound for the term in (3.29):

$$\begin{aligned}
&\left|\mathbb{E}\left[f\left(\frac{X_{C_{1:m-1}}}{t}, \frac{X_{C_m \setminus s}}{t}\right) \mathbb{1}\left(\frac{X_s}{t} < \delta\right) \middle| X_u = t\right]\right. \\
&\left. - \mathbb{E}\left[f\left(\frac{X_{C_{1:m-1}}}{t}, 0\right) \mathbb{1}\left(\frac{X_s}{t} < \delta\right) \middle| X_u = t\right]\right| && (3.34) \\
&+ \left|\mathbb{E}\left[f(A_{u,C_{1:m-1}}, A_{u,C_m \setminus s}) \mathbb{1}(A_{u,s} < \delta)\right] - \mathbb{E}\left[f(A_{u,C_{1:m-1}}, 0) \mathbb{1}(A_{u,s} < \delta)\right]\right| && (3.35) \\
&+ \left|\mathbb{E}\left[f\left(\frac{X_{C_{1:m-1}}}{t}, 0\right) \mathbb{1}\left(\frac{X_s}{t} < \delta\right) \middle| X_u = t\right]\right. \\
&\left. - \mathbb{E}\left[f(A_{u,C_{1:m-1}}, 0) \mathbb{1}(A_{u,s} < \delta)\right]\right|. && (3.36)
\end{aligned}$$

We treat each of the three terms in turn.

Equation (3.36) converges to 0 by the induction hypothesis; note again that the set of discontinuities of the integrand receives zero probability in the limit.

Next we look at expression (3.34). From the assumptions of f, namely that it ranges in $[0,1]$ and that $|f(x) - f(y)| \leq L\|x - y\|_1$ for some constant $L > 0$, where $\|z\|_1 = \sum_j |z_j|$ for a Euclidean vector z, the term in (3.34) is bounded by

$$\begin{aligned}
&\mathbb{E}\left[\left|f\left(\frac{X_{C_{1:m-1}}}{t}, \frac{X_{C_m \setminus s}}{t}\right) - f\left(\frac{X_{C_{1:m-1}}}{t}, 0\right)\right| \mathbb{1}\left(\frac{X_s}{t} < \delta\right) \middle| X_u = t\right] \\
&\qquad \leq \mathbb{E}\left[\mathbb{1}\left(X_s/t < \delta\right) \min\left(1, L\|X_{C_m \setminus s}/t\|_1\right) \mid X_u = t\right].
\end{aligned}$$

We need to show that the upper bound converges to 0 as $t \to \infty$. Because the variables in $C_m \setminus s$ are independent of X_u conditionally on X_s, the previous integral is equal to

$$\int_{[0,\delta]} \mathbb{E}\left[\min\left(1, L\|X_{C_m \setminus s}/t\|_1\right) \mid X_s/t = x_s\right] \mathbb{P}\left(X_s/t \in \mathrm{d}x_s \mid X_u = t\right). \tag{3.37}$$

For $\eta > 0$, the inner expectation is equal to

$$\begin{aligned}
&\mathbb{E}\left[\min\left(1, L\|X_{C_m \setminus s}/t\|_1\right) \mathbb{1}\left\{\forall v \in C_m \setminus s : X_v/t \leq \eta\right\} \mid X_s/t = x_s\right] && (3.38) \\
&+ \mathbb{E}\left[\min\left(1, L\|X_{C_m \setminus s}/t\|_1\right) \mathbb{1}\left\{\exists v \in C_m \setminus s : X_v/t > \eta\right\} \mid X_s/t = x_s\right]. && (3.39)
\end{aligned}$$

The integrand in (3.38) is either zero because of the indicator function or, if the indicator is one, it is bounded by $L\,|C_m \setminus s|\,\eta$. The expression in (3.39) is clearly smaller than or equal to

$$\mathbb{P}\left(\exists v \in C_m \setminus s : X_v/t > \eta \mid X_s/t = x_s\right).$$

Going back to the integral in (3.37) we can thus bound it by

$$\begin{aligned}
\int_{[0,\delta]} &\left[L\,|C_m \setminus s|\,\eta + \mathbb{P}\left(\exists v \in C_m \setminus s : X_v/t > \eta \mid X_s/t = x_s\right)\right] \\
&\mathbb{P}\left(X_s/t \in \mathrm{d}x_s \mid X_u = t\right).
\end{aligned} \tag{3.40}$$

Consider the supremum of the probability in the integrand over the values $x_s \in [0,\delta]$ to bound the integral further. Hence (3.40) is smaller than or equal to

$$L\,|C_m \setminus s|\,\eta + \sup_{x_s \in [0,\delta]} \mathbb{P}\left(\exists v \in C_m \setminus s : X_v/t > \eta \mid X_s/t = x_s\right).$$

Using Assumption 3.3.4 and the fact that η can be chosen arbitrarily small we conclude that (3.34) converges to 0 as $t \to \infty$. Note that in case $\mathbb{P}(X_i = 0) = 0$ for any $i \in V$ the supremum above may exclude zero, so that the supremum can be restricted to all $x_s \in (0, \delta]$. However if $\mathbb{P}(X_i = 0) > 0$ for some $i \in V$ we need the supremum to be over all values in the closed interval $[0, \delta]$.

Finally we look at the term in (3.35). As f has range contained in $[0, 1]$ and is Lipschitz continuous, the expression in (3.35) is smaller than or equal to

$$\mathbb{E}\Big[\mathbb{1}(A_{us} < \delta) \min\Big(1,\, L \textstyle\sum_{v \in C_m \setminus s} A_{uv}\Big)\Big]. \tag{3.41}$$

From $(A_{uv}, v \in C_m \setminus s) = A_{u,C_m \setminus s} = A_{us} Z_{s,C_m \setminus s}$ we can write (3.41) as

$$\mathbb{E}\Big[\mathbb{1}(A_{us} < \delta) \min\Big(1, L A_{us} \textstyle\sum_{v \in C_m \setminus s} Z_{sv}\Big)\Big].$$

The random variable inside the expectation is bounded by 1 for any value of $\delta > 0$ and it converges to 0 as $\delta \downarrow 0$. By the bounded convergence theorem, the expectation in (3.41) converges to 0 as $\delta \downarrow 0$. □

3.6.2 Proof of Proposition 3.3.11

The quantile function of the unit-Fréchet distribution is $u \mapsto -1/\ln(u)$ for $0 < u < 1$. In view of Sklar's theorem and the identity (3.4), the copula, K, of G is

$$K(u) = G(-1/\ln u_1, \ldots, -1/\ln u_d) = \exp\big(-\ell(-\ln u_1, \ldots, -\ln u_d)\big).$$

for $u \in (0, 1)^d$. It follows that the partial derivative $\dot{K}_1$ of K with respect to its first argument exists, is continuous on $(0, 1)^2$ and is given by

$$\dot{K}_1(u) = \frac{K(u)}{u_1}\, \dot{\ell}_1(-\ln u_1, \ldots, -\ln u_d),$$

for $u \in (0, 1)^d$. The stdf is homogeneous: for $t > 0$ and $x \in (0, \infty)^d$, we have

$$\ell(tx_1, \ldots, tx_d) = t\, \ell(x_1, \ldots, x_d).$$

Taking the partial derivative with respect to x_1 on both sides and simplifying yields the identity

$$\dot{\ell}_1(tx_1, \ldots, tx_d) = \dot{\ell}_1(x_1, \ldots, x_d).$$

Let $F(x) = \exp(-1/x)$, for $x > 0$, denote the unit-Fréchet cumulative distribution function. Note that $-\ln F(x) = 1/x$ for $x > 0$. For $t > 0$ and $x = (x_2, \ldots, x_d) \in (0, \infty)^{d-1}$, we find

$$\begin{aligned}
&\mathbb{P}(\forall j \geq 2, X_j \leq t x_j \mid X_1 = t) = \dot{K}_1\big(F(t), F(tx_2), \ldots, F(tx_d)\big) \\
&= \frac{K\big(F(t), F(tx_2), \ldots, F(tx_d)\big)}{F(t)}\, \dot{\ell}_1(1/t, 1/(tx_2), \ldots, 1/(tx_d)) \\
&= \frac{K\big(F(t), F(tx_2), \ldots, F(tx_d)\big)}{F(t)}\, \dot{\ell}_1(1, 1/x_2, \ldots, 1/x_d).
\end{aligned}$$

As $t \to \infty$, the first factor on the right-hand side tends to one, whence

$$\lim_{t \to \infty} \mathbb{P}(\forall j \geq 2, X_j \leq t x_j \mid X_1 = t) = \dot{\ell}_1(1, 1/x_2, \ldots, 1/x_d).$$

To show that the right-hand side of the previous equation is indeed the cumulative distribution function of a $(d-1)$-variate probability measure on Euclidean space, it is

sufficient to show that, for every $j \in \{2, \dots, d\}$, the family of conditional distributions $(X_j/t \mid X_1 = t)$ as t ranges over $[t_0, \infty)$ for some large $t_0 > 0$ is uniformly tight. Indeed, the family of joint conditional distributions $((X_2, \dots, X_d)/t \mid X_1 = t)$ for $t \in [t_0, \infty)$ is then uniformly tight as well, and by Prohorov's theorem (van der Vaart, 1998, Theorem 2.4), we can find a sequence $t_n \to \infty$ such that the joint conditional distributions $(X/t_n \mid X_1 = t_n)$ converge weakly as $n \to \infty$, the limiting cumulative distribution function then necessarily being equal to the one stated above. It suffices to consider the case $d = j = 2$. By the first part of the proof above,

$$\lim_{t\to\infty} \mathbb{P}(X_2/t > x_2 \mid X_1 = t) = 1 - \dot{\ell}_1(1, 1/x_2).$$

Since $\ell : [0, \infty)^2 \to [0, \infty)$ is convex, the functions $y_1 \mapsto \ell(y_1, y_2)$ depend continuously on the parameter $y_2 \geq 0$. Since they are also convex, Attouch's theorem (Rockafellar and Wets, 1998, Theorem 12.35) implies that their derivatives depend continuously in y_2 as well, at least in continuity points y_1. But since $\ell(y_1, 0) = y_1$, we find that $\dot{\ell}_1(1, 1/x_2) \to \dot{\ell}_1(1, 0) = 1$ as $x_2 \to \infty$. For any $\epsilon > 0$, we can thus find $x_2(\epsilon) > 0$ such that $1 - \dot{\ell}_1(1, 1/x_2(\epsilon)) < \epsilon/2$ and then we can find $t(\epsilon) > 0$ such that $\mathbb{P}(X_2/t > x_2(\epsilon) \mid X_1 = t) < \epsilon/2 + 1 - \dot{\ell}_1(1, 1/x_2(\epsilon)) < \epsilon$ for all $t > t(\epsilon)$. The uniform tightness follows. □

3.6.3 Proof of Proposition 3.4.2

By Assumption 3.4.1, the random vector X satisfies Assumption 3.3.1 and it is Markov with respect to the graph $\mathcal{G}$. Assumption 3.3.4 is void (i.e., there is nothing to check), since, for each edge $(i, j) \in E$, the limiting distribution of $X_j/t \mid X_i = t$ as $t \to \infty$ is log-normal by Segers (2020b, Example 4) and Example 3.3.12 here and therefore does not have an atom at zero. We can thus apply Theorem 3.3.5 to conclude that $(X_v/t, v \in V \setminus u \mid X_u = t)$ converges weakly as $t \to \infty$. By the continuous mapping theorem, the same then holds true for $(\ln(X_v/t), v \in V \setminus u \mid X_u = t)$. It remains to calculate the limit distribution.

Calculating the limit in Theorem 3.3.5. By example 3.3.12 we have, as $t \to \infty$,

$$\big(\ln X_v - \ln X_s,\, v \in C \setminus s \mid X_s > t\big) \xrightarrow{d} \mathcal{N}_{|C\setminus s|}\big(\mu_{C,s}(\Delta_C), \Psi_{C,s}(\Delta_C)\big), \tag{3.42}$$

where the mean vector is

$$\big(\mu_{C,s}(\Delta)\big)_v = -2\delta_{sv}^2, \qquad v \in C \setminus s, \tag{3.43}$$

and the covariance matrix $\Psi_{C,s}(\Delta)$ is as in (3.20). It follows that if the random vector $Z_{s,C\setminus s}$ has law $\nu_{C,s}$, then the distribution of $(\ln Z_{sv}, v \in C \setminus s)$ is equal to the limit in (3.42). In particular, $\nu_{C,s}$ is multivariate log-normal.

For fixed $u \in V$, we will identify the limit $A_{u,V\setminus u}$ in Theorem 3.3.5. Let $Z = (Z_{s,C\setminus s}, C \in \mathcal{C})$ with $Z_{s,C} = (Z_{sv}, v \in C \setminus s)$ be the random vector constructed in Definition 3.3.2 by concatenating independent log-normal random vectors with distributions $\nu_{C,s}$. In this concatenation, recall that $s \in C$ and that either s is equal to u or s separates u and $C \setminus s$. We can write $Z = (Z_e, e \in E_u)$ where the E_u is the set of edges $e \in E$ that point away from u: for $e = (s, v) \in E_u$, either s is equal to u or s separates u and v. By construction, the distribution of Z is multivariate log-normal too. By (3.43), we have $\mathbb{E}[\ln Z_e] = -2\delta_e^2$ where $e = (s, v) \in E_u$. The covariance matrix of $(\ln Z_e, e \in E_u)$ has a block structure: for edges $e, f \in E_u$, the variables $\ln Z_e$ and $\ln Z_f$ are uncorrelated (and thus independent) if e and f belong to different cliques, while if they belong to the same clique, i.e., if $e = (s, i)$ and $f = (s, j)$ with $i, j, s \in C$ for some $C \in \mathcal{C}$, then, by (3.20), we have

$$\mathbb{C}\left(\ln Z_e, \ln Z_f\right) = 2(\delta_{si}^2 + \delta_{sj}^2 - \delta_{ij}^2) \tag{3.44}$$

By Theorem 3.3.5, we can express the limit $A_{u,V\setminus u}$ of $(X_v/t, v \in V \setminus u \mid X_u = t)$ as $t \to \infty$ in terms of Z: we have

$$\ln A_{uv} = \ln\left(\prod_{e\in(u\rightsquigarrow v)} Z_e\right) = \sum_{e\in(u\rightsquigarrow v)} \ln Z_e, \qquad v \in V \setminus u. \tag{3.45}$$

The distribution of $(\ln A_{uv}, v \in V \setminus u)$ is thus multivariate Gaussian, being the one of a linear transformation of the multivariate Gaussian random vector $(\ln Z_e, e \in E_u)$. The expectation of $\ln A_{uv}$ is readily obtained from (3.45):

$$\mathbb{E}[\ln A_{uv}] = \sum_{e\in(u\rightsquigarrow v)} \mathbb{E}[\ln Z_e] = \sum_{e\in(u\rightsquigarrow v)} (-2\delta_e^2) = -2p_{uv}, \qquad v \in V \setminus u,$$

which coincides with the element v of the vector $\mu_u(\Delta)$ in (3.23). It remains to show that the covariance matrix of $(\ln A_{uv}, v \in V \setminus u)$ is $\Sigma_u(\Delta)$ in (3.24).

Calculating $\Sigma_u(\Delta)$. Let $i, j \in V \setminus u$. By (3.45) and the bilinearity of the covariance operator, we have

$$\mathbb{C}\left(\ln A_{ui}, \ln A_{uj}\right) = \sum_{e\in(u\rightsquigarrow i)} \sum_{f\in(u\rightsquigarrow j)} \mathbb{C}\left(\ln Z_e, \ln Z_f\right).$$

Each of the paths $(u \rightsquigarrow i)$ and $(u \rightsquigarrow j)$ has at most a single edge in a given clique $C \in \mathcal{C}$; otherwise, they would not be the shortest paths from u to i and j, respectively. Let the node $a \in V$ be such that $(u \rightsquigarrow i) \cap (u \rightsquigarrow j) = (u \rightsquigarrow a)$. It could be that $a = u$, in which case the intersection is empty. Now we need to consider three cases.

(1) If $a = i$, i.e., if i lies on the path from u to j, then the random variables $\ln Z_f$ for $f \in (i \rightsquigarrow j)$ are uncorrelated with the variables $\ln Z_e$ for $e \in (u \rightsquigarrow i)$. By (3.44), the covariance becomes

$$\begin{aligned}\mathbb{C}\left(\ln A_{ui}, \ln A_{uj}\right) &= \sum_{e\in(u\rightsquigarrow i)} \mathbb{V}\left(\ln Z_e\right) \\ &= \sum_{e\in(u\rightsquigarrow i)} 4\delta_e^2 = 4p_{ui} = 2\left(p_{ui} + p_{uj} - p_{ij}\right),\end{aligned}$$

since $p_{ui} + p_{ij} = p_{uj}$, the path from u to j passing by i. This case includes the one where $i = j$, since then $(i \rightsquigarrow j)$ is empty and thus $p_{ij} = 0$.

(2) If $a = j$, the argument is the same as in the previous case.

(3) Suppose a is different from both i and j. Let e_a and f_a be the first edges of the paths $(a \rightsquigarrow i)$ and $(a \rightsquigarrow j)$, respectively. These two edges may or may not belong to the same clique. All other edges on $(a \rightsquigarrow i)$ and $(a \rightsquigarrow j)$, however, must belong to different cliques. It follows that

$$\begin{aligned}\mathbb{C}\left(\ln A_{ui}, \ln A_{uj}\right) &= \sum_{e\in(u\rightsquigarrow a)} \mathbb{V}\left(\ln Z_e\right) + \mathbb{C}\left(\ln Z_{e_a}, \ln Z_{f_a}\right) \\ &= 4p_{ua} + \mathbb{C}\left(\ln Z_{e_a}, \ln Z_{f_a}\right).\end{aligned}$$

Now we need to distinguish between two further sub-cases.

(3.a) Suppose e_a and f_a do not belong to the same clique. Then the covariance between $\ln Z_{e_a}$ and $\ln Z_{f_a}$ is zero, so that

$$\mathbb{C}\left(\ln A_{ui}, \ln A_{uj}\right) = 4p_{ua} = 2\left(\left(p_{ui} - p_{ai}\right) + \left(p_{uj} - p_{aj}\right)\right)$$

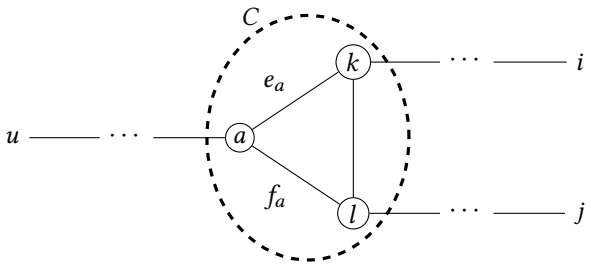


Figure 3.5: Calculation of $\mathbb{C}\left(\ln A_{ui}, \ln A_{uj}\right)$ in the proof of Proposition 3.4.2. The paths from u to i and from u to j have the path from u to a in common. On these two paths, the edges right after node a are $e_a = (a,k)$ and $f_a = (a,l)$, respectively. The picture considers the case (3.b) where the three nodes a, k, l belong to the same clique, say C.

$$\begin{aligned} &= 2\left(p_{ui} + p_{uj} - \left(p_{ai} + p_{aj}\right)\right) \\ &= 2\left(p_{ui} + p_{uj} - p_{ij}\right), \end{aligned}$$

since the shortest path between i and j passes through a.

(3.a) Suppose e_a and f_a belong to the same clique; see Figure 3.5. Writing $e_a = (a,k)$ and $f_a = (a,l)$, we find, in view of (3.44),

$$\begin{aligned} \mathbb{C}\left(\ln A_{ui}, \ln A_{uj}\right) &= 4p_{ua} + 2\left(\delta_{ak}^2 + \delta_{al}^2 - \delta_{kl}^2\right) \\ &= 2\left(\left(p_{ua} + \delta_{ak}^2\right) + \left(p_{ua} + \delta_{al}^2\right) - \delta_{kl}^2\right) \\ &= 2\left(p_{uk} + p_{ul} - \delta_{kl}^2\right) \\ &= 2\left(\left(p_{ui} - p_{ki}\right) + \left(p_{uj} - p_{lj}\right) - \delta_{kl}^2\right) \\ &= 2\left(p_{ui} + p_{uj} - \left(p_{ki} + p_{lj} + \delta_{kl}^2\right)\right) \\ &= 2\left(p_{ui} + p_{uj} - p_{ij}\right), \end{aligned}$$

since the shortest path between i and j passes by k and l.

We conclude that the covariance matrix of $(\ln A_{uv}, v \in V \setminus u)$ is indeed $\Sigma_u(\Delta)$ in (3.24).

Positive definiteness of $\Sigma_u(\Delta)$. Being a covariance matrix, $\Sigma_u(\Delta)$ is positive semi-definite. We need to show it is invertible. The linear transformation in (3.45) can be inverted to give

$$\ln Z_e = \ln A_{ub} - \ln A_{ua}$$

for an edge $e = (a,b)$ in E_u; note that either $a = u$, in which case $A_{ua} = 1$ and thus $\ln A_{ua} = 0$, or a lies on the path from u to b. Also, each edge $e \in E_u$ is uniquely identified by its endpoint v in $V \setminus u$; let $e(v)$ be the unique edge in E_u with endpoint v. It follows that, as column vectors, the random vectors $(\ln Z_{e(v)}, v \in V \setminus u)$ and $(\ln A_{uv}, v \in V \setminus u)$ are related by

$$(\ln A_{uv}, v \in V \setminus u) = M_u\left(\ln Z_{e(v)}, v \in V \setminus u\right),$$

where M_u is a $(|V|-1) \times (|V|-1)$ matrix indexed by $(v,w) \in (V \setminus u)^2$ whose inverse is

$$(M_u^{-1})_{vw} = \begin{cases} 1 & \text{if } w = v, \\ -1 & \text{if } (w,v) \in E_u, \\ 0 & \text{otherwise.} \end{cases}$$

The covariance matrix of $(\ln A_{uv}, v \in V \setminus u)$ is thus

$$\Sigma_u(\Delta) = M_u \, \Sigma_u^Z(\Delta) \, M_u^\top$$

where $\Sigma_u^Z(\Delta)$ is the (block-diagonal) covariance matrix of $(\ln Z_{e(v)}, v \in V \setminus u)$. The blocks in Σ_u^Z are given by (3.44) and are positive definite and thus invertible by the assumption that each parameter matrix Δ_C is conditionally negative definite. As a consequence, $\Sigma_u^Z(\Delta)$ is invertible too. Writing $\Theta_u^Z(\Delta) = \big(\Sigma_u^Z(\Delta)\big)^{-1}$, we find that $\Sigma_u(\Delta)$ is invertible as well with inverse

$$\Theta_u(\Delta) = (M_u^{-1})^\top \, \Theta_u^Z(\Delta) \, M_u^{-1}. \tag{3.46}$$

$P(\Delta)$ is conditionally negative definite. Clearly, $P(\Delta)$ is symmetric and has zero diagonal. For any non-zero vector $a \in \mathbb{R}^V$, we have, since $\Sigma_u(\Delta)$ is positive definite,

$$\begin{aligned}
0 &< a^\top \Sigma_u(\Delta) a \\
&= 2 \sum_{i \in V} \sum_{j \in V} a_i \left(p_{ui} + p_{uj} - p_{ij}\right) a_j \\
&= 2 \sum_{i \in V} a_i p_{ui} \sum_{j \in V} a_j + 2 \sum_{i \in V} a_i \sum_{j \in V} p_{uj} - 2 \sum_{i \in V} \sum_{j \in V} a_i p_{ij} u_j.
\end{aligned}$$

If $\sum_{i \in V} a_i = 0$, the first two terms on the right-hand side vanish. The last term is $-2a^\top P(\Delta) a$. We conclude that $P(\Delta)$ is conditionally negative definite, as required. □

3.6.4 Proof of Proposition 3.4.4

Let $H_{P(\Delta)}$ be the $|V|$-variate max-stable Hüsler–Reiss distribution in (3.21) with parameter matrix $P(\Delta)$ in (3.22). From (3.16) we have

$$-\ln H_{P(\Delta)}(x_v, v \in V) = \ell(1/x_v, v \in V) = \mathbb{E}\left[\max_{v \in V} x_v^{-1} A_{uv}\right].$$

By the maximum–minimums identity, we have

$$\mathbb{E}\left[\max_{v \in V} x_v^{-1} A_{uv}\right] = \sum_{i=1}^{|V|} (-1)^{i-1} \sum_{W \subseteq V : |W| = i} \mathbb{E}\left[\min_{v \in V} x_v^{-1} A_{uv}\right]$$

If W is a singleton, $W = \{u\}$, then the expectation is simply x_u^{-1}, whereas if W has more than one element, we write the expectation as the integral of the tail probability:

$$\begin{aligned}
\mathbb{E}\left[\min_{v \in W} x_v^{-1} A_{uv}\right] &= \int_0^\infty \mathbb{P}\left[\forall v \in W : x_v^{-1} A_{uv} > y\right] \mathrm{d}y \\
&= \int_0^{x_u^{-1}} \mathbb{P}\left[\forall v \in W \setminus u : A_{uv} > x_v y\right] \mathrm{d}y \\
&= \int_{\ln x_u}^\infty \mathbb{P}\left[\forall v \in W \setminus u : \ln A_{uv} > \ln(x_v) - z\right] \, e^{-z} \, \mathrm{d}z.
\end{aligned}$$

If $W = \{u\}$, we interpret the probability inside the integral as equal to one, so that the integral formula is valid for any non-empty $W \subseteq V$. Combining things, we find that

$$\begin{aligned}
&-\ln H_{P(\Delta)}(x_v, v \in V) \\
&\quad = \sum_{i=1}^{|V|} (-1)^{i-1} \sum_{W \subseteq V : |W| = i} \int_{\ln x_u}^\infty \mathbb{P}\left[\forall v \in W \setminus u : \ln A_{uv} > \ln(x_v) - z\right] e^{-z} \, \mathrm{d}z.
\end{aligned}$$

Recall that the distribution of $(\ln A_{uv}, v \in V \setminus u)$ is multivariate normal with mean vector $\mu_u(\Delta)$ and covariance matrix $\Sigma_u(\Delta)$ in (3.23) and (3.24), respectively, hence the expression of $H_{P(\Delta)}$ in (3.21). □

3.6.5 Proof of Proposition 3.4.5

By Proposition 3.4.4, the Markov field X in Assumption 3.4.1 satisfies (3.5) with F its joint cumulative distribution function and $G = H_{P(\Delta)}$. It follows that (3.7) holds too, yielding the weak convergence of high-threshold excesses to a Pareto random vector Y with distribution given in (3.6). It remains to show that Y is an extremal graphical model with respect to the given graph $\mathcal{G}$ in the sense of Definition 2 in Engelke and Hitz (2020).

Proposition 3.4.2 implies

$$(\ln X_v - \ln X_u, v \in V \setminus u \mid X_u > t) \xrightarrow{d} \mathcal{N}_{|V\setminus u|}\big(\mu_u(\Delta), \Sigma_u(\Delta)\big)$$

as $t \to \infty$. The latter is the distribution of $(\ln Y_v - \ln Y_u, v \in V \setminus u \mid Y_u > 1)$ for the multivariate Pareto random vector Y in (3.7) associated to the max-stable Hüsler–Reiss distribution with parameter matrix $P(\Delta)$.

To show that Y is an extremal graphical model with respect to the given block graph $\mathcal{G}$, we apply the criterion in Proposition 3 in Engelke and Hitz (2020). Let $\Theta_u(\Delta) = (\Sigma_u(\Delta))^{-1}$ be the precision matrix of the covariance matrix $\Sigma_u(\Delta)$ in (3.24); see (3.46). For $i, j \in V$ such that i and j are not connected, i.e., (i, j) is not an edge, we need to show that there is $u \in V \setminus \{i, j\}$ such that

$$\big(\Theta_u(\Delta)\big)_{ij} = 0.$$

Indeed, according to the cited proposition, the latter identity implies that

$$Y_i \perp_{\mathrm{e}} Y_j \mid Y_{\setminus\{i,j\}},$$

the relation $\perp_{\mathrm{e}}$ being defined in Definition 1 in Engelke and Hitz (2020), and thus that Y is an extremal graphical model with respect to $\mathcal{G}$.

For two distinct and non-connected nodes $i, j \in V$, let $u \in V \setminus \{i, j\}$. We will show that $(\Theta_u(\Delta))_{ij} = 0$. We have

$$\begin{aligned}\big(\Theta_u(\Delta)\big)_{ij} &= \sum_{a\in V\setminus u}\sum_{b\in V\setminus u}\big((M_u^{-1})^\top\big)_{ia}\big(\Theta_u^Z(\Delta)\big)_{ab}\big(M_u^{-1}\big)_{bj} \\ &= \sum_{a\in V\setminus u}\sum_{b\in V\setminus u}\big(M_u^{-1}\big)_{ai}\big(\Theta_u^Z(\Delta)\big)_{ab}\big(M_u^{-1}\big)_{bj}.\end{aligned}$$

Now, $\big(M_u^{-1}\big)_{ai}$ and $\big(M_u^{-1}\big)_{bj}$ are non-zero only if $a = i$ or $(i, a) \in E_u$ together with $b = j$ or $(j, b) \in E_u$. In neither case can a and b belong to the same clique, because otherwise we would have found a cycle connecting the nodes u, i, a, b, j. But if a and b belong to different cliques, then so do the edges $e(a)$ and $e(b)$ in E_u with endpoints a and b, and thus $\big(\Theta_u^Z(\Delta)\big)_{ab} = 0$, since $\Sigma_u^Z(\Delta)$ and thus $\Theta_u^Z(\Delta)$ are block-diagonal. □

3.6.6 Proof of Proposition 3.4.6

Necessity. Let $v \in \bar{U}$ have clique degree $\mathrm{cd}(v)$ at most two. We show that the full path sum matrix $P(\Delta)$ is not uniquely determined by the restricted path sum matrix $P(\Delta)_U$ and the graph $\mathcal{G}$. There are two cases: $\mathrm{cd}(v) = 1$ and $\mathrm{cd}(v) = 2$.

Suppose first that $\mathrm{cd}(v) = 1$. Then v belongs only to a single clique, say $C \in \mathcal{C}$. For any $i, j \in U$, the shortest path $(i \rightsquigarrow j)$ does not pass through v. Hence the edge weights δ^2_{vw} for $w \in C \setminus v$ do not show up in any path sum p_{ij} appearing in $P(\Delta)_U$. It follows that these edge weights can be chosen freely (subject to specifying a valid Hüsler–Reiss parameter matrix on C) without affecting $P(\Delta)_U$.

Suppose next that $\mathrm{cd}(v) = 2$. Without loss of generality, assume $U = V \setminus v$; this only enlarges the number of visible path sums with respect to the initial problem.

We will show that the path sum sub-matrix $(p_{ab}(\Delta))_{a,b\in V\setminus v}$ does not determine the complete path sum matrix $P(\Delta)$.

By assumption, v is included in two different cliques. Let the set of nodes from one of them, excluding v, be I and let the set of nodes from the other one, excluding v, be J. The sets I and J are non-empty and disjoint. We will show that the edge parameters δ^2_{vi} for $i \in I$ and δ^2_{vj} for $j \in J$ are not uniquely determined by the path sums p_{ab} for $a, b \in V \setminus \{v\}$.

- On the one hand, if the path $(a \rightsquigarrow b)$ does not pass by v, then the path sum p_{ab} does not contain any of the edge parameters δ^2_{vi} or δ^2_{vj} as a summand.
- On the other hand, if the path $(a \rightsquigarrow b)$ passes through v, then for some $i \in I$ and $j \in J$ determined by a and b the path sum p_{ab} contains the sum $\delta^2_{vi} + \delta^2_{vj}$ as a summand. However, sums of the latter form do not change if we decrease each δ^2_{vi} $(i \in I)$ by some small quantity, say $\eta > 0$, and simultaneously increase each δ^2_{vj} $(j \in J)$ by the same quantity, yielding $(\delta^2_{vi} - \eta) + (\delta^2_{vj} + \eta) = \delta^2_{vi} + \delta^2_{vj}$.

Sufficiency. Let every node in $\bar{U}$ have clique degree at least three. Let $a \in \bar{U}$ and let δ^2_{ab} be the parameter attached to the edge $(a, b) \in E$, with $b \in V \setminus a$. We will show that we can solve δ^2_{ab} from the observable path sums p_{ij} for $i, j \in U$.

By assumption there are at least three cliques that are connected to a, say I, J, and Y. Without loss of generality, assume $b \in I$. If $b \in U$ set $\bar{\imath} := b$, while if $b \in \bar{U}$ walk away from b up to the first node $\bar{\imath}$ in U and this along the unique shortest path between b and $\bar{\imath}$; note that $(a, b) \in (a \rightsquigarrow \bar{\imath})$. Apply a similar procedure to the cliques J and Y: choose a node $j \in J \setminus a$ (respectively $y \in Y \setminus a$) and if $j \in U$ $(y \in U)$ set $\bar{\jmath} := j$ $(\bar{y} := y)$, while if $j \in \bar{U}$ (respectively $y \in \bar{U}$) take the first node $\bar{\jmath}$ $(\bar{y})$ such that $(a, j) \in (a \rightsquigarrow \bar{\jmath})$ $[(a, y) \in (a \rightsquigarrow \bar{y})]$. Because the nodes $\bar{\imath}$, $\bar{\jmath}$ and $\bar{y}$ belong to U, the path sums $p_{\bar{\imath}\bar{\jmath}}$, $p_{\bar{\imath}\bar{y}}$, and $p_{\bar{y}\bar{\jmath}}$ are given. By construction, node a lies on the unique shortest paths between the nodes $\bar{\imath}$, $\bar{\jmath}$ and $\bar{y}$; see also Behtoei et al. (2010, Theorem 1(b)). It follows that

$$
\begin{aligned}
p_{\bar{\imath}\bar{\jmath}} &= p_{a\bar{\imath}} + p_{a\bar{\jmath}}, \\
p_{\bar{\imath}\bar{y}} &= p_{a\bar{\imath}} + p_{a\bar{y}}, \\
p_{\bar{y}\bar{\jmath}} &= p_{a\bar{\jmath}} + p_{a\bar{y}}.
\end{aligned}
$$

These are three equations in three unknowns, which can be solved to give, among others, $p_{a\bar{\imath}} = \frac{1}{2}(p_{\bar{\imath}\bar{y}} + p_{\bar{\imath}\bar{\jmath}} - p_{\bar{y}\bar{\jmath}})$. Now we distinguish between two cases, $b \in U$ and $b \in \bar{U}$.

- If $b \in U$ then $\bar{\imath} = b$ and we have written $\delta^2_{ab} = p_{a\bar{\imath}}$ in terms of the given path sums.
- If $b \in \bar{U}$ we repeat the same procedure as above but starting from node b. We keep the node $\bar{\imath}$, but the nodes $\bar{\jmath}$ and $\bar{y}$ may be different from those when starting from a. After having written $p_{b\bar{\imath}}$ in terms of observable path sums, we can compute $\delta^2_{ab} = p_{a\bar{\imath}} - p_{b\bar{\imath}}$. □

Max-linear graphical models with heavy-tailed factors on trees of transitive tournaments

4

This chapter corresponds to an article prepared with Johan Segers, carrying the same name as the chapter and submitted to *Applied Probability Journals* in September 2022.

4.1 Introduction

Dependence in multivariate linear factor models is determined by a collection of independent random variables, called factors, which are shared by the modelled variables. In extreme value analysis there are the max-linear and the additive factor models with heavy-tailed factors. In Einmahl et al. (2012), it is shown that both have the same max-domain of attraction.

In Gissibl and Klüppelberg (2018), a link is made between such factor models and probabilistic graphical models via a max-linear recursively defined structural equation model on a directed acyclic graph (DAG). Each node carries a variable defined as a *weighted maximum* of its parent variables and an independent factor. This leads to a representation of the graphical model as a (max-)factor model as in Einmahl et al. (2012), the factors relevant for a given variable being limited to the set of its ancestors. More recent is the linear causally structured model in Gnecco et al. (2021): each variable is the *weighted sum* of the variables on all its parent nodes plus an independent factor. This leads to a representation where a single variable is a weighted sum of all its ancestral factors.

In this paper, we study a type of graph that, to the best of our knowledge, is not yet known and which we gave a name that reflects its most important properties: a *tree of transitive tournaments (ttt)*, denoted by $\mathcal{T}$. A tournament is a graph obtained by directing a complete graph, while a tournament is said to be transitive if it has no directed cycles. The name reflects the interpretation of such a graph as a competition where every node is a player and a directed edge points from the winner to the loser. Some examples are hierarchical relations between members of animal and bird societies, brand preferences, and votes between two alternative policies (Harary and Moser, 1966). A ttt links up several such transitive tournaments in a tree-like structure. It is acyclic by construction. If there is a directed path from one node to another one, there is a unique shortest such path. Moreover, between any pair of nodes, there is a unique shortest undirected path.

In this paper, we study max-linear graphical models with respect to a ttt as defined in (4.2) below. In particular, for a max-linear random vector $X = (X_v, v \in V)$ with

node set V, we study the limit in distribution

$$(X_v/X_u, v \in V \setminus u \mid X_u > t) \xrightarrow{d} (A_{uv}, v \in V), \qquad t \to \infty. \tag{4.1}$$

It is not hard to show that the limit distribution in (4.1) is discrete (Segers, 2020b, Example 1). We show that if the ttt has a unique node without parents, a so-called source node, the joint distribution of $(A_{uv}, v \in V)$ is determined by products of independent multiplicative increments along the unique shortest undirected paths between the node u at which the high threshold is exceeded on the one hand and the rest of the nodes on the other hand. Such behaviour is analogous to that of Markov random fields on block graphs in Asenova and Segers (2021) and of Markov trees in Segers (2020b, Theorem 1). In turn, these results go back to the extensive literature on the additive or multiplicative structure of extremes for Markov chains (e.g. Smith, 1992; Yun, 1998; Segers, 2007; Janssen and Segers, 2014; Resnick and Zeber, 2013).

An underlying reason for the factorization into independent increments is the fact that a max-linear graphical model with respect to a ttt is a Markov random field with respect to the undirected graph associated to the original, directed graph when the ttt has a unique source. A ttt with unique source has no v-structures, that is, no nodes with non-adjacent parents. Both properties, the factorization of the limiting variables and the Markovianity with respect to the undirected graph, are lost if the graph contains v-structures. To show this, we rely on recent theory of conditional independence in max-linear Bayesian networks based on the notion of $*$-connectedness (Améndola et al., 2021; Améndola et al., 2022). This theory diverges from classical results on conditional independence in Bayesian networks based on the notion of d-separation (Lauritzen, 1996; Koller and Friedman, 2009).

In our paper the graph is given. A significant line of research in the context of extremal dependence is graph discovery. Given observations on a number of variables represented as nodes in a graph, the task is to estimate the edges. For Bayesian networks we can also talk about causality discovery because directed edges show the direction of influence. A first attempt to identify the DAG in the context of max-linear models is Gissibl et al. (2018), followed by several papers focusing on this topic: Klüppelberg and Krali (2021), Buck and Klüppelberg (2021), Gissibl et al. (2021), Tran et al. (2021a) and Tran et al. (2021b). The problems related to identifiability of the true graph and to the estimation of the edge weights are discussed in Klüppelberg and Lauritzen (2019). Gnecco et al. (2021) study a new metric called causal tail coefficient which is shown to reveal the structure of a linear causal recursive model with heavy-tailed noise. Graph discovery for non-directed graphs is studied in Engelke and Hitz (2020), Engelke and Volgushev (2020) and Hu et al. (2022).

Inspired from practice, and more specifically river network applications (Asenova et al., 2021), we study a different identifiability problem. If the structure of the graph is known, it may happen that on some nodes the variables are latent, i.e., unobserved. The identifiability problem in this case is whether two different parameter vectors can still generate the same distribution of the observable part of the model. If this is possible then we cannot uniquely identify all tail dependence parameters that characterize the full distribution. Similarly to Asenova and Segers (2021), the identifiability criterion involves properties of the nodes with latent variables. The criterion is specific for a ttt with unique source and is easy to check. Our identifiability problem resembles the "method of path coefficients" of Sewall Wright which uses a system of equations involving correlations to solve for the edge coefficients (Wright, 1934).

The novelty of the paper lies in several directions. First, a new class of graphs is introduced, called a tree of transitive tournaments (ttt), which is the directed acyclic analogue of a block graph. It can be seen as a generalization of a directed tree, where

edges are replaced by transitive tournaments. Second, we show that a max-linear graphical model over a ttt with unique source exhibits properties known for other graphical models, namely Markov trees (Segers, 2020b) and Markov block graphs (Asenova and Segers, 2021). In particular, when the ttt has a unique source, the model is Markov with respect to the skeleton of the graph. This property underlies the factorization of the tail limit into independent increments along the unique shortest trails. Finally, we study a problem of identifiability of the edge weights from the angular measure both when all variables are observed and also when some of them are latent.

The structure of the paper is as follows. In Section 4.2 we introduce the ttt, the max-linear model, and its angular measure, which plays a key role in almost all proofs. In Section 4.3 we discuss the limiting distribution of (4.1) and give four equivalent characterizations of a max-linear graphical model with respect to a ttt with unique source. The identifiability problem is covered in Section 4.4. The discussion summarizes the main points of the paper. The appendices contain some additional lemmas and the proofs that are not presented in the main text.

4.2 Notions and definitions

4.2.1 Tree of transitive tournaments

Let $\mathcal{T} = (V, E)$ be a directed acyclic graph (DAG) with finite vertex (node) set V and edge set $E \subset V \times V$. An edge $e := (u, v) \in E$ is directed meaning $(u, v) \neq (v, u)$; it is outgoing with respect to the parent node u and incoming with respect to the child node v. The graph $\mathcal{T}$ excludes loops, i.e., edges of the form (u, u), and as $\mathcal{T}$ is directed, we cannot have both $(u, v) \in E$ and $(v, u) \in E$. Two nodes u and v are adjacent if (u, v) or (v, u) is an edge. A cycle is a sequence of edges $e_1, \ldots, e_n$ with $e_k = (u_k, u_{k+1})$ and $u_1 = u_{n+1}$ for some nodes $u_1, \ldots, u_n$. The property that $\mathcal{T}$ is acyclic means that it does not contain any cycle. The graph $\mathcal{T}$ is assumed connected, i.e., for any two distinct nodes u and v we can find nodes $u_1 = u, u_2, \ldots, u_{n+1} = v$ such that u_k and u_{k+1} are adjacent for every $k = 1, \ldots, n$; we call the associated edge sequence an undirected path or a trail between u and v. If all edges are directed in the same sense, i.e., $(u_k, u_{k+1}) \in E$ for all $k = 1, \ldots, n$, we talk about a (directed) path from the ancestor u to the descendant v.

A graph, directed or not, is complete if there is an edge between any pair of distinct nodes. A subgraph of a graph is biconnected if the removal of any of its nodes will not disconnect the subgraph. A maximal biconnected subgraph, also known as a biconnected component, is a subgraph that cannot be extended by adding one adjacent node without violating this principle.

A directed complete graph is called a tournament. A tournament $\tau = (V_\tau, E_\tau)$ is transitive if $(u, v), (v, w) \in E_\tau$ implies $(u, w) \in E_\tau$. A transitive tournament is necessarily acyclic. The graph-theoretic properties of transitive tournaments are studied in Harary and Moser (1966). The property most used here is that the set of out-degrees of the d nodes of a transitive tournament is $\{d-1, d-2, \ldots, 0\}$; the in- and out-degrees of a node are the numbers of incoming and outgoing edges, respectively.

A subgraph of a graph is a maximal transitive tournament if it is not properly contained in another subgraph which is also a transitive tournament. The set of maximal transitive tournaments that are subgraphs of a DAG $\mathcal{T}$ will be denoted by $\mathbb{T}$. For brevity we will just write tournament when we mean a maximal transitive tournament and denote it by τ.

A block graph is an undirected graph where every maximal biconnected subgraph is a complete graph (Le and Tuy, 2010). Let T denote the non-directed version of $\mathcal{T}$, also called the skeleton of $\mathcal{T}$. It shares the same node set as $\mathcal{T}$, and for every edge

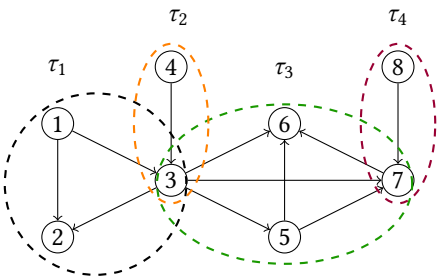


Figure 4.1: A tree of four maximal transitive tournaments: τ_1, τ_2, τ_3, and τ_4. The skeleton graph is the same, but with arrow heads removed. Node 3 is a separator node between tournaments τ_1, τ_2 and τ_3, while node 7 is a separator node between tournaments τ_3 and τ_4. Nodes 1, 4 and 8 are source nodes, i.e., have no parents. The subgraph with node set $\{1, 3, 4\}$ is a v-structure: node 3 has non-adjacent parents 1 and 4. Other v-structures within the ttt are the subgraphs with node sets $\{3, 7, 8\}$ and $\{5, 7, 8\}$. Between any pair of distinct nodes there is a unique shortest trail; for instance, nodes 4 and 8 are connected by the trail passing through nodes 3 and 7. There is no undirected cycle encompassing several tournaments.

(u, v) in the original graph $\mathcal{T}$, the reverse edge (v, u) is added to form the edge set of the skeleton graph T, after which each pair of edges $\{(u, v), (v, u)\}$ is identified with the undirected edge $\{u, v\}$ of T.

Definition 4.2.1 (Tree of transitive tournaments (ttt))**.** *A tree of transitive tournaments is a connected directed acyclic graph whose skeleton is a block graph.*

A tree of transitive tournaments (ttt) is illustrated in Figure 4.1. A ttt is acyclic: each tournament is acyclic and we cannot find a cycle passing by different tournaments either. This is why we call such a graph a *tree* of transitive tournaments.

A ttt enjoys three key properties. They all follow from the link with block graphs, whose characteristics can be found in Le and Tuy (2010).

(P1) two or more maximal transitive tournaments may have only one node in common, referred to as a *separator node*;

(P2) there is no undirected cycle that passes through nodes in different maximal transitive tournaments;

(P3) there is a unique shortest trail (undirected path) between every pair of nodes.

Similarly to block graphs (Le and Tuy, 2010) a ttt can be seen as a tree whose edges are replaced by transitive tournaments.

Recall that a path is directed by convention, so when we need non-directed paths this will be indicated explicitly. Between a pair of nodes there may be several paths. The set of all paths between two nodes $u, v \in V$ is denoted by $\pi(u, v)$. An element, say p, of $\pi(u, v)$ is a collection of edges, $\{(v_1, v_2), (v_2, v_3), \ldots, (v_{n-1}, v_n)\}$ for a path that involves the non-repeating nodes $\{v_1 = u, v_2, \ldots, v_{n-1}, v_n = v\}$. Note that $\pi(u, u) = \varnothing$ in an acyclic graph. In a ttt, if there is at least one path between distinct nodes u and v, there is a *unique shortest path* (see Lemma 4.6.1-1) between them, which we denote by $p(u, v)$, and which belongs to $\pi(u, v)$. We also set $p(u, u) = \varnothing$ by convention.

A source is a node without parents. If a DAG has a unique source, this node is an ancestor of every other node. This property follows from the following reasoning: let u_0 denote the unique source node of the DAG, and let v be any other node different from u_0. Then v must have a parent, say u. If $u = u_0$, we are done. Otherwise, replace v by u and restart. Since the graph is finite and has no cycles, this chain must stop at

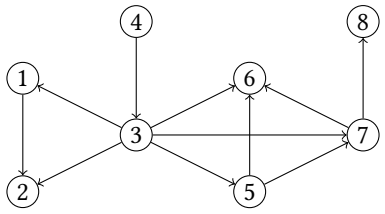


Figure 4.2: A tree of four maximal transitive tournaments. The skeleton graph is the same as the one in Fig. 4.1, but the graph here has a single source, on node 4. We see that there are no v-structures anymore in the graph.

some moment at a node without parents. But this node is necessarily equal to u_0 by assumption.

A key object in the paper is a ttt with unique source. In Lemma 4.6.1-2 it is shown that in this case there are no nodes with parents that are not adjacent or 'married', also known as a v-structure (Koller and Friedman, 2009). There are three v-structures in Fig. 4.1: one on nodes 1, 3, 4, one on 3, 7, 8 and another on 5, 7, 8. The main results in this paper are based on a ttt without v-structures. According to Lemma 4.6.1, there are no v-structures in a ttt with unique source. This is illustrated in Fig. 4.2. Other properties of a ttt with unique source are stated in Section 4.6.1.

Considered on its own, every tournament in a ttt has a unique source; this is due to the ordering of the out-degrees due to Harary and Moser (1966) mentioned earlier. When we talk about a source node, we will always state if we refer to the whole graph or with respect to a particular tournament.

4.2.2 Max-linear structural equation model on a ttt

In a general directed graph (V, E), let $\mathrm{pa}(v) \in V$ denote the set of parents of $v \in V$ and put $\mathrm{Pa}(v) = \mathrm{pa}(v) \cup \{v\}$. In a similar way let $\mathrm{an}(v)$, $\mathrm{desc}(v)$, and $\mathrm{ch}(v)$ denote the sets of ancestors, descendants, and children, respectively, excluding v, while $\mathrm{An}(v)$, $\mathrm{Desc}(v)$, and $\mathrm{Ch}(v)$ denote the same sets but including v. To each edge $e = (i, j) \in E$ we associate a weight $c_e = c_{ij} \in [0, \infty)$. The product of the edge parameters over a directed path $p = \{e_1, \ldots, e_m\}$ is denoted by

$$c_p = \prod_{r=1}^{m} c_{e_r}.$$

When the product is over the unique shortest path from u to v, we write $c_{p(u,v)}$. The product over the empty set being one by convention, we have $c_{p(i,i)} = 1$.

Let $(Z_i, i \in V)$ be a vector of independent unit-Fréchet random variables, i.e., $\mathbb{P}(Z_i \leq z) = \exp(-1/z)$ for $z > 0$. In the spirit of (Gissibl and Klüppelberg, 2018), a recursive max-linear model on a ttt, $\mathcal{T}$, is defined by

$$X_v = \bigvee_{i \in \mathrm{pa}(v)} c_{iv} X_i \vee c_{vv} Z_v, \qquad v \in V. \tag{4.2}$$

where the parameters c_e, for $e \in E$, and c_{vv}, for $v \in V$, are positive. We interpret this constraint as follows: if $c_{ij} = 0$, the variable X_j cannot be influenced by X_i through edge (i, j), and the edge could be removed from the graph. If $c_{vv} = 0$, the factor variable Z_v does not influence X_v. We don't want to deal with such border cases, so we assume that all parameters in the model definition (4.2) are positive. According to Gissibl and Klüppelberg (2018, Theorem 2.2) the expression in (4.2) is equal also to

$$X_v = \bigvee_{i \in V} b_{vi} Z_i, \tag{4.3}$$

with

$$b_{vi} = \begin{cases} 0 & \text{if } i \notin \mathrm{An}(v), \\ c_{vv} & \text{if } i = v, \\ c_{ii} \max_{p \in \pi(i,v)} c_p & \text{if } i \in \mathrm{an}(v). \end{cases} \tag{4.4}$$

The cumulative distribution function (cdf) of X_v is given by $\mathbb{P}(X_v \leq x) = \exp(-\sum_{i \in V} b_{vi}/x)$ for $x > 0$. We assume that $(X_v, v \in V)$ are unit-Fréchet, yielding the constraint

$$\sum_{i \in V} b_{vi} = \sum_{i \in \mathrm{An}(v)} b_{vi} = 1, \qquad \forall v \in V, \tag{4.5}$$

since $b_{vi} = 0$ whenever $i \notin \mathrm{An}(v)$. It is thus necessary and sufficient to have

$$c_{vv} = 1 - \sum_{i \in \mathrm{an}(v)} c_{ii} \max_{p \in \pi(i,v)} c_p, \tag{4.6}$$

with $c_{vv} = 1$ if $\mathrm{an}(v) = \emptyset$. By (4.6), the coefficients c_{vv} for $v \in V$ are determined recursively by the edge weights c_e for $e \in E$. If $c_{iv} \geq 1$ for some $(i, v) \in E$, then (4.2) implies that $X_v \geq X_i \vee c_{vv} Z_v$, and the constraint that X_i and X_v are unit-Fréchet distributed implies that $c_{vv} = 0$, a case we want to exclude, as explained above. This is why we impose $0 < c_e < 1$ for all $e \in E$ from the start, yielding the parameter space

$$\mathring{\Theta} = \left\{ \theta = (c_e, e \in E) \in (0,1)^E : \ \forall v \in V, \ c_{vv} > 0 \right\}. \tag{4.7}$$

The notion of *criticality* is important for max-linear structural equation models. We refer to Gissibl and Klüppelberg (2018), Améndola, Klüppelberg, Lauritzen, and Tran (2022), Gissibl, Klüppelberg, and Lauritzen (2021) and Klüppelberg and Lauritzen (2019) for examples where different conditional independence relations arise depending on which path is critical, or for illustrations in the context of graph learning. According to Gissibl and Klüppelberg (2018, Definition 3.1), a path $p \in \pi(i, v)$ is *max-weighted* under $\theta \in \Theta$ if it realizes the maximum $\max_{p' \in \pi(i,v)} c_{p'}$, where p' is any path in $\pi(i, v)$. In Améndola et al. (2022) the term *critical* is preferred.

If there is a (directed) path between two nodes, there is a unique shortest (directed) path between them (Lemma 4.6.1-1). This is crucial for our parametric model. We define the *critical parameter space* $\Theta_* \subset (0,1)^E$ as the set of parameters $\theta = (c_e, e \in E)$, such that for every $v \in V$ and every $i \in \mathrm{an}(v)$, the unique shortest directed path from i to v is the only critical path. Therefore we have $c_{p(i,v)} > c_p$, with strict inequality for any $p \in \pi(i, v)$ different from $p(i, v)$. Formally,

$$\Theta_* = \Big\{ \theta \in (0,1)^E : \ \forall v \in V, \ \forall i \in \mathrm{an}(v), \forall p \in \pi(i,v) \setminus \{p(i,v)\}, \\ c_{p(i,v)} > c_p \Big\}.$$

Next, we consider the intersection of the two spaces as an appropriate parameter space for our max-linear structural equation model:

$$\mathring{\Theta}_* = \mathring{\Theta} \cap \Theta_*. \tag{4.8}$$

For $\theta \in \mathring{\Theta}_*$, every element of the max-linear coefficient matrix $B_\theta = (b_{vi})_{v,i \in V}$ can be rewritten using an edge weight product over the unique shortest path $p(i, v)$ via

$$b_{vi} = \begin{cases} 0 & \text{if } i \notin \mathrm{An}(v), \\ c_{vv} & \text{if } i = v, \\ c_{ii} c_{p(i,v)} & \text{if } i \in \mathrm{an}(v), \end{cases} \qquad \text{and} \qquad c_{vv} = 1 - \sum_{i \in \mathrm{an}(v)} c_{ii} c_{p(i,v)}. \tag{4.9}$$

Also, note that $b_{ii} = c_{ii}$, leading to the frequently used expression

$$b_{vi} = c_{p(i,v)} b_{ii}, \qquad i \in \mathrm{an}(v).$$

Now, all elements are in place to describe our main object of interest.

Assumption 4.2.1 (Max-linear structural equation model on a ttt)**.** *The random vector* $X = (X_v, v \in V)$ *has the max-linear representation in* (4.3) *and* (4.9) *with respect to the ttt* $\mathcal{T} = (V, E)$ *(Definition 4.2.1) where* $(Z_v, v \in V)$ *is a vector of independent unit-Fréchet random variables and the edge weight vector* $\theta = (c_e, e \in E)$ *belongs to* $\mathring{\Theta}_*$ *in* (4.8).

The following identity for nodes with a unique parent will be useful:

$$\mathrm{pa}(v) = \{i\} \implies b_{vv} = 1 - c_{iv}. \tag{4.10}$$

Indeed, if i is the only parent of v, then $X_v = c_{iv}X_i \vee c_{vv}Z_v$ by (4.2). The variables X_v, X_i, Z_v are unit-Fréchet distributed and X_i is independent of Z_v, since X_i is a function of $(Z_u, u \in \mathrm{An}(i))$ and $v \notin \mathrm{An}(i)$. Hence $c_{iv} + c_{vv} = 1$, and because $c_{vv} = b_{vv}$, Eq. (4.10) follows.

A notational convention: in case of double subscripts, we may also write x_{i_1,i_2} instead of $x_{i_1 i_2}$.

4.2.3 The angular measure

Let X follow a max-linear model with parameter vector θ as in Assumption 4.2.1. The joint distribution P_θ of X on $[0,\infty)^V$ is max-stable and has unit-Fréchet margins. It is determined by

$$P_\theta([0,z]) = \mathbb{P}(X \le z) = \exp\left(-l_\theta\left((1/z_v)_{v\in V}\right)\right), \qquad z \in (0,\infty]^V,$$

where the stable tail dependence function (stdf) $l_\theta : [0,\infty)^V \to [0,\infty)$ is

$$l_\theta(x) = \sum_{i\in V} \max_{v\in V} (b_{vi}x_v) \tag{4.11}$$

for $x = (x_v)_{v\in V} \in [0,\infty)^V$ (Einmahl et al., 2012).

Let H_θ be the angular measure on the unit simplex $\Delta_V = \{a \in [0,1]^V : \sum_{v\in V} a^{(v)} = 1\}$ corresponding to the stdf l_θ. The link between the stdf and the angular measure is detailed in de Haan and Ferreira (2007) for the bivariate case and in Resnick (1987, Chapter 5) and Beirlant et al. (2004, Chapters 7–8) for higher dimensions: we have

$$l_\theta(x) = \int_{\Delta_V} \max_{v\in V} (a^{(v)}x_v)\, \mathrm{d}H_\theta(a).$$

In view of the expression of l_θ in (4.11), the angular measure is discrete and satisfies

$$H_\theta = \sum_{i\in V} m_i \delta_{a_i}, \tag{4.12}$$

with masses $m_i = \sum_{v\in V} b_{vi}$ and atoms $a_i = (b_{vi}/m_i)_{v\in V} \in \Delta_V$ for $i \in V$ (Einmahl, Krajina, and Segers, 2012, page 1779). The notation δ_x refers to a unit point mass at x.

If X follows a max-linear model, the angular measure of X is identifiable from its distribution P_θ via the limit relation

$$t\,\mathbb{P}\left(\frac{1}{\|X\|_1} X \in \cdot,\ \|X\|_1 > t\right) \xrightarrow{w} H_\theta(\,\cdot\,), \qquad t \to \infty,$$

where $\|x\|_1 = \sum_i |x_i|$ for a vector x in Euclidean space, while the arrow $\xrightarrow{w}$ denotes weak convergence of finite Borel measures, in this case on Δ_V.

When we discuss latent variables and identifiability in Section 4.4, we have to deal with the angular measure of a subvector of X, say $X_U = (X_v)_{v\in U}$, for non-empty

$U \subset V$. Its stdf $l_{\theta,U}$ arises from l_θ by setting $x_v = 0$ for all $v \notin U$: for $x \in [0,\infty)^U$ we have

$$l_{\theta,U}(x) = \sum_{i \in V} \max_{v \in U} (b_{vi} x_v) = \int_{\Delta_U} \max_{v \in U} (a^{(v)} x_v) \, \mathrm{d}H_{\theta,U}(a).$$

The distribution of X_U is max-linear too, so that its angular measure $H_{\theta,U}$ on Δ_U has a similar form as the one of X:

$$H_{\theta,U} = \sum_{i \in V} m_{i,U} \delta_{a_{i,U}}, \tag{4.13}$$

with masses $m_{i,U} = \sum_{v \in U} b_{vi}$ and atoms $a_{i,U} = (b_{vi}/m_{i,U})_{v \in U} \in \Delta_U$ for $i \in V$.

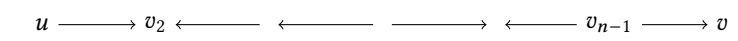
4.3 Conditional tail limit and the ttt with unique source

Here we study the limit distribution of

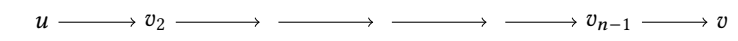
$$\left(\frac{X_v}{X_u}, v \in V \;\middle|\; X_u > t \right), \qquad t \to \infty, \tag{4.14}$$

when X is a max-linear model with respect to a ttt $\mathcal{T} = (V, E)$ as in Assumption 4.2.1. In particular, we are interested to know whether the elements of the limiting vector of (4.14) can be factorized into products of independent increments, similarly to other models with this property as in Segers (2020a) and Asenova and Segers (2021). In Proposition 4.3.1 below, we show that the limit variables factorize according to the unique shortest trails under the condition that the ttt has a unique source (node without parents). Moreover, by Proposition 4.3.2, the latter criterion is necessary and sufficient for X to satisfy the global Markov property with respect to the skeleton graph associated to $\mathcal{T}$, i.e., the undirected counterpart of $\mathcal{T}$.

According to property (P3), any pair of distinct nodes in a ttt is connected by a unique shortest trail. Let $t(u,v)$ denote the set of edges along the unique shortest trail between two distinct nodes u and v. Suppose the node sequence associated to the trail $t(u,v)$ is $\{u = v_1, \ldots, v = v_n\}$. Note that in the trail $t(u,v)$, the actual directions of the edges in the graph are preserved. In contrast, consider $t_u(u,v) = \{(v_1,v_2),(v_2,v_3),\ldots,(v_{n-1},v_n)\}$, i.e., the set of edges incident to the same set of nodes, $\{u = v_1, \ldots, v = v_n\}$, but directed from u to v, irrespective of their original directions. If for instance the graph of $t(u,v)$ is given by



then the graph of $t_u(u,v)$ is given by



For a given node $u \in V$, let E_u be the set of all edges in such unique shortest paths directed away from u, that is,

$$E_u = \bigcup_{v \in V \setminus u} t_u(u,v). \tag{4.15}$$

Recall from Section 4.2 that $\mathbb{T}$ denotes the set of tournaments within the ttt $\mathcal{T}$. For fixed $u \in V$ there is for every tournament $\tau = (V_\tau, E_\tau) \in \mathbb{T}$ a node, say $w_{u,\tau}$, which is the unique node in V_τ such that the trail $t(u, w_{u,\tau})$ is the shortest one among all trails between u and a node v in V_τ. As an example, consider Figure 4.3: starting from node $u = 8$, the closest node from the node set $V_{\tau_1} = \{1,2,3\}$ is 3, hence $w_{8,\tau_1} = 3$. With these definitions we are ready to state the condition under which the limiting variables factorize into independent increments.

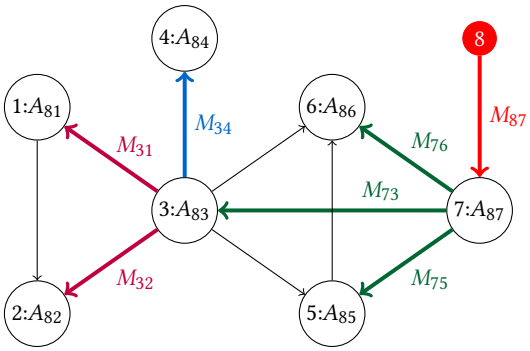


Figure 4.3: A ttt on four tournaments: τ_1 on node set $\{1,2,3\}$, τ_2 on $\{3,4\}$, τ_3 on $\{3,5,6,7\}$ and τ_4 on $\{8,7\}$. The variable exceeding a high threshold is at node 8. The set E_u is composed of the coloured edges which do not necessarily have the same directions in the original graph. On the nodes we have $A^{(8)} = (A_{8i}, i = 1,\dots,7)$ and on the edges we have the multiplicative increments $(M_e, e \in E_u)$. Increments in different colours are mutually independent, while those in the same color are dependent according to Lemma 4.6.3.

Proposition 4.3.1 (Factorization in max-linear model). *Let $(X_v, v \in V)$ follow a max-linear model as in Assumption 4.2.1. Fix $u \in V$. Let E_u be as in* (4.15) *and let $(M_e, e \in E_u)$ be a random vector composed of mutually independent subvectors $M^{(u,\tau)} = \big(M_{w_{u,\tau},j} : j \in V_\tau, (w_{u,\tau}, j) \in E_u\big)$, one for every transitive tournament $\tau \in \mathbb{T}$, and with marginal distribution as in Lemma 4.6.3.*

The following statements are equivalent:

(i) $\mathcal{T}$ has a unique source.

(ii) For every $u \in V$, we have, as $t \to \infty$, the weak convergence

$$\mathcal{L}(X_v/X_u, v \in V \mid X_u > t) \xrightarrow{d} \mathcal{L}(A^{(u)}) = \mathcal{L}(A_{uv}, v \in V) \tag{4.16}$$

with

$$A_{uv} = \prod_{e \in t_u(u,v)} M_e, \qquad v \in V. \tag{4.17}$$

(iii) There exists $u \in V$ such that the limit in (4.16) *and* (4.17) *holds.*

According to Proposition 4.3.1, the factorization property (4.17) holds either for all nodes or for no node at all, a necessary and sufficient condition being that the ttt has a unique source. The principle of (4.17) is illustrated in Figure 4.3 for $u = 8$. The limit $A^{(u)} = (A_{uv}, v \in V \setminus u)$ is given by

$$\begin{aligned} A_{87} &= M_{87}, & A_{83} &= M_{87}M_{73}, & A_{82} &= M_{87}M_{73}M_{32}, \\ & & A_{86} &= M_{87}M_{76}, & A_{81} &= M_{87}M_{73}M_{31}, \\ & & A_{85} &= M_{87}M_{75}, & A_{84} &= M_{87}M_{73}M_{34}, \end{aligned}$$

where $M^{(8,\tau_4)} = M_{87}$, $M^{(8,\tau_3)} = (M_{76}, M_{73}, M_{75})$, $M^{(8,\tau_2)} = M_{34}$ and $M^{(8,\tau_1)} = (M_{31}, M_{32})$ are independent sub-vectors by construction.

What underlies the link between the factorization of the limiting variables from Proposition 4.3.1 on the one hand and the uniqueness of the source of the ttt on the other hand is the Markovianity of X with respect to the skeleton graph T. The Markov property states that for any three non-empty and disjoint sets $A, B, C \subset V$ such that in the graph T the nodes in A are separated from the nodes in B by the nodes in C, the

vector $X_A = (X_v, v \in A)$ is conditionally independent from X_B given X_C (Lauritzen, 1996). Another equivalence condition can be added to the list in Proposition 4.3.1.

Proposition 4.3.2. *Let X follow a max-linear model with respect to the ttt $\mathcal{T}$ as in Assumption 4.2.1. Then X satisfies the global Markov property with respect to the skeleton graph T if and only if $\mathcal{T}$ has a unique source.*

The proof of Proposition 4.3.2 is based on notions and results from Améndola et al. (2022) which provides an extensive study of conditional independence properties of max-linear models. In particular, the notion of $*$*-connecting* path between two nodes in a DAG is introduced, a notion which is similar to the one of an *active* path (Koller and Friedman, 2009, Definition 3.6) between two nodes.

4.4 Latent variables and parameter identifiability

In practice, it is possible that on some of the nodes, the variables of interest are not observed (latent). Examples from the literature are water heights on certain locations on the river networks of the Danube in Asadi et al. (2015) and the Seine in Asenova et al. (2021). We look at the problem of recovering all parameters of the distribution of the complete vector, based on the distribution of the observed variables only. If this is possible, we can study the parametric model as if all variables were observed: in particular, we are able to compute measures of tail dependence for sets including the unobserved variables. The latter is important as it may be the only possible way to quantify tail dependence, because non-parametric estimates are not available when dealing with unobserved variables.

Consider for instance the network in Figure 4.4. The max-linear model on $\mathcal{T} = (V, E)$ has eight variables and eleven parameters $\theta = (c_e, e \in E)$. By Proposition 4.4.2 below, the parameter $\theta \in \mathring{\Theta}_*$ can be uniquely identified in case X_1, X_3, X_7 are not observed on the basis of the joint distribution of the remaining five variables, $X_U = (X_2, X_4, X_5, X_6, X_8)$.

The problem of parameter identifiability will be formalized on the level of the angular measure H_θ and is presented in detail in the next two subsections.

4.4.1 Graph-induced characteristics of the angular measure

In this subsection, we argue that the condition $\theta = (c_e, e \in E) \in \mathring{\Theta}_*$ guarantees that all edge weights in θ are uniquely identifiable from the angular measure H_θ of $X = (X_v, v \in V)$ and thus from the distribution P_θ of X. Recall from (4.12) that H_θ is discrete with atoms $a_i = (a_{vi})_{v \in V} \in \Delta_V$ and masses $m_i > 0$.

Thanks to the assumption $\theta \in \mathring{\Theta}_*$, we have

$$a_{vi} > 0 \iff b_{vi} > 0 \iff i \in \mathrm{An}(v) \iff v \in \mathrm{Desc}(i). \tag{4.18}$$

For any DAG, all nodes have a different set of descendants, i.e.,

$$\forall i, j \in V : i \neq j \implies \mathrm{Desc}(i) \neq \mathrm{Desc}(j). \tag{4.19}$$

Indeed, if $i \neq j$ and $\mathrm{Desc}(i) \subseteq \mathrm{Desc}(j)$, then $i \in \mathrm{desc}(j)$ and hence $j \notin \mathrm{desc}(i)$, so that $\mathrm{Desc}(j) \not\subseteq \mathrm{Desc}(i)$.

Lemma 4.4.1. *Let $(X_v, v \in V)$ follow a max-linear model as in Assumption 4.2.1, with parameter vector $\theta \in \mathring{\Theta}_*$ and induced coefficient matrix $(b_{vi})_{i,v \in V}$. Let $H_\theta = \sum_{i \in V} m_i \delta_{a_i}$ in* (4.12) *be its angular measure. Then*

(1) $m_i > 0$ for all $i \in V$;

(2) for any atom $a_i = (a_{vi})_{v\in V}$, we have $a_{vi} > 0$ if and only if $v \in \mathrm{Desc}(i)$. Specifically, all $|V|$ vectors a_i are different and every atom can be matched uniquely to a node in V;

(3) for each edge $(i, v) \in E$, we have $c_{iv} = b_{vi}/b_{ii} = a_{vi}/a_{ii}$.

In particular, $\theta \in \mathring{\Theta}_*$ is identifiable from H_θ and thus from P_θ, i.e., for $\theta_1 \neq \theta_2 \in \mathring{\Theta}_*$ we have $H_{\theta_1} \neq H_{\theta_2}$ and thus $P_{\theta_1} \neq P_{\theta_2}$.

In Lemma 4.4.1, if the edge (i, v) is not critical, then there is another path, say p', from i to v with path product $c_{p'} \geq c_{iv}$, and then we can further lower the value of c_{iv} without changing the coefficients in (4.4), because they involve $c_{p'}$ rather than c_{iv}, thus yielding the same measure H_θ. This shows that without the criticality assumption, some edge weights may not be identifiable from H_θ.

Example 4.4.1 (Unique zero patterns.)**.** *In dimension $d = 3$, consider an angular measure given by the following atoms and masses:*

$$\omega_1 = \frac{1}{2.2}(0.8, 1, 0.4), \mu_1 = 2.2, \qquad \omega_2 = \frac{1}{0.5}(0, 0, 0.5), \mu_2 = 0.5,$$
$$\omega_3 = \frac{1}{0.3}(0.2, 0, 0.1), \mu_3 = 0.3.$$

Consider the vectors $\beta_j = \mu_j\omega_j$ for $j \in \{1, 2, 3\}$. By Lemma 4.4.1, the unordered collection $\{\beta_1, \beta_2, \beta_3\} = \{(0.8, 1, 0.4)^\top, (0, 0, 0.5)^\top, (0.2, 0, 0.1)^\top\}$ permits to recover the values of the coefficients in the max-linear model

$$X_1 = c_{11}Z \vee c_{21}c_{22}Y, \quad X_2 = c_{22}Y, \quad X_3 = c_{13}c_{11}Z \vee c_{13}c_{21}c_{22}Y \vee c_{33}T.$$

with (known) edge set $E = \{(2, 1), (1, 3)\}$, and this due the presence of zeroes in the vectors. For the current example, argue as follows. The angular measure H_θ of (X_1, X_2, X_3) has three atoms: atom $a_Z = b_Z/m_Z$ with $b_Z = (c_{11}, 0, c_{13}c_{11})^\top$, atom $a_Y = b_Y/m_Y$ with $b_Y = (c_{21}c_{22}, c_{22}, c_{13}c_{21}c_{22})^\top$, and atom $a_T = b_T/m_T$ with $b_T = (0, 0, c_{33})^\top$. As unordered sets, $\{\beta_1, \beta_2, \beta_3\}$ and $\{b_Z, b_Y, b_T\}$ are equal, but the question is which vector β_j corresponds to which vector b_*. From an inspection of the zero entries of the vectors, it is easily seen that the only possibility to identify the three coefficient vectors $\beta_1, \beta_2, \beta_3$ with the vectors b_Z, b_Y, b_T of the angular measure H_θ is

$$\beta_1 = (0.8, 1, 0.4) = (c_{21}c_{22}, c_{22}, c_{13}c_{21}c_{22}) = b_Y,$$
$$\beta_2 = (0, 0, 0.5) = (0, 0, c_{33}) = b_T,$$
$$\beta_3 = (0.2, 0, 0.1) = (c_{11}, 0, c_{13}c_{11}) = b_Z.$$

Solving the equations yields $(c_{11}, c_{21}, c_{22}, c_{13}, c_{33}) = (0.2, 0.8, 1, 0.5, 0.5)$.

4.4.2 Identifiability issues with the angular measure of a subvector

When we deal with latent variables, we know the distribution of the observable variables only, $X_U = (X_v, v \in U)$ for non-empty $U \subset V$. The angular measure, say $H_{\theta,U}$, of X_U in (4.13) is discrete and takes the form

$$H_{\theta,U} = \sum_{r=1}^{s} \mu_r \delta_{\omega_r}, \tag{4.20}$$

with masses $\mu_r > 0$ and s distinct atoms $\omega_r \in \Delta_U$. Combining (4.13) and (4.20), we should have

$$\sum_{r=1}^{s} \mu_r \delta_{\omega_r} = \sum_{i\in V} m_{i,U} \delta_{a_{i,U}}, \tag{4.21}$$

which means that, as sets, we should have $\{\omega_1, \ldots, \omega_s\} = \{a_{i,U} : i \in V\}$. In contrast to the situation in Lemma 4.4.1, the subvectors $a_{i,U}$ for $i \in V$ are not necessarily all different. Any atom ω_r of $H_{\theta,U}$ is of the form $a_{i,U} = (b_{vi}/m_{i,U})_{v\in U}$ for one or possibly several indices $i \in V$. For $r = 1, \ldots, s$ and $i \in V$ such that $\omega_r = a_{i,U}$, we know from (4.18) that

$$\{v \in U : \omega_{r,v} > 0\} = \mathrm{Desc}(i) \cap U. \tag{4.22}$$

The (unordered) collection of vectors $\{(b_{vi})_{v\in U} : i \in V\}$ will be denoted by $\mathcal{B}_{\theta,U}$.

With unobservable variables, there are several issues with the angular measure and its expression on the right hand-side of (4.21).

- *Zero masses.* We have $m_{i,U} = \sum_{v\in U} b_{vi}$, so that if all components of $(b_{vi})_{v\in U}$ are zero, then $m_{i,U} = 0$. This happens when $\mathrm{Desc}(i) \cap U = \emptyset$. In this case, we have $s < |V|$, i.e., $H_{\theta,U}$ has less atoms than H_θ.
- *Equal atoms.* We may have $a_{i,U} = a_{j,U}$ for some indices $i, j \in V$ and $i \neq j$. In this case, the terms i and j in (4.13) are to be aggregated and again, $H_{\theta,U}$ has less than $|V|$ atoms, $s < |V|$. This happens when the vectors $(b_{vi}, v \in U)$ and $(b_{vj}, v \in U)$ are proportional for some distinct $i, j \in V$.
- *Zeroes on the same positions.* A more subtle problem occurs when for two distinct vectors $b, b' \in \mathcal{B}_{\theta,U}$, the supports $\{v \in U : b_v > 0\}$ and $\{v \in U : b'_v > 0\}$ are equal. Such a situation arises when two distinct nodes $i, j \in V$ satisfy $\mathrm{Desc}(i) \cap U = \mathrm{Desc}(j) \cap U$. The latter equality is only possible in the presence of latent variables and is to be contrasted with property (4.19) when all variables are observable.

4.4.3 Identifiability criterion

For a max-linear model with respect to a ttt $\mathcal{T} = (V, E)$ with unique source, we need conditions that ensure that the minimal representation of the angular measure of X_U is the one in (4.13). Consider the following two conditions for the set of nodes $\bar{U} = V \setminus U$ carrying latent variables:

(I1) any $u \in \bar{U}$ has at least two children;

(I2) any $u \in \bar{U}$ is the source of some tournament in $\mathcal{T}$.

Proposition 4.4.2. *Let X follow a max-linear model as in Assumption 4.2.1 with respect to a ttt $\mathcal{T} = (V, E)$ with unique source. For a non-empty node set $U \subset V$, the parameter $\theta \in \mathring{\Theta}_*$ is uniquely identifiable from the distribution of $(X_v, v \in U)$ if and only if conditions (I1) and (I2) are satisfied.*

Figure 4.4 illustrates the identifiability criterion.

4.5 Discussion

In this paper we have considered a Bayesian max-linear network over a special type of graph which we called a tree of transitive tournaments (ttt). It is a graph which collects in an acyclic manner transitive tournaments which are themselves complete DAGs. The max-linear model is defined on a particular parameter space which ensures that the impact from one variable to another takes place along the shortest path, a consideration that has been defined in the literature as the path's criticality. It turns out that a ttt with unique source leads to a graph without v-structures, or nodes with non-adjacent parents. The limit of the scaled random vector, conditional on the event that a high threshold is exceeded at a particular node, is shown to be factorizable in

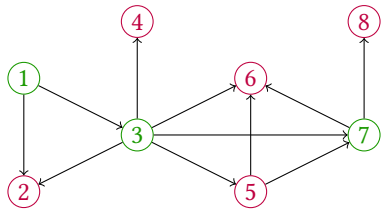


Figure 4.4: In the following ttt, the nodes that are allowed to contain a latent variable while the edge parameters remain identifiable are $1, 3, 7$. These are the only nodes where each of them satisfies both (I1) and (I2). For instance, if node 2 has unobserved variables, the parameters attached to edges $(1, 2), (3, 2)$ are not identifiable. This is because the edge weights c_{12}, c_{32} take part only in products over paths ending at 2. But if $2 \in \bar{U}$ these coefficients disappear from the atoms of the angular measure $a_{i,U} = (b_{vi}/m_{i,U}, v \in U)$ and accordingly from the collection of vectors $\mathcal{B}_{\theta,U}$.

independent multiplicative increments if and only if the ttt has a unique source. This result is analogous to that for Markov trees in Segers (2020b) and for Markov random fields on undirected block graph in Asenova and Segers (2021). The property that the Bayesian max-linear model on a ttt with unique source shares with these two other models is that it satisfies the global Markov property with respect to the undirected counterpart or skeleton graph of the ttt.

Upon appropriate modifications, we expect the results presented in this paper to hold equally for the linear additive causal model introduced in Gnecco et al. (2021). One of the reasons is that the max-domain of attraction of a linear model with heavy-tailed factors is the same as that of a max-linear one (Einmahl et al., 2012). However, the relation between the edge weights $\theta = (c_e)_{e \in E}$ and the coefficient matrix $B_\theta = (b_{ij})_{i,j \in V}$ is different between the max-linear and additive linear versions, and this may ask for different approaches in showing the same properties for the additive version.

4.6 Supplement

4.6.1 Trees of transitive tournaments

Recall that in a directed acyclic graph, a v-structure refers to a node with parents that are not adjacent, see Figure 4.1.

Lemma 4.6.1 (Properties I)**.** *Let $\mathcal{T}$ be a tree of transitive tournaments as in Definition 4.2.1. We have the following statements:*

1. *If there is a path between two nodes, then there is a unique shortest path between them.*
2. *The ttt $\mathcal{T}$ has a unique source if and only if it possesses no v-structures.*
3. *If $\mathcal{T}$ has a unique source, then for any two nodes $i \neq j$, the sets $\mathrm{Desc}(i)$ and $\mathrm{Desc}(j)$ are either disjoint or one contains the other, that is, i is an ancestor of j or vice versa.*

Proof. 1. Let $a, b \in V$. If a and b share the same tournament, they must be connected by an arrow, which is then the unique shortest path between them, since all other possible paths have length larger than one.

Let a, b be nonadjacent. If there is a unique directed path between a and b then this is the unique shortest path. Suppose now there are two shortest paths: $p_1, p_2 \in \pi(a, b)$. Let the path p_1 be along the vertices $\{v_1 = a, v_2, \ldots, v_n = b\}$ and the path p_2 on along the vertices $\{u_1 = a, u_2, \ldots, u_n = b\}$.

We will proceed by contradiction. Assume $v_2 \neq u_2$. If v_2 and u_2 belong to two different tournaments, then there exists a non-directed cycle through nodes in different tournaments, namely $\{a, v_2, \ldots, b, \ldots, u_2, a\}$. But this is impossible by property (P2) of a ttt. Hence, v_2 and u_2 must belong to the same tournament, say τ_a, because a is part of the same tournament too. Now consider u_3 and v_3. Then either $u_3 = v_3$ or they share a tournament, say τ_3, because otherwise there exists a non-directed cycle through nodes in different tournaments. Since $(v_2, v_3) \in E$ and $(u_2, u_3) \in E$ and by the assumption that $v_2 \neq u_2$, all four nodes $\{a, v_2, u_2, v_3 = u_3\}$ or all five nodes $\{a, v_2, u_2, v_3, u_3\}$ belong to τ_a. This is because by property (P1), two tournaments can share only one node, hence it is impossible to have $\tau_3 \cap \tau_a = \{v_2, u_2\}$. Because all four or five nodes belong to the same tournaments and since $(a, v_2), (v_2, v_3), (a, u_2), (u_2, u_3) \in E$ we must have $(a, v_3) \in E$ and $(a, u_3) \in E$ to avoid inter-tournament undirected cycles. Hence the paths $\{a = v_1, v_3, \ldots, v_n = b\}$ and $\{u_1 = a, u_3, \ldots, u_n = b\}$ are shorter than p_1 and p_2, a contradiction. Hence we must have $v_2 = u_2$.

We apply the same strategy to the nodes v_3, u_3 and v_4, u_4 to find that $v_3 = u_3$. Proceeding recursively, we conclude that $p_1 = p_2$.

2. First we show that if the ttt has a unique source, there cannot be a v-structure. We proceed by contraposition. Assume that there is a node, v, with parents in two different tournaments τ_a and τ_b. Let a and b be the sources of τ_a and τ_b respectively (Harary and Moser, 1966, Corollary 5a). Note that we definitely have $v \neq a$ and $v \neq b$. From node v go to node a. If a doesn't have a parent from another tournament we have found one node with zero in-degree with respect to the whole graph. If a has parent(s) from another tournament, say τ'_a, then go to the node that within τ'_a has in-degree zero, say node a'. Keep on going until you find a node with in-degree zero within the whole graph—such a node must exist because the graph is finite. Repeat the same for τ_b, yielding two different nodes having zero in-degree with respect to whole graph. These nodes must be different because of the definition of $\mathcal{T}$: since we have started in two different tournaments τ_a and τ_b we cannot end up in the same node, or otherwise there would be a non-directed cycle passing through v and that node. Hence we have found two nodes with zero in-degree, hence $\mathcal{T}$ does not have a unique source node.

Next we show that if $\mathcal{T}$ has two or more source nodes, u and v, then there is a v-structure. Because u and v are sources they have in-degree zero, so that they cannot belong to the same tournament, and thus they belong to two different tournaments. Consider the unique shortest trail between u, v on a sequence of nodes $\{u = v_1, v_2, \ldots, v_n = v\}$. Such a trail exists as, by definition of a ttt, the skeleton of $\mathcal{T}$ is a block graph and the fact that in a block graph there is a unique shortest path between every two nodes (Behtoei et al., 2010, Theorem 1). For every two consecutive nodes in the shortest path, v_i, v_{i+1}, we have either $(v_i, v_{i+1}) \in E$ or $(v_{i+1}, v_i) \in E$. Because u and v are sources of $\mathcal{T}$, we have $(u, v_2) \in E$ and $(v, v_{n-1}) \in E$. Note that $n \geq 3$, since u and v cannot be adjacent. We need three nodes v_i, v_{i+1}, v_{i+2} such that $(v_i, v_{i+1}) \in E$ and $(v_{i+2}, v_{i+1}) \in E$. If $n = 3$, then the triple (u, v_2, v) already fulfils the requirement. If $n \geq 4$, then continue from v_2 as follows. Let $i = \max\{j = 1, \ldots, n-2 : (v_j, v_{j+1}) \in E\}$; then $(v_i, v_{i+1}) \in E$ and $(v_{i+2}, v_{i+1}) \in E$, as required. Because this is the shortest trail, v_i and v_{i+2} cannot belong to the same tournament, since otherwise there would exist a shorter trail passing only through v_i and v_{i+2}.

3. Suppose that $v \in \mathrm{Desc}(i) \cap \mathrm{Desc}(j)$ but also both $i \notin \mathrm{an}(j)$ and $j \notin \mathrm{an}(i)$; in particular, i and j do not belong to the same tournament. Consider the paths $p(i, v)$ and $p(j, v)$. Along each path, continue walking upwards considering successive parents. Since the graph is finite, this walk must end for both paths to a node without parents. By assumption, this must be the same unique source node of the ttt, say u_0. We will thus have found two different paths from u_0 to v, one passing via i and the

other one via j. However, as i and j do not belong to the same tournament, this is in contradiction to property (P2) of a ttt. □

Lemma 4.6.2 (Properties II)**.** *Consider a ttt $\mathcal{T} = (V, E)$ as in Definition 4.2.1 with unique source.*

1. *If $\{v_1, v_2, \ldots, v_n\}$ is the node sequence of a unique shortest path between nodes v_1 and v_n, then all nodes except for possibly v_1 and v_n are the source node of the tournament shared with the next node in the sequence.*
2. *Between any two distinct nodes u, v in $\mathcal{T}$ the unique shortest trail between them is either $p(u, v)$ or $p(v, u)$ or there exists a node $w \in V \setminus \{u, v\}$ such that the trail is composed of the two shortest paths $p(w, u)$ and $p(w, v)$.*

Proof. 1. Suppose that there is a node v_r, for $r \in \{2, \ldots, n-1\}$, which is not the source node in the tournament shared with v_{r+1}, say τ. Let $\bar{v}$ be a parent of v_r in τ. Note that $\bar{v}$ must be a parent of v_{r+1} too, because of the out-degree ordering in a tournament. Because v_{r-1} is a parent of v_r too, both v_{r-1} and $\bar{v}$ must belong to τ, since otherwise v_r would have parents from different tournaments, which is impossible according to Lemma 4.6.1-2. Hence $v_{r-1}, v_r, \bar{v}, v_{r+1}$ all belong to the same tournament, i.e., to τ. Necessarily v_{r-1} is a parent of v_{r+1}, because otherwise there would be a directed cycle $\{v_{r-1}, v_r, v_{r+1}, v_{r-1}\}$. But then $\{v_1, \ldots, v_{r-1}, v_{r+1}, \ldots, v_n\}$ is a shorter path between v_1 and v_n, in contradiction to the hypothesis.

2. Let the shortest trail between u and v be the one along the node sequence $\{v_1 = u, \ldots, v_n = v\}$. It is sufficient to show that there cannot exist a node v_r for $r \in \{2, \ldots, n-1\}$ such that $(v_{r-1}, v_r) \in E$ and $(v_{r+1}, v_r) \in E$. Suppose indeed that the converse were true, i.e., there exists $r \in \{2, \ldots, n-1\}$ such that both v_{r-1} and v_{r+1} are parents of v_r. Then v_{r-1} and v_{r+1} must be adjacent because v-structures are excluded by statement 2 of Lemma 4.6.1. But then $\{v_1, \ldots, v_{r-1}, v_{r+1}, \ldots, v_n\}$ is a shorter trail between u and v, yielding a contradiction. □

4.6.2 Proofs and additional results for Section 4.3

Lemma 4.6.3. *Let $(X_v, v \in V)$ follow a max-linear model as in Assumption 4.2.1. Let $\tau \in \mathbb{T}$ be a transitive tournament on nodes V_τ. Then for $u \in V_\tau$, we have*

$$\begin{aligned} \mathcal{L}\left(\frac{X_v}{X_u}, v \in V_\tau \,\middle|\, X_u > t\right) \xrightarrow{d} \mathcal{L}(M^{(u,\tau)}) &= \mathcal{L}(M_{uv}, v \in V_\tau) \\ &= \sum_{j \in \mathrm{An}(u)} b_{uj} \delta_{\left\{\frac{c_{p(j,v)}}{c_{p(j,u)}}, v \in V_\tau\right\}}. \end{aligned} \tag{4.23}$$

The vector $M^{(u,\tau)} = (M_{uv}, v \in V_\tau)$ has dependent variables and the distribution of a single element is as follows.

1. *The distribution of M_{uv} when $(u, v) \in E$.*
   (a) *If u is the source node of τ, the distribution is given by $\mathcal{L}(M_{uv}) = \delta_{\{c_{uv}\}}$.*
   (b) *If u is not the source node of τ, the distribution is given by*
   $$\mathcal{L}(M_{uv}) = \sum_{j \in \mathrm{An}(u)} b_{uj} \delta_{\left\{\frac{c_{p(j,v)}}{c_{p(j,u)}}\right\}}.$$
2. *The distribution of M_{uv} when $(v, u) \in E$.*
   (a) *If v is the source node of τ, the distribution is given by*
   $$\mathcal{L}(M_{uv}) = c_{vu} \delta_{\{1/c_{vu}\}} + (1 - c_{vu}) \delta_{\{0\}}.$$
   (b) *If v is not the source node of τ, the distribution is given by*
   $$\mathcal{L}(M_{uv}) = \sum_{j \in \mathrm{An}(v)} b_{uj} \delta_{\left\{\frac{c_{p(j,v)}}{c_{p(j,u)}}\right\}} + \sum_{j \in \mathrm{An}(u) \setminus \mathrm{An}(v)} b_{uj} \delta_{\{0\}}.$$

Proof. From Segers (2020b, Example 1) we have the limit

$$\sum_{j\in V} b_{uj}\delta_{\left\{\frac{b_{vj}}{b_{uj}}, v\in V_\tau\right\}}.$$

Adapting this representation to a model where we have $b_{uj} = 0$ for $j \notin \mathrm{An}(u)$ and $b_{ij} = c_{p(j,i)}b_{jj}$ for $j \in \mathrm{An}(i)$ we obtain

$$\sum_{j\in \mathrm{An}(u)} b_{uj}\delta_{\left\{\frac{c_{p(j,v)}b_{jj}}{c_{p(j,u)}b_{jj}}, v\in V_\tau\right\}}.$$

Recall that $c_{p(i,i)} = 1$ and $c_{p(i,j)} = 0$ if $i \notin \mathrm{An}(j)$.

Next we show that $(M_{uv}, v \in V_\tau)$ are mutually dependent. When u is the source of τ then for every $j \in \mathrm{An}(u)$ the atom

$$\left(\frac{c_{p(j,v)}}{c_{p(j,u)}}, v \in V_\tau\right) = \left(\frac{c_{p(j,u)}c_{uv}}{c_{p(j,u)}}, v \in V_\tau\right) = (1; c_{uv}, v \in V_\tau \setminus u)$$

gets probability $\sum_{j\in \mathrm{An}(u)} b_{uj} = 1$. Hence $(M_{uv}, v \in V_\tau)$ are at the same time perfectly dependent and independent.

For u which is not the source node the general idea is to take a collection of coordinates with joint probability zero, and positive product of the marginal probabilities, thus showing that the joint probability does not equal the product of marginal probabilities for selected possible value of the random vector.

Let for brevity $V_\tau = \{1, 2, \ldots, m\}$: the nodes are labelled according to their order of out-degrees within τ: the source node of τ has $m - 1$ (largest) out-degree and is labelled by 1, the node with out-degree $m - 2$ is labelled as 2, etc.

Consider u being the node 2. We have, thanks to the no-cycle property within a tournament $\mathrm{An}(2) = \mathrm{An}(1) \cup \{2\}$. For all $j \in \mathrm{An}(1)$ we have

$$\left(\frac{c_{p(j,v)}}{c_{p(j,2)}}, v = 1, \ldots, m\right) = \left(\frac{1}{c_{12}}; 1; \frac{c_{p(j,1)}c_{1v}}{c_{p(j,1)}c_{12}}, v = 3, \ldots, m\right), \tag{4.24}$$

which is an atom of $(M_{2v}, v = 1, \ldots, m)$ with mass $\sum_{j\in \mathrm{An}(1)} b_{2j}$. This means that for the marginal distribution of M_{21} we have the probability $\mathbb{P}(M_{21} = 1/c_{12}) \geq \sum_{j\in \mathrm{An}(1)} b_{2j}$. For $j = 2$ we have an atom $(0, 1, c_{23}, \ldots, c_{2m})$ with mass b_{22}. This means that for the marginal probabilities of $(M_{23}, \ldots, M_{2m})$ we have $\mathbb{P}(M_{2v} = c_{2v}) \geq b_{22}$ for all $v = 3, \ldots, m$. Take a vector of coordinates $(1/c_{12}, 1, c_{23}, \ldots, c_{2m})$. Note that this vector cannot be the same as the one in (4.24). For any $v = 3, \ldots, m$ we cannot have $c_{1v}/c_{12} = c_{2v}$ because of the criticality assumption, according to which $c_{1v} > c_{12}c_{2v}$ for any $v = 3, \ldots, m$. The joint probability of this vector of coordinates is

$$\mathbb{P}(M_{21} = 1/c_{12}, M_{22} = 1, M_{23} = c_{23}, \ldots, M_{2m} = c_{2m}) = 0.$$

However the product of marginal probabilities is positive:

$$\mathbb{P}(M_{21} = 1/c_{12})\,\mathbb{P}(M_{22} = 1)\prod_{v=3}^{m}\mathbb{P}(M_{2v} = c_{2v}) \geq \sum_{j\in \mathrm{An}(1)} b_{2j} \times b_{22}^{m-1} > 0.$$

Now let $u \geq 3$. Take the vector of coordinates in (4.23) corresponding to $j = 1$ which is equal to $(1/c_{1u}, c_{12}/c_{1u}, \ldots, c_{1m}/c_{1u})$ and has probability at least b_{u1}. Consider also the vector of coordinates for $j = u$ which is $(0, \ldots, 0, 1; c_{uv}, v = u + 1, \ldots, m)$ with mass at least b_{uu}. Replace the first coordinate by $1/c_{1u}$. The vector obtained in this way has joint probability zero. For every $j \in \mathrm{pa}(u)$ we have $b_{vj}/b_{uj} = 0$ when v is not child of j or equivalently, given the order in the node labelling, when $v < j$.

So for fixed $u \geq 3$, for $j = 1$ the vector $(b_{vj}/b_{uj}, v = 1, \ldots, m)$ has no zeros. For $j = 2$ the vector $(b_{vj}/b_{uj}, v = 1, \ldots, m)$ has one zero, namely $(0; b_{vj}/b_{uj}, v = 2, \ldots, m)$, for $j = 3$ the vector $(b_{vj}/b_{uj}, v = 1, \ldots, m)$ has two zeros, namely $(0, 0; b_{vj}/b_{uj}, v = 3, \ldots, m)$ and so on until $j = u$ with the corresponding vector $(b_{vj}/b_{uj}, v = 1, \ldots, m) = (0, \ldots, 0; b_{vj}/b_{uj}, v = u, \ldots, m)$. By replacing the first coordinate by a non-zero value in this vector we get an impossible value for the random vector $(M_{uv}, v = 1, \ldots, m)$ or a value with probability zero. Considering the univariate marginal distributions of $(M_{uv}, v = 1, \ldots, m)$ we obtain for the product of marginal probabilities a positive value:

$$\mathbb{P}(M_{u1} = 1/c_{1u}) \left[\prod_{v=2}^{u-1} \mathbb{P}(M_{uv} = 0)\right] \mathbb{P}(M_{uu} = 1) \prod_{v=u+1}^{m} \mathbb{P}(M_{uv} = c_{uv}) \geq b_{u1} \times b_{uu}^{m-1}.$$

This shows that for any $u \in V_\tau$ the vector $(M_{u1}, \ldots, M_{um})$ has jointly dependent elements.

Next we show the distribution of a single element $M_{uv}, v \in V_\tau \setminus u$.

1. Consider first when u is the source node in τ. Since $(u, v) \in E$, we have $\mathrm{An}(u) \subset \mathrm{An}(v)$ and thus $\mathrm{An}(v) \cap \mathrm{An}(u) = \mathrm{An}(u)$. We have $b_{vj} > 0, j \in \mathrm{An}(u)$, hence zero is not a possible value of M_{uv}. For $j \in \mathrm{An}(u)$

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{p(j,u)} c_{uv} b_{jj}}{c_{p(j,u)} b_{jj}} = c_{uv},$$

and since $\sum_{j \in \mathrm{An}(u)} b_{uj} = 1$ we obtain the desired result under 1.(a).

When u is not the source node in τ not all shortest paths to v pass through u hence for $j \in \mathrm{An}(u)$ we have

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{p(j,v)} b_{jj}}{c_{p(j,u)} b_{jj}} = \frac{c_{p(j,v)}}{c_{p(j,u)}} > 0$$

with mass b_{uj}. Hence the result in 1.(b). Note that zero is not possible value as we still have $\mathrm{An}(u) \subset \mathrm{An}(v)$. Also $c_{p(u,u)} = 1$ by convention.

2. Let us have now $(v, u) \in E$. In this case $\mathrm{An}(u) \setminus \mathrm{An}(v)$ is not empty because it contains at least the node u, so zero is a possible value of M_{uv}. We need to distinguish only the zero atoms from the non-zero ones. When v is a source node in τ, we have, for $j \in \mathrm{An}(v)$

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{p(j,v)} b_{jj}}{c_{p(j,u)} b_{jj}} = \frac{c_{p(j,v)}}{c_{p(j,v)} c_{vu}} = \frac{1}{c_{vu}} > 0,$$

which is an atom with probability

$$\sum_{j \in \mathrm{An}(v)} b_{uj} = \sum_{j \in \mathrm{An}(v)} c_{p(j,v)} c_{vu} b_{jj} = c_{vu} \sum_{j \in \mathrm{An}(v)} c_{p(j,v)} b_{jj} = c_{vu} \sum_{j \in \mathrm{An}(v)} b_{vj} = c_{vu}.$$

The probability of the zero atom is

$$\sum_{j \in \mathrm{An}(u) \setminus \mathrm{An}(v)} b_{uj} = \sum_{j \in \mathrm{An}(u)} b_{uj} - \sum_{j \in \mathrm{An}(v)} b_{uj} = 1 - c_{vu}.$$

This shows 2.(a).

When v is not a source node of τ we have for $j \in \mathrm{An}(v)$

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{p(j,v)} b_{jj}}{c_{p(j,u)} b_{jj}} = \frac{c_{p(j,v)}}{c_{p(j,u)}} > 0,$$

an atom with mass b_{uj} and zero atom with probability $\sum_{j \in \mathrm{An}(u) \setminus \mathrm{An}(v)} b_{uj}$. This shows 2.(b). □

Remark. From the results in Lemma 4.6.3 we see that a multiplicative increment does not have a degenerate distribution at zero, so that a product of several such multiplicative increments cannot be degenerate at zero either. This is an important observation that we will use in further proofs.

Lemma 4.6.4. *Let $(X_v, v \in V)$ follow a max-linear model as in Assumption 4.2.1.Let $\mathcal{T}$ have a unique source. For any $u \in V$ we have*

$$\mathcal{L}\left(\frac{X_v}{X_u}, v \in V \;\middle|\; X_u > t\right) \xrightarrow{d} \mathcal{L}(A_{uv}, v \in V) = \sum_{j \in \mathrm{An}(u)} b_{uj} \delta_{\left\{\frac{c_{p(j,v)}}{c_{p(j,u)}}, v \in V\right\}}. \tag{4.25}$$

The distribution of the random variable A_{uv} depends on the three types of possible trails according to Lemma 4.6.2-2. In what follows we assume $(u,v) \notin E$. For the case $(u,v) \in E$ see Lemma 4.6.3.

1. *Distribution of A_{uv} on a path $\{u = v_1, r = v_2, \ldots, v = v_n\}$ with $u, r \in \tau$, one of the tournaments of $\mathcal{T}$.*
   (a) *If u is a source node in τ then $\mathcal{L}(A_{uv}) = \delta_{\{c_{p(u,v)}\}}$.*
   (b) *If u is not a source node in τ we have*

$$\mathcal{L}(A_{uv}) = \sum_{j \in \mathrm{An}(u)} b_{uj} \delta_{\left\{\frac{c_{p(j,r)}}{c_{p(j,u)}} c_{p(r,v)}\right\}}.$$

2. *Distribution of A_{uv} on a path $\{v = v_1, r = v_2, \ldots, u = v_n\}$ with $v, r \in \tau$.*
   (a) *If v is a source node in τ then*

$$\mathcal{L}(A_{uv}) = c_{p(v,u)} \delta_{\left\{\frac{1}{c_{p(v,u)}}\right\}} + (1 - c_{p(v,u)}) \delta_{\{0\}}.$$

   (b) *If v is not a source node in τ then*

$$\mathcal{L}(A_{uv}) = \sum_{j \in \mathrm{An}(v)} c_{p(r,u)} b_{rj} \delta_{\left\{\frac{c_{p(j,v)}}{c_{p(j,r)} c_{p(r,u)}}\right\}} + \sum_{j \in \mathrm{An}(u) \setminus \mathrm{An}(v)} b_{uj} \delta_{\{0\}}.$$

3. *The distribution of A_{uv} on a trail composed of two paths $p(r,u)$ and $p(r,v)$. Let the trail be on nodes $\{u, \ldots, m, r, n, \ldots, v\}$. Let also τ_m, τ_n be two tournaments with $r, m \in \tau_m$ and $r, n \in \tau_n$.*
   (a) *If r is source in both τ_m and τ_n, then*

$$\mathcal{L}(A_{uv}) = c_{p(r,u)} \delta_{\left\{\frac{c_{p(r,v)}}{c_{p(r,u)}}\right\}} + (1 - c_{p(r,u)}) \delta_{\{0\}}.$$

   (b) *If r is source in τ_m, but not in τ_n, then*

$$\mathcal{L}(A_{uv}) = \sum_{j \in \mathrm{An}(r)} c_{p(r,u)} b_{rj} \delta_{\left\{\frac{c_{p(j,n)} c_{p(n,v)}}{c_{p(j,r)} c_{p(r,u)}}\right\}} + \sum_{j \in \mathrm{An}(u) \setminus \mathrm{An}(r)} b_{uj} \delta_{\{0\}}.$$

   (c) *If r is source in τ_n, but not in τ_m, then*

$$\mathcal{L}(A_{uv}) = \sum_{j \in \mathrm{An}(r)} c_{p(m,u)} b_{mj} \delta_{\left\{\frac{c_{p(j,r)} c_{p(r,v)}}{c_{p(j,m)} c_{p(m,u)}}\right\}} + \sum_{j \in \mathrm{An}(u) \setminus \mathrm{An}(r)} b_{uj} \delta_{\{0\}}.$$

Proof. We have already seen that from Segers (2020b, Example 1) we have the limit

$$\mathcal{L}\left(\frac{X_v}{X_u}, v \in V \mid X_u > t\right) \xrightarrow{d} \sum_{j \in V} b_{uj} \delta_{\left\{\frac{b_{vj}}{b_{uj}}, v \in V\right\}}.$$

Adapting this representation to a model where we have $b_{uj} = 0$ for $j \notin \mathrm{An}(u)$ and $b_{ij} = c_{p(j,i)} b_{jj}$ for $j \in \mathrm{An}(i)$ we obtain

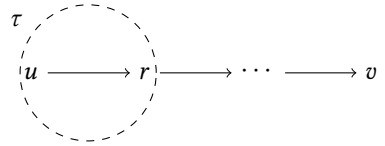
$$\sum_{j \in \mathrm{An}(u)} b_{uj} \delta_{\left\{\frac{c_{p(j,v)}}{c_{p(j,u)}}, v \in V\right\}}.$$

Recall that $c_{p(i,i)} = 1$ and $c_{p(i,j)} = 0$ if $i \notin \mathrm{An}(j)$. For a single $v \in V \setminus u$ we have the marginal distribution

$$\mathcal{L}(A_{uv}) = \sum_{j \in \mathrm{An}(u)} b_{uj} \delta_{\left\{\frac{b_{vj}}{b_{uj}}\right\}}. \tag{4.26}$$

The distribution of A_{uv} depends deterministically on properties of the ttt. When $\mathcal{T}$ has a unique source, according to Lemma 4.6.2-2 there are three possible shortest trails between two nodes. In addition we have also the property under Lemma 4.6.2-1. We look at the different distributions of A_{uv} that arise due to these two properties of the ttt.

First we deal with 1.(a). Since $\mathrm{An}(u) \subset \mathrm{An}(v)$ all atoms in (5.31) are positive and zero is not a possible value of A_{uv}. All paths from $\mathrm{An}(u)$ to v pass through u because u is source in τ and because by property (P2) of a ttt no cycle involving several tournaments is allowed. The case is illustrated by the graph below.

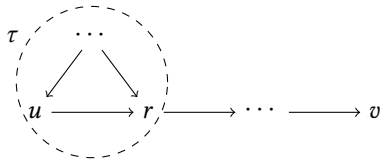


Hence for all $j \in \mathrm{An}(u)$ we have

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{p(j,v)} b_{jj}}{c_{p(j,u)} b_{jj}} = \frac{c_{p(j,u)} c_{p(u,v)}}{c_{p(j,u)}} = c_{p(u,v)} > 0,$$

with mass $\sum_{j \in \mathrm{An}(u)} b_{uj} = 1$.

Next we show 1.(b). Because $\mathrm{An}(u) \subset \mathrm{An}(v)$, zero is not possible value of A_{uv}. Not all shortest paths from $\mathrm{An}(u)$ to v pass through u because u is not source in τ. However all paths from $\mathrm{An}(u)$ to v pass through r, as shown in the picture. Paths from $\mathrm{An}(u)$ to v other than these passing through u or r are impossible because of the property (P2) of a ttt.

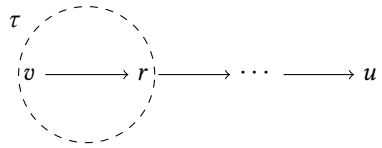


We have for $j \in \mathrm{An}(u)$

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{p(j,v)} b_{jj}}{c_{p(j,u)} b_{jj}} = \frac{c_{p(j,r)}}{c_{p(j,u)}} c_{p(r,v)} > 0,$$

with mass b_{uj}, hence the expression in 1.(b).

Next we show 2.(a). When the directed path is from v to u the set $\mathrm{An}(u) \setminus \mathrm{An}(v)$ contains at least u hence we have $b_{vj} = 0$ for all $j \in \mathrm{An}(u) \setminus \mathrm{An}(v)$. This means that zero is a possible value of A_{uv}. All shortest paths from $j \in \mathrm{An}(v)$ to u pass through v as v is source in τ. Otherwise, there would be cycle encompassing multiple tournaments, which is not allowed under property (P2) of a ttt.



For $j \in \mathrm{An}(v)$ the non-zero atom is given by

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{p(j,v)} b_{jj}}{c_{p(j,u)} b_{jj}} = \frac{c_{p(j,v)}}{c_{p(j,v)} c_{p(v,u)}} = \frac{1}{c_{p(v,u)}} > 0, \qquad j \in \mathrm{An}(v),$$

with mass

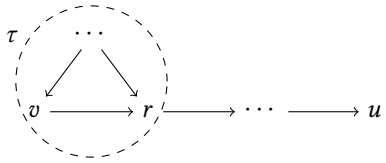
$$\begin{aligned} \sum_{j \in \mathrm{An}(v)} b_{uj} &= \sum_{j \in \mathrm{An}(v)} c_{p(j,v)} c_{p(v,u)} b_{jj} = c_{p(v,u)} \sum_{j \in \mathrm{An}(v)} c_{p(j,v)} b_{jj} \\ &= c_{p(v,u)} \sum_{j \in \mathrm{An}(v)} b_{vj} = c_{p(v,u)}. \end{aligned}$$

For the zero atom we have probability

$$\sum_{j \in \mathrm{An}(u) \setminus \mathrm{An}(v)} b_{uj} = \sum_{j \in \mathrm{An}(u)} b_{uj} - \sum_{j \in \mathrm{An}(v)} b_{uj} = 1 - c_{p(v,u)}.$$

This shows 2.(a).

To show 2.(b) we note that when v is not a source node of τ not all shortest paths from $j \in \mathrm{An}(v)$ to u pass through v. However all paths from $j \in \mathrm{An}(v)$ to u pass through r, as it can be seen from the figure here.



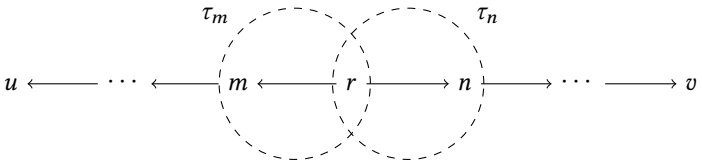
Hence for $j \in \mathrm{An}(v)$ we have

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{p(j,v)} b_{jj}}{c_{p(j,u)} b_{jj}} = \frac{c_{p(j,v)}}{c_{p(j,r)} c_{p(r,u)}} > 0,$$

which is an atom with mass $b_{uj} = c_{p(j,r)} c_{p(r,u)} b_{jj} = c_{p(r,u)} b_{rj}$. The zero atom comes from the fact that $b_{vj} = 0$ for all $j \in \mathrm{An}(u) \setminus \mathrm{An}(v)$, and it has probability $\sum_{j \in \mathrm{An}(u) \setminus \mathrm{An}(v)} b_{uj}$. This shows the distribution under 2.(b).

By Lemma 4.6.2-1 the node r, as part of the paths $p(r,u)$ is allowed not be a source node in τ_m. Similarly considering the path $p(r,v)$. However when we combine $p(r,u)$ and $p(r,v)$ in one trail $t(u,v)$ the node r should be a source in at least one of τ_m and τ_n. If r is not source of both τ_m and τ_n then there would be indeed a v-structure. However, Lemma 4.6.1-2 excludes v-structures when $\mathcal{T}$ has a unique source, hence node r should be source in at least one tournament, τ_m and/or τ_n.

To show 3.(a) we note that all paths from $j \in \mathrm{An}(r)$ to u and to v pass through r, as r is source in both τ_n and τ_m. The case is depicted in the following picture.



Also we have $b_{vj} = 0$ for all $j \in \mathrm{An}(u) \setminus \mathrm{An}(r)$. For $j \in \mathrm{An}(r)$ we have

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{p(j,r)}c_{p(r,v)}b_{jj}}{c_{p(j,r)}c_{p(r,u)}b_{jj}} = \frac{c_{p(r,v)}}{c_{p(r,u)}} > 0,$$

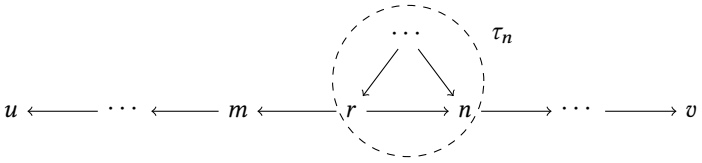
with probability

$$\sum_{j\in\mathrm{An}(r)} b_{uj} = \sum_{j\in\mathrm{An}(r)} c_{p(j,r)}c_{p(r,u)}b_{jj} = c_{p(r,u)} \sum_{j\in\mathrm{An}(r)} c_{p(j,r)}b_{jj}$$
$$= c_{p(r,u)} \sum_{j\in\mathrm{An}(r)} b_{rj} = c_{p(r,u)}.$$

The probability of the zero atom is

$$\sum_{j\in\mathrm{An}(u)\setminus\mathrm{An}(r)} b_{uj} = \sum_{j\in\mathrm{An}(u)} b_{uj} - \sum_{j\in\mathrm{An}(r)} b_{uj} = 1 - c_{p(r,u)}.$$

Next we show 3.(b). Because r is not a source in τ_n not all paths from $\mathrm{An}(r)$ to v pass through r, but they do all pass through n. Also all paths from $\mathrm{An}(r)$ to u pass through r because r is source in τ_m.

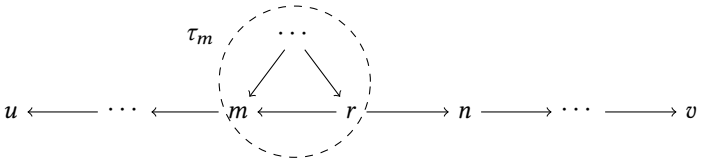


Hence for $j \in \mathrm{An}(r)$

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{p(j,n)}c_{p(n,v)}b_{jj}}{c_{p(j,r)}c_{p(r,u)}b_{jj}} = \frac{c_{p(j,n)}c_{p(n,v)}}{c_{p(j,r)}c_{p(r,u)}} > 0,$$

which is an atom with mass $b_{uj} = c_{p(j,r)}c_{p(r,u)}b_{jj} = b_{rj}c_{p(r,u)}$. The zero atom has probability equal to $\sum_{j\in\mathrm{An}(u)\setminus\mathrm{An}(r)} b_{uj}$.

Next we show 3.(c). When r is source in τ_n it means that all paths from $\mathrm{An}(r)$ to v pass through r. Because r is not source in τ_m not all paths from $\mathrm{An}(r)$ to u pass through r, but they do all pass through m.



For $j \in \mathrm{An}(r)$ we have

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{p(j,r)}c_{p(r,v)}b_{jj}}{c_{p(j,m)}c_{p(m,u)}b_{jj}} = \frac{c_{p(j,r)}c_{p(r,v)}}{c_{p(j,m)}c_{p(m,u)}} > 0,$$

which is an atom with mass $b_{uj} = c_{p(j,m)}c_{p(m,u)}b_{jj} = b_{mj}c_{p(m,u)}$. The zero atom comes from $b_{uj} = 0$ for all $j \in \mathrm{An}(u) \setminus \mathrm{An}(r)$. It gets probability $\sum_{j\in\mathrm{An}(u)\setminus\mathrm{An}(r)} b_{uj}$. □

Proof of Proposition 4.3.1

Proof. First we prove that (i) implies (ii). Assume $\mathcal{T}$ has a unique source. We have to prove that for any $u \in V$ an element from the limiting vector in (4.16) is given by (4.17).

In Lemma 4.6.4 we have seen a number of cases for the distribution of A_{uv} depending on deterministic properties of the trail between u and v. Below we consider each of these cases again.

Case 1. Let the unique shortest trail between u and v be a path on node sequence $\{u = v_1, r = v_2, \ldots, v_n = v\}$. Let τ be the tournament containing u, r.

Case 1.(a). – Let u be source in τ. From Lemma 4.6.4-1.(a) we have $P(A_{uv} = c_{p(u,v)}) = 1$. Consider the variables $(M_e, e \in p(u,v))$ which are by construction independent between each other because they belong to different tournaments. Note that in this case all nodes $v_1, \ldots, v_{n-1}$ are source nodes in the tournament containing that node and the next one in the sequence. This follows from Lemma 4.6.2-1. Then according to Lemma 4.6.3 1.(a) for every $M_e, e \in p(u,v)$ we have $\mathbb{P}(M_e = c_e) = 1$ and hence

$$\mathbb{P}\left(\prod_{e\in p(u,v)} M_e = c_{p(u,v)}\right) = \prod_{e\in p(u,v)} \mathbb{P}(M_e = c_e) = 1,$$

which shows $A_{uv} = \prod_{e\in p(u,v)} M_e$.

Case 1.(b). – If u is not the source in τ, the distribution of M_{ur} is as in Lemma 4.6.3-1.(b). As in the case 1.(a) all nodes $r = v_2, v_3, \ldots, v_{n-1}$ are source nodes in the tournament containing that node and the next one in the sequence. The variables $M_e, e \in p(r,v)$ are degenerate at c_e. As the case 1.(a) above the variables $(M_e, e \in p(u,v))$ are by construction independent between each other because they are indexed by edges which belong to different tournaments. Then we have

$$\begin{aligned}\mathcal{L}\left(\prod_{e\in p(u,v)} M_e\right) = \mathcal{L}\left(M_{ur}\prod_{e\in p(r,v)} M_e\right) &= \left(\sum_{j\in \mathrm{An}(u)} b_{uj}\delta_{\left\{\frac{c_{p(j,r)}}{c_{p(j,u)}}\right\}}\right) \otimes \delta_{\{c_{p(r,v)}\}} \\ &= \sum_{j\in \mathrm{An}(u)} b_{uj}\delta_{\left\{\frac{c_{p(j,r)}}{c_{p(j,u)}} c_{p(r,v)}\right\}}.\end{aligned} \tag{4.27}$$

The sign $\otimes$ denotes multiplication between two discrete probability measures, say μ and ν of two independent variables, say ξ_1, ξ_2 respectively. For two possible values a_1, a_2 of ξ_1, ξ_2 respectively we have $\mu(\{a_1\})\nu(\{a_1\})$ as a measure of the event $\{\xi_1\xi_2 = a_1a_2\} = \{\xi_1 = a_1, \xi_2 = a_2\}$. The last one expression in (4.27) is the distribution of A_{uv} in Lemma 4.6.4-1.(b).

Case 2. Let the unique shortest trail between u and v be a path from v to u on the node sequence $\{v = v_1, r = v_2, \ldots, v_n = u\}$. Let τ be the tournament containing v, r.

Case 2.(a). – Let v be source in τ. Consider the random variables $M_{v_{i+1},v_i}, i = 1, \ldots, n-1$ whose distributions are as in Lemma 4.6.3-2.(a). Since this is the unique shortest trail from v to u, all edges on it belong to different tournaments and the vector $(M_{v_{i+1},v_i}, i = 1, \ldots, n-1)$ contains independent variables by definition. Then

$$\mathbb{P}\left(\prod_{i=1}^{n-1} M_{v_{i+1},v_i} = \frac{1}{c_{p(v,u)}}\right) = \prod_{i=1}^{n-1} \mathbb{P}\left(M_{v_{i+1},v_i} = \frac{1}{c_{v_i,v_{i+1}}}\right) = \prod_{i=1}^{n-1} c_{v_i,v_{i+1}} = c_{p(v,u)}. \tag{4.28}$$

For the zero atom we have

$$\begin{aligned}\mathbb{P}\left(\prod_{i=1}^{n-1} M_{v_{i+1},v_i} = 0\right) &= 1 - \prod_{i=1}^{n-1} \mathbb{P}(M_{v_{i+1},v_i} > 0) \\ &= 1 - \prod_{i=1}^{n-1} \mathbb{P}\left(M_{v_{i+1},v_i} = \frac{1}{c_{v_i,v_{i+1}}}\right) = 1 - c_{p(v,u)}.\end{aligned} \tag{4.29}$$

The expressions in (4.28) and (4.29) represent indeed the distribution of A_{uv} in Lemma 4.6.4-2.(a).

Case 2.(b). – If v is not the source in τ, consider a random variable M_{rv} with distribution as in Lemma 4.6.3-2.(b) and a random variable A_{ur} constructed as in 2.(a) here above, i.e., as the product $\prod_{i=2}^{n-1} M_{v_{i+1},v_i}$. By construction M_{rv} is independent from A_{ur} with the same argument as above. We have

$$\begin{aligned}\mathcal{L}(A_{ur}M_{rv}) = &\left(c_{p(r,u)}\delta_{\left\{\frac{1}{c_{p(r,u)}}\right\}} + (1 - c_{p(r,u)})\delta_{\{0\}}\right) \\ &\otimes \left(\sum_{j\in \mathrm{An}(v)} b_{rj}\delta_{\left\{\frac{c_{p(j,v)}}{c_{p(j,r)}}\right\}} + \sum_{j\in \mathrm{An}(r)\setminus \mathrm{An}(v)} b_{rj}\delta_{\{0\}}\right)\end{aligned}$$

which gives non-zero atoms $c_{p(j,v)}/(c_{p(j,r)}c_{p(r,u)})$, $j \in \mathrm{An}(v)$ with masses $b_{rj}c_{p(r,u)}$, $j \in \mathrm{An}(v)$. To show the probability of the zero atom, consider

$$\begin{aligned}&\mathbb{P}(A_{ur}M_{rv} = 0) = 1 - \mathbb{P}(A_{ur} > 0) \cdot \mathbb{P}(M_{rv} > 0) \\ &= 1 - c_{p(r,u)} \sum_{j\in \mathrm{An}(v)} b_{rj} = \sum_{j\in \mathrm{An}(u)} b_{uj} - \sum_{j\in \mathrm{An}(v)} c_{p(j,r)}b_{jj}c_{p(r,u)} \\ &= \sum_{j\in \mathrm{An}(u)} b_{uj} - \sum_{j\in \mathrm{An}(v)} b_{uj} = \sum_{j\in \mathrm{An}(u)\setminus \mathrm{An}(v)} b_{uj},\end{aligned}$$

which is what we need to confirm $A_{uv} = A_{ur}M_{rv}$ where A_{uv} is as in Lemma 4.6.4-2.(b).

Case 3. In the three cases that follow let the unique shortest trail from u to v be given by two paths $p(r,u)$ and $p(r,v)$. Let the trail be on nodes $\{u,\dots,m,r,n,\dots,v\}$. Let also τ_m, τ_n be two tournaments with $r, m \in \tau_m$ and $r, n \in \tau_n$.

Case 3.(a). – Let r be source in both τ_m and τ_n. Consider random variables A_{rv} as in Lemma 4.6.4-1.(a) and A_{ur} as in Lemma 4.6.4-2.(a). Above we have shown in cases 1.(a) and 2.(a) that A_{rv} and A_{ur} are factorizable in independent multiplicative increments. By construction A_{rv} and A_{ur} are independent from each other, because the multiplicative increments are independent. We have

$$\mathbb{P}\left(A_{ur}A_{rv} = \frac{c_{p(r,v)}}{c_{p(r,u)}}\right) = \mathbb{P}\left(A_{ur} = \frac{1}{c_{p(r,u)}}\right)\mathbb{P}(A_{rv} = c_{p(r,v)}) = c_{p(r,u)}.$$

For the probability of the zero atom we have

$$\mathbb{P}(A_{ur}A_{rv} = 0) = P(A_{ur} = 0) = (1 - c_{p(r,u)}).$$

The two displays above represent indeed the distribution of A_{uv} in Lemma 4.6.4-3.(a).

Case 3.(b). – Let r be source in τ_m, but not source in τ_n. Consider three random variables A_{ur}, M_{rn}, A_{nv} with distributions as in Lemma 4.6.4-2.(a), Lemma 4.6.3-1.(b) and Lemma 4.6.4-1.(a) respectively. For A_{ur} and A_{nv} we have shown in cases 2.(a)

and 1.(a) in this proof that they are factorizable in independent multiplicative increments. By construction M_{rn} is independent from the increments in A_{ur} and A_{nv}. Then

$$\begin{aligned}
&\mathcal{L}(A_{ur}M_{rn}A_{nv}) \\
&= \left(c_{p(r,u)}\delta_{\left\{\frac{1}{c_{p(r,u)}}\right\}} + (1-c_{p(r,u)})\delta_{\{0\}} \right) \otimes \left(\sum_{j\in \mathrm{An}(r)} b_{rj}\delta_{\left\{\frac{c_{p(j,n)}}{c_{p(j,r)}}\right\}} \right) \otimes \delta_{\{c_{p(n,v)}\}} \\
&= \sum_{j\in \mathrm{An}(r)} b_{rj}c_{p(r,u)}\delta_{\left\{\frac{c_{p(j,n)}c_{p(n,v)}}{c_{p(j,r)}c_{p(r,u)}}\right\}} + (1-c_{p(r,u)})\delta_{\{0\}}.
\end{aligned}$$

Note that

$$\begin{aligned}
\sum_{j\in \mathrm{An}(u)\setminus \mathrm{An}(r)} b_{uj} &= \sum_{j\in \mathrm{An}(u)} b_{uj} - \sum_{j\in \mathrm{An}(r)} b_{uj} = 1 - \sum_{j\in \mathrm{An}(r)} c_{p(j,r)}c_{p(r,u)}b_{jj} \\
&= 1 - c_{p(r,u)} \sum_{j\in \mathrm{An}(r)} b_{rj} = 1 - c_{p(r,u)}.
\end{aligned}$$

This shows that the distribution of $A_{ur}M_{rn}A_{nv}$ is the one of A_{uv} in Lemma 4.6.4-3.(b).

Case 3.(c). – Let r be source in τ_n, but not in τ_m. Consider variables A_{um}, M_{mr}, A_{rv} with distributions as in Lemma 4.6.4-2.(a), Lemma 4.6.3-2.(b) and Lemma 4.6.4-1.(a) respectively. The variables A_{um} and A_{rv} have been shown to factorize in independent increments in cases 2.(a) and 1.(a) of this proof respectively, hence they are independent from each other too. By construction M_{mr} is independent from A_{um} and A_{rv}. Then we have

$$\begin{aligned}
\mathcal{L}(A_{um}M_{mr}A_{rv}) &= \left(c_{p(m,u)}\delta_{\left\{\frac{1}{c_{p(m,u)}}\right\}} + (1-c_{p(m,u)})\delta_{\{0\}} \right) \\
&\otimes \left(\sum_{j\in \mathrm{An}(r)} b_{mj}\delta_{\left\{\frac{c_{p(j,r)}}{c_{p(j,m)}}\right\}} + \sum_{j\in \mathrm{An}(m)\setminus \mathrm{An}(r)} b_{mj}\delta_{\{0\}} \right) \otimes \delta_{\{c_{p(r,v)}\}}.
\end{aligned}$$

The non-zero atoms are $c_{p(j,r)}c_{p(r,v)}/(c_{p(j,m)}c_{p(m,u)})$ for $j \in \mathrm{An}(r)$ with masses $c_{p(m,u)}b_{mj} = c_{p(j,m)}c_{p(m,u)}b_{jj} = b_{uj}$ for $j \in \mathrm{An}(r)$. The probability of the zero atom is given by

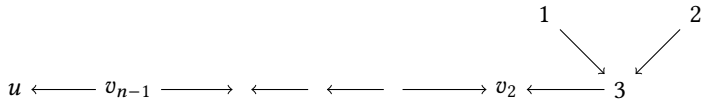
$$\begin{aligned}
\mathbb{P}(A_{um}M_{mr}A_{rv} = 0) &= 1 - \mathbb{P}(A_{um} > 0)\,\mathbb{P}(M_{mr} > 0) \\
&= 1 - c_{p(m,u)} \sum_{j\in \mathrm{An}(r)} b_{mj} \\
&= 1 - \sum_{j\in \mathrm{An}(r)} c_{p(j,m)}c_{p(m,u)}b_{jj} \\
&= \sum_{j\in \mathrm{An}(u)} b_{uj} - \sum_{j\in \mathrm{An}(r)} b_{uj} = \sum_{j\in \mathrm{An}(u)\setminus \mathrm{An}(r)} b_{uj}.
\end{aligned}$$

Hence the distribution of $A_{um}M_{mr}A_{rv}$ is the one of A_{uv} in Lemma 4.6.4-3.(c). This completes the proof that the statement in (i) implies (ii).

The statement in (iii) holds trivially from (ii).

Next we prove that (iii) implies (i) by contraposition: we assume that $\mathcal{T}$ has at least two sources and we will show that it is not possible to obtain the factorization in (4.17). If $\mathcal{T}$ has at least two sources, then by Lemma 4.6.1-2 there is at least one v-structure, say on nodes $1, 2, 3$ and involving edges $(1,3), (2,3) \in E$. Consider the nodes $1, 2$. For every $u \in V$ we have two possibilities:

(a) When the v-structure belongs to only one of the two trails $t(u,1)$ or $t(u,2)$.



(b) When each node of the v-structure belongs to one of the two trails $t(u,1)$ and $t(u,2)$.

Figure 4.5: The two possible configurations of the trails $t(u,1)$ and $t(u,2)$ when nodes $1,2,3$ form a v-structure.

(a) the v-structure belongs to only one of the trails $t(u,1)$ and $t(u,2)$: w.l.o.g. $(1,3),(2,3) \in t(u,2)$ and $(1,3),(2,3) \notin t(u,1)$;
(b) each trail $t(u,1)$ and $t(u,2)$ contains one edge of the v-structure: w.l.o.g. $(1,3) \in t(u,1)$ and $(2,3) \in t(u,2)$.

If $u \in \{1,2\}$, then we are in case 1, while if $u = 3$, we are in case 2. If $u \notin \{1,2,3\}$, then node 3 must belong to at least one of the two trails $t(u,1)$ or $t(u,2)$, because otherwise the skeleton graph would have a cycle connecting nodes $u,1,2,3$ and passing through more than one block. The latter is impossible according to property (P2). The two possibilities are illustrated in Figure 4.5.

Case 3.(c-i). Consider first the case when, w.l.o.g., the v-structure belongs to $t(u,2)$ but not to $t(u,1)$, see Figure 4.5(a). Let the trail from 1 to u be on nodes $\{v_1 = 1, v_2, \ldots, v_n = u\}$. We can have any direction on the edges of $t(1,u)$. Recall the distribution of A_{u2}:

$$\mathcal{L}(A_{u2}) = \sum_{j \in \mathrm{An}(u)} b_{uj} \delta_{\{b_{2j}/b_{uj}\}}.$$

We have $b_{2j} = 0$ for all $j \notin \mathrm{An}(2)$. We claim that $\mathrm{An}(u) \cap \mathrm{An}(2) = \emptyset$. According to property (P2) of a ttt, $\mathcal{T}$ does not contain an undirected cycle involving several tournaments. This means that it is impossible to find a node from which there leave paths to u and to 2. Also it is not possible to find a path passing through 3 and going to 2, because otherwise there would be either an undirected cycle involving several tournaments, or a cycle within a tournament. Both are impossible for a ttt. This leads to the conclusion that A_{u2} is degenerate at zero. Now we look at the variables $(M_{v_{i+1},v_i}, i = n-1,\ldots,1; M_{13}, M_{32})$ which we take by construction to be independent as they belong to different tournaments. Each of them is one of the variables in Lemma 4.6.3, and none of these is degenerate at zero. Hence their product cannot be degenerate at zero too.

Case 3.(c-ii). Next we consider the second case, when w.l.o.g. $(1,3) \in t(u,1)$ and $(2,3) \in t(u,2)$, see Figure 4.5(b). Let the trail from node 3 to u be on nodes $\{v_1 = 3, v_2, \ldots, v_n = u\}$. First we consider the case when we have at least one $i = 1,\ldots,n-1$ for which $(v_{i+1}, v_i) \in E$, i.e., we have at least one edge with direction from u to 3. Because $t(u,3)$ is a shortest trail, the edges incident to the nodes on the trail belong to different tournaments. The distribution of (A_{u1}, A_{u2}) is given by

$$\mathcal{L}(A_{u1}, A_{u2}) = \sum_{j \in \mathrm{An}(u)} b_{uj} \delta_{\left\{\frac{b_{1j}}{b_{uj}}, \frac{b_{2j}}{b_{uj}}\right\}},$$

where $b_{1j} = 0$ and $b_{2j} = 0$ if $j \notin \mathrm{An}(1)$ and $j \notin \mathrm{An}(2)$ respectively. When for some $i = 1,\ldots,n-1$ we have $(v_{i+1}, v_i) \in E$ then necessarily $\mathrm{An}(1) \cap \mathrm{An}(u) = \emptyset$

and $\mathrm{An}(2) \cap \mathrm{An}(u) = \emptyset$. There cannot be a path from $\mathrm{An}(1)$ or $\mathrm{An}(2)$ to any of the nodes $\{v_2, \ldots, v_n = u\}$, because otherwise there would be a cycle involving several tournaments in contradiction to the definition of a ttt. Because of the edge $(v_{i+1}, v_i) \in E$ all nodes in $\mathrm{An}(1) \cup \mathrm{An}(2)$ are not ancestors of u. And also because of the edges $(1,3), (2,3) \in E$ all nodes in $\mathrm{An}(u)$ cannot be ancestors of nodes 1 or 2. Thus when for some $i = 1, \ldots, n-1$ there is a directed edge $(v_{i+1}, v_i) \in E$ we have $\mathcal{L}(A_{u1}, A_{u2}) = \delta_{\{0,0\}}$. We have found a node $v \in V$ such that $A_{uv} = 0$ almost surely, but then the factorisation (4.16)–(4.17) cannot hold, because these never degenerate at zero.

Now let the trail from node 3 to u be actually a path. Let also nodes 1 and 2 be sources with respect to the tournaments shared with node 3, say $1,3 \in V_{\tau_1}$ and $2,3 \in V_{\tau_2}$. It is always possible to choose 1 and 2 in such a way they are the sources of τ_1 and τ_2. This is because node 3 obviously is not a source in τ_1 and τ_2, so the sources of these must point to 3. We can decompose $\mathrm{An}(u)$ into three disjoint sets, $\mathrm{An}(1)$, $\mathrm{An}(2)$ and the rest, $\mathrm{An}(u) \setminus \{\mathrm{An}(1) \cup \mathrm{An}(2)\}$. For the distribution of (A_{u1}, A_{u2}) we have

$$\begin{aligned}\mathcal{L}(A_{u1}, A_{u2}) &= \sum_{j \in \mathrm{An}(u)} b_{uj} \delta_{\left\{\frac{b_{1j}}{b_{uj}}, \frac{b_{2j}}{b_{uj}}\right\}} \\ &= \sum_{j \in \mathrm{An}(1)} b_{uj} \delta_{\left\{\frac{b_{1j}}{b_{uj}}, \frac{b_{2j}}{b_{uj}}\right\}} + \sum_{j \in \mathrm{An}(2)} b_{uj} \delta_{\left\{\frac{b_{1j}}{b_{uj}}, \frac{b_{2j}}{b_{uj}}\right\}} \\ &\quad + \sum_{j \in \mathrm{An}(u) \setminus \{\mathrm{An}(1) \cup \mathrm{An}(2)\}} b_{uj} \delta_{\left\{\frac{b_{1j}}{b_{uj}}, \frac{b_{2j}}{b_{uj}}\right\}}.\end{aligned}$$

For the atoms in the first summation we have

$$\frac{b_{1j}}{b_{uj}} = \frac{c_{p(j,1)} b_{jj}}{c_{p(j,u)} b_{jj}} = \frac{c_{p(j,1)}}{c_{p(j,1)} c_{13} c_{p(3,u)}} = \frac{1}{c_{13} c_{p(3,u)}}$$

and $b_{2j}/b_{uj} = 0$ as $b_{2j} = 0$ for all $j \in \mathrm{An}(1)$. Hence we have an atom that does not depend on $j \in \mathrm{An}(1)$, i.e., $\left(1/(c_{13} c_{p(3,u)}), 0\right)$ and its mass is

$$\sum_{j \in \mathrm{An}(1)} b_{uj} = \sum_{j \in \mathrm{An}(1)} c_{p(j,1)} c_{13} c_{p(3,u)} b_{jj} = c_{13} c_{p(3,u)} = c_{p(1,u)}.$$

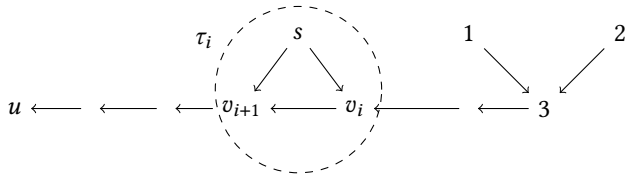
In a similar way, from the second summation in the last display we have an atom $\left(0, 1/(c_{23} c_{p(3,u)})\right)$ with mass $c_{23} c_{p(3,u)} = c_{p(2,u)}$. In the third summation term the atom is $(0,0)$ as $b_{1j} = b_{2j} = 0$ for all $j \in \mathrm{An}(u) \setminus \{\mathrm{An}(1) \cup \mathrm{An}(2)\}$ and its mass is $1 - c_{13} c_{p(3,u)} - c_{23} c_{p(3,u)} = 1 - c_{p(3,u)}(c_{13} + c_{23})$. Consider now the multiplicative increments $(M_{31}; M_{32}, M_{v_{i+1}, v_i} i = 1, \ldots, n-1)$ which are mutually independent since they belong to different tournaments. Because node 1 is a source node in the tournament τ_1 the distribution of M_{31} is $c_{13}\delta_{\{1/c_{13}\}} + (1 - c_{13})\delta_{\{0\}}$ by Lemma 4.6.3-2.(a). Similarly for M_{32}. We have

$$\begin{aligned}&\mathbb{P}\left(M_{31} \prod_{i=1}^{n-1} M_{v_{i+1}, v_i} = 0, M_{32} \prod_{i=1}^{n-1} M_{v_{i+1}, v_i} = 0\right) = 1 - \prod_{i=1}^{n-1} \mathbb{P}(M_{v_{i+1}, v_i} > 0) \\ &+ \mathbb{P}(M_{31} = 0)\, \mathbb{P}(M_{32} = 0) \qquad (4.30) \\ &- \left(1 - \prod_{i=1}^{n-1} \mathbb{P}(M_{v_{i+1}, v_i} > 0)\right) \mathbb{P}(M_{31} = 0)\, \mathbb{P}(M_{32} = 0).\end{aligned}$$

After some rearranging of the expression above we obtain

$$1 - \prod_{i=1}^{n-1} \mathbb{P}(M_{v_{i+1}, v_i} > 0)(c_{13} + c_{23} - c_{13} c_{23}). \qquad (4.31)$$

There are two further sub-cases: either all nodes $v_1, \dots, v_{n-1}$ are source nodes with respect to the tournament involving the next node in the sequence, or not. In the first sub-case, namely when all nodes in $\{v_1 = 3, v_2, \dots, v_{n-1}\}$ are source nodes with respect to the tournament involving the next node in the sequence, then $\mathbb{P}(M_{v_{i+1},v_i} > 0) = c_{v_i,v_{i+1}}$ for $i = 1, \dots, n-1$. This means that the probability in (4.31) and accordingly in (4.30) equals $1 - c_{p(3,u)}(c_{13} + c_{23} - c_{13}c_{23})$, which is different than $\mathbb{P}(A_{u1} = 0, A_{u2} = 0) = 1 - c_{p(3,u)}(c_{13} + c_{23})$. In the second sub-case, i.e., if at least one node from $\{v_1 = 3, v_2, \dots, v_{n-1}\}$ is not source with respect to the tournament involving the next node in the sequence then the possible values for $M_{31}\prod_{i=1}^{n-1} M_{v_{i+1},v_i}$ are not only $\{0, 1/c_{p(1,u)}\}$, which are the only possible values of A_{u1} as we showed in the previous paragraph. Let $i \in \{1, \dots, n-1\}$ be such that node v_i is not the source node in the tournament shared with v_{i+1}, say τ_i. This is depicted in the following graph:



Recall the distribution of M_{v_{i+1},v_i} from Lemma 4.6.3-2.(b):

$$\mathcal{L}(M_{v_{i+1},v_i}) = \sum_{j \in \mathrm{An}(v_i)} b_{v_{i+1},j}\delta_{\{b_{v_i,j}/b_{v_{i+1},j}\}} + \sum_{j \in \mathrm{An}(v_{i+1}) \setminus \mathrm{An}(v_i)} \delta_{\{0\}}.$$

Take for instance a node, say s, a parent of v_i and accordingly in $\mathrm{An}(v_i)$. Then

$$\frac{b_{v_i,s}}{b_{v_{i+1},s}} = \frac{c_{sv_i}}{c_{sv_{i+1}}}$$

is a possible value of M_{v_{i+1},v_i} with positive probability, namely at least $b_{v_{i+1},s}$. Another possible positive value is for $j = v_i \in \mathrm{An}(v_i)$, namely

$$\frac{b_{v_i,v_i}}{b_{v_{i+1},v_i}} = \frac{1}{c_{v_i,v_{i+1}}}$$

with probability at least b_{v_{i+1},v_i}. The criticality assumption on edge weights guarantees $\frac{c_{sv_i}}{c_{sv_{i+1}}} \neq 1/c_{v_i,v_{v+1}}$. This means that the product $M_{31}\prod_{i=1}^{n-1} M_{v_{i+1},v_i}$ has at least two different positive values - one involving $\frac{c_{sv_i}}{c_{sv_{i+1}}}$ and another $1/c_{v_i,v_{i+1}}$. However A_{u1} has only one possible positive value. □

Proof of Proposition 4.3.2

Proof. Sufficiency. Assume $\mathcal{T}$ has a unique source. We need to show that, for any disjoint and nonempty sets A, B, S, we have $X_A \perp\!\!\!\perp X_B \mid X_S$, whenever S is a separator of A and B in the skeleton T of $\mathcal{T}$. We would like to use Theorem 5.15 in Améndola et al. (2022), by which we need to show $A \perp_* B \mid S$ in $\mathcal{D}_S^*$, that is, there are no $*$-connecting paths between any pair of nodes in A and B in the *conditional reachability* DAG $\mathcal{D}_S^*$. We will explain these notions further.

Let $A, B, S \subset V$ be nonempty disjoint node sets, such that S is a separator of A and B in the skeleton T. Consider Figure 4.6. According to Definition 5.4 of Améndola et al. (2022), a $*$-connecting path between $a \in A$ and $b \in B$ is one of the five configurations therein. Our goal is to show that for $a \in A$ and $b \in B$ it is impossible to find a $*$-connecting path in a certain graph $\mathcal{D}_S^*$, which is not $\mathcal{T}$, neither T, but constructed under particular rules given in Améndola et al. (2022, Definition 5.1).

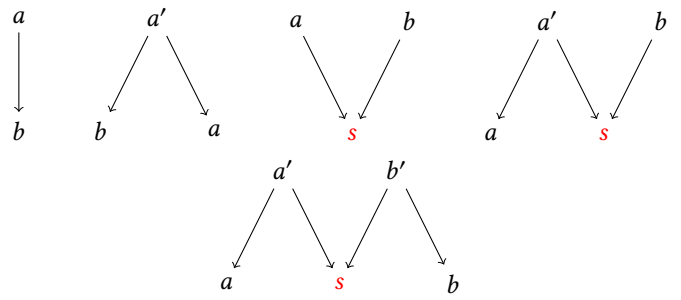


Figure 4.6: According to Definition 5.4 of Améndola et al. (2022), a $*$-connected path between a and b relative to S is one of the five configurations above. In the last three graphs we have $s \in S$.

According to this definition, the conditional reachability graph $\mathcal{D}_S^*$ is on the same vertex set, V. Between two nodes i and j in V there is an edge (i, j) in $\mathcal{D}_S^*$ if and only if there is a directed path from i to j in $\mathcal{T}$ such that no node on that path belongs to S, except possibly for i and j themselves.

Consider Figure 4.6. We need to show that in the conditional reachability graph $\mathcal{D}_S^*$, there is no $*$-connecting path between a node $a \in A$ and a node $b \in B$.

To obtain the first configuration in $\mathcal{D}_S^*$, there must be, in the skeleton T, nodes $a \in A$ and $b \in B$ such that no node on the path from a to b passes through S. But this is impossible, because we assumed that S is a separator of A and B in T. Similarly for the second configuration.

For the other three configurations in Figure 4.6 consider Figure 4.7.

In Figure 4.7, the left-hand and right-hand trails in the original graph $\mathcal{T}$ are the only possible one that give rise to the middle path in Figure 4.6 with respect to the graph $\mathcal{D}_S^*$: both the left-hand and right-hand graphs in Figure 4.7 show existing trails between a and b in $\mathcal{T}$, trails composed of a directed path from a to s, and a directed path from b to s. The only node on these trails which belongs to S is s. Hence in $\mathcal{D}_S^*$ we put a directed edge from a to s and from b to s. This gives the third $*$-connecting path in Figure 4.6. But this configuration cannot occur, for the following reason. On the left-hand trail in Figure 4.7, the separator node s has parents u_r and v_q in different tournaments. This leads to a v-structure between the nodes u_r, s, v_q, in contradiction to Lemma 4.6.1-2 and the hypothesis that $\mathcal{T}$ has a unique source.

On the right-hand trail in Figure 4.7, the node s shares a tournament with its parents u_r and v_q, but only s belongs to S; on the trail

$$\{a = u_1, u_2, \ldots, u_r, v_q, \ldots, v_2, v_1 = b\}$$

none of the nodes are in S. In T, this means that there is a path between A and B that does not pass through S. This is in contradiction to the assumption that S separates A and B in T.

To show that the fourth type of $*$-connecting path in Figure 4.6 cannot occur, we can use the reasoning used for the third one by setting $a = a'$ in Figure 4.7. Then either s has parents from two different tournaments or there is a non-directed path from a' to b which does not pass through S. The first case is excluded by Lemma 4.6.1-2 and the assumption that $\mathcal{T}$ has a unique source, and the second one by the assumption that S is a separator of A and B in T. The impossibility of the fifth $*$-connected configuration follows analogously.

Necessity. We will show that if $\mathcal{T}$ has multiple source nodes, there is a triple of disjoint, non-empty sets $A, B, S \subset V$ such that S is a separator of A and B in T, but X_A and X_B are conditionally dependent given X_S. In case $\mathcal{T}$ has at least two sources,

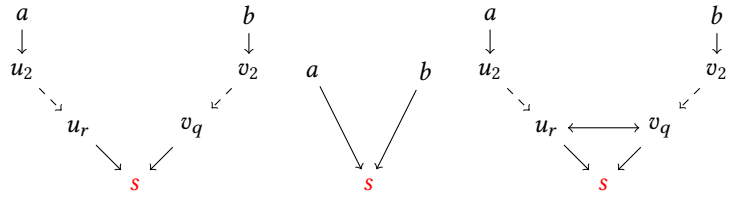


Figure 4.7: The left and right trails in the original graph $\mathcal{T}$ are the only possible trails that give rise to the middle path in the graph $\mathcal{D}_S^*$.

we have at least one v-structure in $\mathcal{T}$ by Lemma 4.6.1-2. Take a triple of nodes in a v-structure, say u, v, w, with u and w being parents of v. Then node v separates nodes u and w in T, i.e., $S = \{v\}$ is a separator of $A = \{u\}$ and $B = \{w\}$ in T. All references below are from Améndola et al. (2022).

To show $X_u \not\perp\!\!\!\perp X_w \mid X_v$ we will use Theorem 6.18 (Context free completeness) of Améndola et al. (2022). We need to show that there is an *effective* $*$-connecting path in the critical DAG $\mathcal{D}_S^*(\theta)$ between nodes u and w as in their Definitions 5.2 and 6.5.

The subgraph on nodes u, v, w of $\mathcal{D}_S^*(\theta)$ is a v-structure, $u \longrightarrow v \longleftarrow w$, according to the definition of $\mathcal{D}_S^*(\theta)$. According to Definition 6.4, the $|S|\times|S|$ substitution matrix of $(u, v) \in E$ relative to $S = \{v\}$ is zero, because S is a singleton and by definition all diagonal entries of the substitution matrix are zero, i.e., $\Xi_S^{vu} = 0$. Similarly, $\Xi_S^{vw} = 0$. Because the edges $(u, v), (w, v)$ form a $*$-connecting path between u, w in $\mathcal{D}_S^*(\theta)$, the substitution matrix of this path relative to S, say Ξ_S, is zero too:

$$\Xi_S = \max(\Xi_S^{vu}, \Xi_S^{vw}) = 0.$$

To find out if the edges $(u, v), (w, v)$ form an *effective* $*$-connecting path between u, w, we need to compute the tropical eigenvalue of $\max(\Gamma_{SS}, \Xi_S)$ where Γ is as in Equation (2.3) in Améndola et al. (2022) and Γ_{SS} is the vv-element of Γ, i.e., $\{\Gamma\}_{vv}$. Because $\{\Gamma\}_{ij} > 0$ if and only if there is a directed path from j to i, we have $\Gamma_{SS} = \{\Gamma\}_{vv} = 0$ and so

$$\max(\Gamma_{SS}, \Xi_S) = 0.$$

The tropical eigenvalue (Améndola et al., 2022, Equation (2.7)) of the above matrix is trivially equal to zero and thus smaller than one. By Definition 6.5 in the cited reference, there is indeed an *effective* $*$-connecting path between u, w in $\mathcal{D}_S^*(\theta)$. In view of their Theorem 6.18, we conclude $X_u \not\perp\!\!\!\perp X_w \mid X_v$.

□

4.6.3 Proofs and additional results for Section 4.4

Proof of Lemma 4.4.1. The point masses satisfy $m_i > 0$ for all $i \in V$ because we have $m_i = 0$ if and only if $c_{ii} = 0$. However $c_{ii} = 0$ is impossible in view of the definition in (4.6). Therefore we cannot have undefined atoms, which would happen when $m_i = 0$. This shows (i).

Next we show (ii). To see why $a_i \neq a_j$ for $i \neq j$, let $i, v \in V$ and recall b_{vi} in (4.9). From the line below (4.12), recall that we also have $b_{vi} = m_i a_{vi}$ for $i, v \in V$. Thanks to the assumption $\theta \in \mathring{\Theta}_*$, we have (4.18). We also have for any DAG (4.19).

The combination of the last two equations implies that in (4.12), all vectors a_i for $i \in V$ are different and thus that H_θ has $|V|$ distinct atoms. Also, for every node $i \in V$, we can find out which of the $|V|$ atoms of H_θ is a_i because it is the unique one that satisfies $\mathrm{Desc}(i) = \{v \in V : a_{vi} > 0\}$. Note that similarly, among the $|V|$ vectors in the set $\mathcal{B}_\theta = \{(b_{vj})_{v\in V} : j \in V\}$, the vector b_i is the unique one such that $\mathrm{Desc}(i) = \{v \in V : b_{vi} > 0\}$.

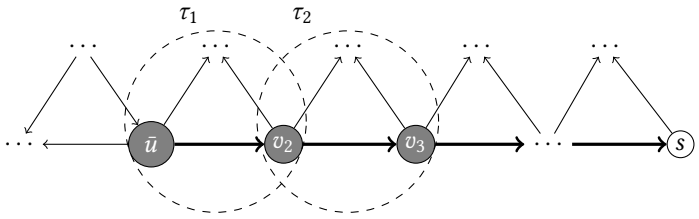


Figure 4.8: A unique path on nodes $\{\bar{u} = v_1, v_2, v_3, \ldots, s = v_n\}$ under Lemma 4.6.3.1. Each of the nodes $\bar{u}, v_2, \ldots, v_{n-1}$ belongs to $\bar{U}$. Each of the nodes $v_2, \ldots, v_{n-1}, s$ has a unique parent. The node $\bar{u} \in \bar{U}$ may have parents as illustrated here, but then there is at least one tournament with respect to which it is a source node, e.g., τ_1. Let v_2 be the node with unique parent $\bar{u}$ in τ_1. When v_2 belongs to $\bar{U}$, it must participate in at least one another tournament, say τ_2. In τ_2 the node with unique parent v_2 is v_3. In this principle the path continues until we find a node in U, which is $v_n = s$ in this case.

Finally consider (iii). By the criticality assumption, every edge is critical, because it is the shortest path between any pair of adjacent nodes. Since $(i, v) \in E$ is critical, we have $b_{vi} = b_{ii}c_{iv}$ and thus $c_{iv} = b_{vi}/b_{ii} = a_{vi}/a_{ii}$.

In summary, the angular measure H_θ possesses $|V|$ distinct atoms that can be uniquely matched to the nodes. As a consequence, we can reconstruct the matrix $(b_{vi})_{i,v\in V}$. Thanks to (iii), this matrix allows us to recover all edge weights c_{vi}. □

Lemma 4.6.3.1. *Let $\mathcal{T} = (V, E)$ be a ttt as in Definition 4.2.1 and let $\mathcal{T}$ have a unique source, u_0. Let $U \subset V$ be non-empty and suppose that $\bar{U} = V \setminus U$ satisfies conditions (I1) and (I2). For every $\bar{u} \in \bar{U}$ there exists $s \in \mathrm{desc}(\bar{u}) \cap U$ such that $\pi(\bar{u}, s)$ is a singleton and the unique path p from $\bar{u}$ to s satisfies the following two properties:*

1. *all nodes on p except for s are in $\bar{U}$;*
2. *all nodes on p except possibly for $\bar{u}$ have only one parent.*

As a consequence, any path with destination s must either start in one of the nodes of p or contain p as a sub-path.

Proof. Let $\bar{u} \in \bar{U}$ and suppose $\bar{u}$ has no parents, so $\bar{u} = u_0$. Take a node whose unique parent is $\bar{u}$, say v_2. By Harary and Moser (1966, Corollary 5.a) such a node exists in every tournament in which $\bar{u}$ takes part. If $v_2 \in U$ then $s = v_2$ and we are done. If $v_2 \in \bar{U}$ then by (I2) v_2 must be a source of at least one another tournament. In each of these, there is a node whose only parent is v_2. Take such a node, say v_3. If $v_3 \in U$ then $v_3 = s$ and we are done; otherwise continue in the same way until we find a node which is in U. Because the graph is finite and because of condition (I2) such a node must exist. It is clear that the path constructed in this way has the stated properties.

Next suppose that $\bar{u}$ belongs to $\bar{U}$ and that $\bar{u}$ has at least one parent. By (I2) it must be a source of at least one tournament In each of these tournaments there is a node with single parent $\bar{u}$. Take one of them, say v_2, and if $v_2 \in U$ then we are done, otherwise repeat the same procedure as above until we find a node which is in U. Because the graph is finite and because of condition (I2) such a node must exist. It is clear that the path constructed in this way has the stated properties too. □

Lemma 4.6.3.2. *Let $\mathcal{T} = (V, E)$ be a ttt as in Definition 4.2.1 and let $\mathcal{T}$ have a unique source, u_0. Let $U \subset V$ be non-empty and suppose that $\bar{U} = V \setminus U$ satisfies conditions (I1) and (I2). Let $i, j \in V$ be two distinct nodes. Upon switching the roles of i and j if needed, the equality $\mathrm{Desc}(i) \cap U = \mathrm{Desc}(j) \cap U$ implies the following properties:*

1. *$i \in \bar{U}$;*
2. *$\mathrm{desc}(i) = \mathrm{Desc}(j)$;*
3. *$\{i\} = \mathrm{pa}(j)$;*

4. *there exists* $u \in V$ *such that* $i, j \in \mathrm{pa}(u)$*;*
5. *for* $k \in V \setminus \{i, j\}$*, the set* $\mathrm{Desc}(k) \cap U$ *is different from* $\mathrm{Desc}(i) \cap U = \mathrm{Desc}(j) \cap V$*;*
6. $|\mathrm{Desc}(i) \cap U| = |\mathrm{Desc}(j) \cap U| \geq 2$.

Proof. 1. Note first that $\mathrm{Desc}(i) \cap U$ cannot be empty, for otherwise, we would have $\mathrm{Desc}(i) \subseteq \bar{U}$, but this is impossible, since $\mathrm{Desc}(i)$ contains at least one leaf node (a node without children), in contradiction to (I1).

Since $\mathrm{Desc}(i) \cap U = \mathrm{Desc}(j) \cap U$ and since this set is non-empty, the intersection $\mathrm{Desc}(i) \cap \mathrm{Desc}(j)$ is not empty too. In relation to Lemma 4.6.1-3 this means either $\mathrm{Desc}(i) \subseteq \mathrm{Desc}(j)$ or $\mathrm{Desc}(j) \subseteq \mathrm{Desc}(i)$. In the remainder of the proof, we suppose $\mathrm{Desc}(j) \subseteq \mathrm{Desc}(i)$. Then we must have $i \notin \mathrm{Desc}(j)$, since otherwise also $\mathrm{Desc}(i) \subseteq \mathrm{Desc}(j)$ and thus $\mathrm{Desc}(i) = \mathrm{Desc}(j)$, which is impossible since i and j are distinct; see (4.19). From $\mathrm{Desc}(j) \subseteq \mathrm{Desc}(i)$ and $\mathrm{Desc}(i) \cap U = \mathrm{Desc}(j) \cap U$ it follows that

$$\mathrm{Desc}(i) \setminus \mathrm{Desc}(j) \subseteq \bar{U}.$$

Because $i \notin \mathrm{Desc}(j)$ we get $i \in \bar{U}$.

2–4. First we show that all elements in $\mathrm{Desc}(i) \setminus \mathrm{Desc}(j)$ are ancestors of j. Let $v \in \mathrm{Desc}(i) \setminus \mathrm{Desc}(j)$. Because j and v are two different nodes, Lemma 4.6.1-3 implies that one of three cases must occur: $\mathrm{Desc}(v) \subset \mathrm{Desc}(j)$; $\mathrm{Desc}(j) \subset \mathrm{Desc}(v)$; or $\mathrm{Desc}(j) \cap \mathrm{Desc}(v) = \emptyset$. The first case, $\mathrm{Desc}(v) \subset \mathrm{Desc}(j)$, is impossible, since $v \notin \mathrm{Desc}(j)$. The third case, $\mathrm{Desc}(j) \cap \mathrm{Desc}(v) = \emptyset$, is impossible too, since it would imply that $\mathrm{Desc}(v) \subseteq \mathrm{Desc}(i) \setminus \mathrm{Desc}(j) \subseteq \bar{U}$, but this cannot happen since $\mathrm{Desc}(v)$ contains at least one leaf node while $\bar{U}$ does not contain any. Only the second case, $\mathrm{Desc}(j) \subset \mathrm{Desc}(v)$, remains. As a consequence, v is an ancestor of j, and so all nodes of $\mathrm{Desc}(i) \setminus \mathrm{Desc}(j)$ are ancestors of j. By the proof of point 1, we get

$$\mathrm{Desc}(i) \setminus \mathrm{Desc}(j) \subseteq \mathrm{an}(j) \cap \bar{U}. \tag{4.32}$$

Let again $v \in \mathrm{Desc}(i) \setminus \mathrm{Desc}(j)$. We show that there exists a unique path from v to j and that j has only a single parent. Since $v \in \bar{U}$, there exists, by Lemma 4.6.3.1, a node $s(v) \in U$ such that there is only directed path $p(v, s(v))$ from v to $s(v)$; moreover, this path satisfies properties 1 and 2 of the statement. Necessarily,

$$s(v) \in \mathrm{Desc}(v) \cap U \subseteq \mathrm{Desc}(i) \cap U = \mathrm{Desc}(j) \cap U.$$

As $s(v)$ is a descendant of j while j is a descendant of v, the path $p(v, s(v))$ passes by j. As a consequence, there is a unique path $p(v, j)$ from v to j; otherwise, there would be more than one path from v to $s(v)$. Moreover, by Lemma 4.6.3.1-2, all nodes on the path $p(v, s(v))$, except possibly for v, have only one parent. In particular, j has only one parent.

Take again any $v \in \mathrm{Desc}(i) \setminus \mathrm{Desc}(j)$. By (I1), v has at least two children. They cannot both be ancestors of j, since then there would be two paths from v to j, in contradiction to the previous paragraph. Let u be a child of v that is not an ancestor of j; then $u \in \mathrm{Desc}(j)$ because of (4.32). This means there are two paths from v to u: the edge (v, u) and a path passing through j. These paths must belong to the same tournament, as the skeleton of $\mathcal{T}$ is a block graph. But then v and j are adjacent, and thus v, which we already knew to be an ancestor of j, is actually a parent of j. But j has only one parent, and so the set $\mathrm{Desc}(i) \setminus \mathrm{Desc}(j)$ must be a singleton. As this set obviously contains node i, we get $v = i$ and thus $\mathrm{Desc}(i) \setminus \mathrm{Desc}(j) = \mathrm{pa}(j) = \{i\}$. Since $\mathrm{Desc}(i) = \{i\} \cup \mathrm{desc}(i)$ and $\mathrm{Desc}(j) \subset \mathrm{Desc}(i)$, it follows that $\mathrm{desc}(i) = \mathrm{Desc}(j) = \mathrm{Desc}(i) \setminus \{i\}$.

5. Let $k \in V \setminus \{i, j\}$ be such that $\mathrm{Desc}(k) \cap U = \mathrm{Desc}(i) \cap U = \mathrm{Desc}(j) \cap U$. By point 3 we have $\{i\} = \mathrm{pa}(j)$. From $\mathrm{Desc}(k) \cap U = \mathrm{Desc}(i) \cap U$ we can have

either $\{k\} = \mathrm{pa}(i)$ or $\{i\} = \mathrm{pa}(k)$, while from $\mathrm{Desc}(k) \cap U = \mathrm{Desc}(j) \cap U$ we have either $\{k\} = \mathrm{pa}(j)$ or $\{j\} = \mathrm{pa}(k)$. Because already $\{i\} = \mathrm{pa}(j)$, we cannot also have $\{k\} = \mathrm{pa}(j)$, whence we must have $\{j\} = \mathrm{pa}(k)$. But then $\{i\} = \mathrm{pa}(k)$ is impossible, so that necessarily $\{k\} = \mathrm{pa}(i)$. From $\{i\} = \mathrm{pa}(j)$, $\{j\} = \mathrm{pa}(k)$, and $\{k\} = \mathrm{pa}(i)$ we get a cycle between the three nodes i, j, k which is a contradiction to the definition of a DAG.

6. We already know that $\mathrm{Desc}(i) \cap U = \mathrm{Desc}(j) \cap U$. We need to show that this set contains at least two elements. Consider the triple $\{i, j, u\}$ from point 4 that forms a triangle with directed edges (i, j), (i, u), and (j, u). By point 1 we have also $i \in \bar{U}$. There are four cases, according to whether j and u belong to U or not.

- If $j, u \in U$, they are two distinct elements of $\mathrm{Desc}(j) \cap U$.
- If $j \in U$ but $u \in \bar{U}$, then take $r \in \mathrm{Desc}(u) \cap U$ [which is non-empty by (I1): all leaf nodes in $\mathrm{Desc}(u)$ are in U], and note that j and r are two distinct elements in $\mathrm{Desc}(j) \cap U$.
- If $j \in \bar{U}$ and $u \in U$, then, as in Lemma 4.6.3.1, let $s \in U$ be such that there is unique path $p(j, s)$ from j to s, this path satisfying properties 1-2 in the same lemma. Then u does not belong to that path (since $u \in U$ and u has at least two parents, i and j), so that s is different from u, and both are members of $\mathrm{Desc}(j) \cap U$.
- If $j, u \in \bar{U}$, then we can find by Lemma 4.6.3.1 nodes $s \in \mathrm{desc}(j) \cap U$ and $r \in \mathrm{desc}(u) \cap U$ with paths $p(j, s)$ and $p(u, r)$ which satisfy the characteristics in this lemma. Nodes s, r clearly belong to $\mathrm{Desc}(j) \cap U$. Moreover, they are distinct: node u, having at least two parents, cannot belong to the unique path $p(j, s)$ between j and s, while by construction, there is a directed path from j to r that passes along u. □

Lemma 4.6.3.3. *Let X be a max-linear model with respect to a ttt with unique source. The coefficient b_{vv} depends only on the edge weights of the tournament shared by node $v \in V$ and its parents, and it is given by*

$$b_{vv} = 1 + \sum_{u \in \mathrm{pa}(v)} \sum_{p \in \pi(u,v)} (-1)^{|p|} c_p. \tag{4.33}$$

Proof. Consider the node v. If v has no parents we have $\mathrm{an}(v) = \emptyset$ and by (4.6) we have $b_{vv} = 1$. If v has at least one parent then there is a tournament which contains the parents, say, $\tau = (V_\tau, E_\tau)$. Let the nodes in τ be labelled according to their in-/out-degree ordering in τ: the node with $|V_\tau| - 1$ children in τ (the source of τ) has index 1, the node with $|V_\tau| - 2$ children in τ has index 2, and so on. We can partition the set $\mathrm{an}(v)$ into $\mathrm{An}(1)$ and $\mathrm{pa}(v) \setminus \{1\}$. For $i \in \mathrm{An}(1)$ the shortest path from i to v passes necessarily through 1, so $b_{vi} = c_{p(i,1)} c_{1v} b_{ii}$. Then we have by (4.4) and (4.6)

$$\begin{aligned} b_{vv} &= 1 - \sum_{i \in \mathrm{an}(v)} b_{vi} = 1 - \sum_{i \in \mathrm{An}(1)} c_{p(i,1)} c_{1v} b_{ii} - \sum_{i \in \mathrm{pa}(v) \setminus 1} b_{vi} \\ &= 1 - c_{1v} - \sum_{i \in \mathrm{pa}(v) \setminus 1} c_{iv} b_{ii}. \end{aligned} \tag{4.34}$$

Let $C = \{c_{ij}\}_{i,j \in V_\tau, i<j}$ be the matrix of edge weights within τ: it is lower triangular and has zero diagonal. Let I_m denote the $m \times m$ identity matrix, write $\boldsymbol{b} = (1, b_{22}, \ldots, b_{|V_\tau|,|V_\tau|})^\top$ (a column vector) and let $1_{|V_\tau|}$ be a column vector of ones of length $|V_\tau|$. Consider the system of linear equations

$$(I_{|V_\tau|} + C)\, \boldsymbol{b} = 1_{|V_\tau|}. \tag{4.35}$$

For $v \geq 2$ the expression in (4.34) is equivalent to the v-th equation in (4.35). A solution for $\boldsymbol{b}$ is

$$\boldsymbol{b} = (I_{|V_\tau|} + C)^{-1} 1_{|V_\tau|} = \big(I_{|V_\tau|} - (-C)\big)^{-1}\ 1_{|V_\tau|}.$$

From the equality

$$\big(I_{|V_\tau|} + (-C) + (-C)^2 + \cdots + (-C)^k\big)\big(I_{|V_\tau|} - (-C)\big) = I_{|V_\tau|} - (-C)^{k+1}$$

and the fact that for $|V_\tau|$-square lower triangular matrix with zero diagonal powers of $k \geq |V_\tau|$ are zero matrices we obtain

$$\big(I_{|V_\tau|} + (-C) + (-C)^2 + \cdots + (-C)^{|V_\tau|-1}\big) = \big(I_{|V_\tau|} - (-C)\big)^{-1}.$$

If the matrix on the left is denoted by K we have as solution $\boldsymbol{b} = K1_{|V_\tau|}$. For all $b_{vv}, v \geq 2$, it can be shown that (4.33) equals the vv-th element of this solution for $\boldsymbol{b}$. For b_{11} consider the corresponding solution when the tournament τ is the one which node 1 shares with its parents. If node 1 has no parents in the ttt, then we have the solution $b_{11} = 1$ which is indeed the case. □

Lemma 4.6.3.4. *Let X follow a max-linear model as in Assumption 4.2.1 with respect to a ttt $\mathcal{T}$ consisting of a single tournament $\tau = (V, E)$. If the node $v \in V$ has at least one parent, then the parameter vector $\theta = (c_e)_{e\in E} \in \mathring{\Theta}_*$ is not identifiable from the distribution of $X_{V\setminus v}$. Specifically, there exists $\theta' = (c'_e)_{e\in E} \in \mathring{\Theta}_*$ such that $\theta' \neq \theta$ and the distribution of $X_{V\setminus v}$ is the same under θ' as under θ.*

Proof. Let $n = |V|$ denote the number of nodes. For convenience, rename the nodes to $V = \{1, \ldots, n\}$ in the ordering induced by the DAG, i.e., node i has $i - 1$ parents, for $i \in V$. The number of edges is $|E| = n(n-1)/2 =: m$, and the parameter set $\mathring{\Theta}_*$ is an open subset of $\mathbb{R}^E$. The distribution of X is max-linear and is given by

$$X_j = \bigvee_{i=1}^{j} b_{ji} Z_i, \qquad j \in V, \tag{4.36}$$

where $b_{11} = 1$, $b_{jj} = 1 - \sum_{i=1}^{j-1} b_{ji}$ for $j \in V \setminus 1$, and where the m coefficients $b = (b_{ji} : 1 \leq i < j \leq n)$ are determined by the edge parameters $\theta = (c_{ij} : 1 \leq i < j \leq n)$.

Discarding the variable X_v for some $v \in V\setminus 1$ yields the vector $X_{V\setminus v}$, the distribution of which is determined by the $m - (v-1)$ coefficients $(b_{ji} : 1 \leq i < j \leq n, j \neq v)$. For convenience, identify $\mathbb{R}^E$ with $\mathbb{R}^m$. Let $\pi : \mathbb{R}^m \to \mathbb{R}^{m-v+1}$ be the projection that sends $x = (x_{ij} : 1 \leq i < j \leq m)$ to $\pi(x) = (x_{ij} : 1 \leq i < j \leq m, j \neq v)$, i.e., the effect of π is to leave out the coordinates (i, v) with $i = 1, \ldots, v-1$. By (4.36) with $j = v$ removed, the distribution of $X_{V\setminus v}$ is determined by $\pi(b)$.

The max-linear coefficients b are a function of the edge parameters θ. Formally, there exists a map $f : \mathring{\Theta}_* \to \mathbb{R}^m$ such that

$$b = f(\theta).$$

The function f can be reconstructed from (4.9) with $p(i, j) = (i, j)$ for $1 \leq i < j \leq n$. Clearly, f is continuous. Since the parameter θ is identifiable from the distribution of X (Lemma 4.4.1), the function f is also injective, i.e., $\theta \neq \theta'$ implies $f(\theta) \neq f(\theta')$. By the Invariance of Domain Theorem (see, e.g. Kulpa, 1998), the image $f(\mathring{\Theta}_*)$ is therefore an open subset of $\mathbb{R}^m$. But then, for any coefficient vector $b \in f(\mathring{\Theta}_*)$, there exists another coefficient vector $b' \in f(\mathring{\Theta}_*)$ such that $b' \neq b$ but still $b_{ji} = b'_{ji}$ for all $1 \leq i < j \leq n$ and $j \neq v$ — in other words, such that $\pi(b) = \pi(b')$. Since f is injective,

the vectors b and b' originate from different edge parameter vectors $\theta = f^{-1}(b)$ and $\theta' = f^{-1}(b')$ in $\mathring{\Theta}_*$. But

$$\pi(f(\theta)) = \pi(b) = \pi(b') = \pi(f(\theta')),$$

so that the edge weight vectors θ and θ' induce the same distribution of $X_{V\setminus v}$. We conclude that the parameter θ is not identifiable from the distribution of $X_{V\setminus v}$. □

Proof of Proposition 4.4.2 When reading the proof, the following perspective may help. Recall the notation in equations (4.13) and (4.20). The knowledge of the (simple max-stable) distribution of X_U implies the knowledge of its angular measure H_U and thus of the unordered collection of pairs of atoms and masses (ω_r, μ_r) for $r = 1, \ldots, s$. The vector X_U can itself be represented as a max-linear model with s independent factors and coefficient vectors $\beta_r = \mu_r \omega_r$ for $r = 1, \ldots, s$. We first need to ensure that we can match those vectors β_r in a unique way to the max-linear coefficient vectors $(b_{vi})_{v\in U}$ for $i \in V$; note that the coordinates v of those vectors are restricted to U. Next, from the latter vectors, we need to recover the edge coefficients $\theta = (c_e)_{e\in E}$.

Proof of sufficiency (if) part of Proposition 4.4.2. We assume (I1) and (I2). In the first step of the proof we show that the angular measure of X_U in (4.20) is composed of $|V|$ distinct atoms and that we can associate every atom in $\{\omega_r : r = 1, \ldots, |V|\}$ to some node $v \in V$ and accordingly be able to associate it to one of the atoms $a_{i,U} = (b_{vi}/m_{i,U})_{v\in U}$ for $i \in V$. For this, we focus on the nature of the atoms $\{a_{i,U}\}$, given the conditions (I1) and (I2). As a consequence, the max-linear coefficient matrix $b_{U\times V} = (b_{vi})_{v\in U, i\in V}$ can be recovered from the distribution of X_U. In Step 2, we show how to recover from this matrix the edge parameters $\theta = (c_e)_{e\in E}$.

Step 1. Recall the representation $H_U = \sum_{i\in V} m_{i,U}\delta_{a_{i,U}}$ in (4.13) of the angular measure of X_U. We shall show that all $|V|$ masses $m_{i,U}$ are positive and that all $|V|$ atoms $a_{i,U}$ are distinct. Moreover, we will show how to match the atoms to the nodes, that is, given an atom $\omega \in \{\omega_r : r = 1, \ldots, |V|\}$ how to identify the node $i \in V$ such that $\omega = a_{i,U}$.

All $|V|$ vectors $\{a_{i,U}\}$ have positive masses $\{m_{i,U}\}$. Recall $m_{i,U} = \sum_{v\in U} b_{vi}$ and recall from (4.18) that $b_{vi} > 0$ if and only if $v \in \mathrm{Desc}(i)$. It follows that $m_{i,U} = 0$ if and only if $\mathrm{Desc}(i) \cap U = \emptyset$ or, in other words, $\mathrm{Desc}(i) \subseteq \bar{U}$. But this is impossible since $\mathrm{Desc}(i)$ contains at least one leaf node, that is, a node without children, and such a node belongs to U by (I1). We conclude that $m_{i,U} > 0$ for all $i \in V$.

All $|V|$ vectors $\{a_{i,U}\}$ are distinct. By (4.22) it follows that whenever for two different nodes $i, j \in V$ we have $\mathrm{Desc}(i) \cap U \neq \mathrm{Desc}(j) \cap U$ then we can find two atoms, say ω' and ω'', within the set $\{\omega_r\}$ such that $\omega' = a_{i,U}$ and $\omega'' = a_{j,U}$. Because $\mathrm{Desc}(i) \cap U \neq \mathrm{Desc}(j) \cap U$ then necessarily $a_{i,U} \neq a_{j,U}$. Suppose however for two different nodes $i, j \in V$ we have $a_{i,U} = a_{j,U}$. This means that for the u-th and j-th elements of these vectors we have

$$a_{i,u;U} = a_{j,u;U} \iff \frac{b_{ui}}{m_i} = \frac{b_{uj}}{m_j} \qquad \text{and} \qquad a_{i,j;U} = a_{j,j;U} \iff \frac{b_{ji}}{m_i} = \frac{b_{jj}}{m_j}$$

Considering the ratios above, we should also have

$$\frac{a_{i,u;U}}{a_{i,j;U}} = \frac{a_{j,u;U}}{a_{j,j;U}} \qquad \Longleftrightarrow \qquad \frac{b_{ui}}{b_{ji}} = \frac{b_{uj}}{b_{jj}}. \tag{4.37}$$

Because $a_{i,U} = a_{j,U}$ necessarily $\mathrm{Desc}(i) \cap U = \mathrm{Desc}(j) \cap U$. By Lemma 4.6.3.2 there exists a node u such that one of the edge sets $\{(i, j), (i, u), (j, u)\}$ or $\{(j, i), (i, u), (j, u)\}$

is contained in E. Without loss of generality, suppose this holds for the first triple. Also, by Lemma 4.6.3.2 there cannot be another node k with $\mathrm{Desc}(k) \cap U = \mathrm{Desc}(i) \cap U = \mathrm{Desc}(j) \cap U$.

Suppose first $j, u \in U$. From the identities

$$b_{ji} = c_{ij} b_{ii}, \qquad b_{ui} = c_{iu} b_{ii}, \qquad b_{uj} = c_{ju} b_{jj}, \tag{4.38}$$

and the criticality requirement

$$c_{iu} > c_{ij} c_{ju}$$

we have

$$\frac{b_{ui}}{b_{ji}} > \frac{b_{uj}}{b_{jj}}. \tag{4.39}$$

This inequality shows that (4.37) cannot happen, hence we cannot have $a_{i,U} = a_{j,U}$. This means that all $|V|$ atoms are distinct.

Next suppose $j, u \in \bar{U}$. By Lemma 4.6.3.1, there exist nodes $j', u' \in U$ such that $j' \in \mathrm{desc}(j)$ and $u' \in \mathrm{desc}(u)$ and the paths $p(j, j')$ and $p(u, u')$ satisfy the properties in the said lemma. Because all nodes on the path $p(j, j')$ except possibly for j have a unique parent, the path $p(j, j')$ cannot pass through u (which has parents i and j), and thus $j' \neq u'$. As i is a parent of j, the shortest (and in fact the only) path from i to j' is the one that concatenates the edge (i, j) with $p(j, j')$ (Lemma 4.6.3.1). It follows that

$$b_{j'i} = c_{p(j,j')} c_{ij} b_{ii} \qquad \text{and} \qquad b_{j'j} = c_{p(j,j')} b_{jj}. \tag{4.40}$$

By a similar argument, the path that concatenates the edge (i, u) with the path $p(u, u')$ is the unique shortest path from i to u', while the path that concatenates (j, u) with $p(u, u')$ is the unique shortest path from j to u'. It follows that

$$b_{u'i} = c_{p(u,u')} c_{iu} b_{ii} \qquad \text{and} \qquad b_{u'j} = c_{p(u,u')} c_{ju} b_{jj}. \tag{4.41}$$

Combining these equalities, $c_{iu} > c_{ij} c_{ju}$ implies that we should have

$$\frac{b_{u'i}}{b_{u'j}} > \frac{b_{j'i}}{b_{j'j}}. \tag{4.42}$$

However from $a_{i,U} = a_{j,U}$ we have for the u'-th and j'-th elements of these vectors

$$a_{i,u';U} = a_{j,u';U} \iff \frac{b_{u'i}}{m_i} = \frac{b_{u'j}}{m_j}$$

and

$$a_{i,j';U} = a_{j,j';U} \iff \frac{b_{j'i}}{m_i} = \frac{b_{j'j}}{m_j}.$$

Considering the ratios above, we should also have

$$\frac{a_{i,u';U}}{a_{i,j';U}} = \frac{a_{j,u';U}}{a_{j,j';U}} \qquad \iff \qquad \frac{b_{u'i}}{b_{u'j}} = \frac{b_{j'i}}{b_{j'j}}. \tag{4.43}$$

Because of (4.42) the equalities in (4.43) cannot happen, hence we cannot have $a_{i,U} = a_{j,U}$.

The analysis of the cases $(j, s) \in U \times \bar{U}$ and $(j, s) \in \bar{U} \times U$ is similar. This shows that all $|V|$ vectors $a_{i,U}$ for $i \in V$ are different and because of (4.21), all vectors ω_r for $r \in \{1, \ldots, |V|\}$ are different too.

Distinguishing all atoms $a_{i,U}$ with zeroes on the same positions. For two different nodes $i, j \in V$, the atoms $a_{i,U}$ and $a_{j,U}$ have the same supports $\{v \in U : a_{vi} > 0\} = \{v \in U : a_{vj} > 0\}$ when $\mathrm{Desc}(i) \cap U = \mathrm{Desc}(j) \cap U$. By Lemma 4.6.3.2-5 there cannot be any other node $k \in V \setminus \{i, j\}$ with the same descendants in U. In the representation

$$H_{\theta,U} = \sum_{t \in V} m_{t,U} \delta_{a_{t,U}} = \sum_{r=1}^{|V|} \mu_r \delta_{\omega_r}$$

there are thus exactly two atoms, ω and ω', say, with the same indices of non-zero coordinates as $a_{i,U} = (b_{vi}/m_{i,U})_{v \in V}$ and $a_{j,U} = (b_{vj}/m_{j,U})_{v \in V}$. The question is then how to know whether $\omega = a_{i,U}$ and $\omega' = a_{j,U}$ or vice versa, $\omega = a_{j,U}$ and $\omega' = a_{i,U}$. Let μ and μ' be the masses of ω and ω', respectively, and consider the vectors $\beta = \mu\omega$ and $\beta' = \mu'\omega'$. An equivalent question is then how to identify β and β' with the two max-linear coefficient vectors $(b_{vi})_{v \in U}$ and $(b_{vj})_{v \in U}$.

By Lemma 4.6.3.2-2 to 4, we can suppose that i is the unique parent of j and that i and j have a common child u. The analysis is now to be split up into different cases, according to whether j and u belong to U or not. Recall that $b_{vi} = c_{p(i,v)} b_{ii}$ and $b_{vj} = c_{p(j,v)} b_{jj}$ for $v \in \mathrm{Desc}(j) \subset \mathrm{Desc}(i)$.

Suppose first that $j, u \in U$. From (4.38) we deduce (4.39) thanks to the criticality assumption. In order to make the correct assignment of the two vectors $\beta = (\beta_v)_{v \in U}$ and $\beta' = (\beta'_v)_{v \in U}$ to the nodes i and j, we need to check the inequality (4.39). If $\beta_u/\beta_j > \beta'_u/\beta'_j$ then we assign the vector β to the node i and the vector β' to the node j. If the equality is reversed, we do the assignment the other way around.

Next suppose that $j, u \in \bar{U}$. According to Lemma 4.6.3.1, there exist nodes $j', u' \in U$ so that there is a unique path from j to j' and from u to u'. By Lemma 4.6.3.1, the paths from i to u' and j' and from j to j to u' are

$$\begin{aligned} p(i, u') &= \{(i, u)\} \cup p(u, u'), \\ p(i, j') &= \{(i, j)\} \cup p(j, j'), \\ p(j, u') &= \{(j, u)\} \cup p(u, u'). \end{aligned}$$

We have the same identities in (4.40) and (4.41) which, together with the criticality assumption, lead to the inequality (4.42). In order to make the correct assignment of the two vectors $\beta = (\beta_v)_{v \in U}$ and $\beta' = (\beta'_v)_{v \in U}$ we do as above for the case $j, u \in U$.

For the cases $(j, u) \in U \times \bar{U}$ and $(j, u) \in \bar{U} \times U$, we combine methods from the cases $(j, u) \in U \times U$ and $(j, u) \in \bar{U} \times \bar{U}$.

With this we finish the proof that we can learn the structure of every atom $\{\omega_r : r = 1, \ldots, |V|\}$, i.e., for every $r = 1, \ldots, |V|$ we can identify the unique node $i \in V$ such that $\omega_r = a_{i,U} = (b_{vi}/m_{i,U})_{v \in U}$. This means that we can also match every element β in the collection of vectors $\{\beta_r : r = 1, \ldots, |V|\}$ to the correct node $i \in V$ such that $\beta = (b_{vi})_{v \in U}$.

Step 2. In the previous step, we have shown that the distribution of X_U (together with the knowledge of the graph structure) determines the max-linear coefficient matrix $b_{U \times V} = (b_{vi})_{v \in U, i \in V}$. Here, we show that this matrix suffices to reconstruct the vector of edge coefficients $\theta = (c_e)_{e \in E}$.

If v is a child of i, then $p(i, v) = \{(i, v)\}$ and thus $b_{vi} = c_{iv} b_{ii}$. If both i and v belong to U, then, clearly, we can identify $c_{iv} = b_{vi}/b_{ii}$.

Let $i \in U$ with child $v \in \mathrm{ch}(i) \cap \bar{U}$. By Lemma 4.6.3.1 there exists a node $v' \in U$ such that there is a unique path from v to v'. Rewrite $v = v_1$ and $v' = v_n$ and consider the node set $\{v_1, v_2, \ldots, v_n\}$ on that unique path. Using the fact that if a node ℓ has a single parent k, then $b_{\ell\ell} = 1 - c_{k\ell}$ (see (4.10)), we find the following identities for the

max-linear coefficients $b_{v_n,j}$ for $j \in \{v_n, \ldots, v_2, i\}$:

$$\begin{aligned} b_{v_n,v_n} &= 1 - c_{v_{n-1},v_n}, \\ b_{v_n,v_{n-1}} &= c_{p(v_{n-1},v_n)} b_{v_{n-1},v_{n-1}} = c_{v_{n-1},v_n} \left(1 - c_{v_{n-2},v_{n-1}}\right), \\ &\vdots \\ b_{v_n,v_2} &= c_{p(v_2,v_n)} b_{v_2,v_2} = c_{v_2,v_3} \cdots c_{v_{n-1},v_n} \left(1 - c_{v_1,v_2}\right), \\ b_{v_n,i} &= c_{p(i,v_n)} b_{ii} = c_{i,v_1} c_{v_1,v_2} c_{v_2,v_3} \cdots c_{v_{n-1},v_n} b_{ii}. \end{aligned} \tag{4.44}$$

From the first equation we identify c_{v_{n-1},v_n}, from the second $c_{v_{n-2},v_{n-1}}$ and so on until we identify c_{v_1,v_2} from the penultimate equation. From the last equation we can identify c_{i,v_1} because b_{ii} is available from $b_{U\times V}$ in view of $i \in U$.

The next step of the proof is to extract the edge parameters between a node with latent variable and its children.

Let $i \in \bar{U}$. We will show that we can identify all edge weights c_{ij} for $j \in \mathrm{ch}(i)$. Because i belongs to $\bar{U}$, it should have at least two children, say v and $\bar{v}$. Take v to be a node whose only parent is i. Note that we can always find such a node by Lemma 4.6.3.1. Let us first assume $v, \bar{v} \in U$. Because v has only one parent, we have $b_{vv} = 1 - c_{iv}$ and thus $c_{iv} = 1 - b_{vv}$. We also know $b_{vi} = c_{i\bar{v}} b_{ii}$ and from here $b_{ii} = b_{vi}/c_{iv} = b_{vi}/(1 - b_{vv})$. From $b_{\bar{v}i} = c_{i\bar{v}} b_{ii}$ we deduce $c_{i\bar{v}} = b_{\bar{v}i}(1 - b_{vv})/b_{vi}$. Hence we have identified all the edge parameters related to children of i which are observable, provided i has two or more children in U.

Next assume that both $v, \bar{v} \in \bar{U}$. By Lemma 4.6.3.1, there exists a node $v' \in \mathrm{desc}(v) \cap U$ such that there is a unique path from v to v' and which has the properties in the cited statement. Let the sequence of nodes along which the path passes be denoted by $\{v_1 = v, v_2, \ldots, v_n = v'\}$. Using again that for a node ℓ with single parent k, $b_{\ell\ell} = 1 - c_{k\ell}$ (see (4.10)), we find the same identities as in (4.44) for the max-linear coefficients $b_{v_n,j}$ for $j \in \{v_n, \ldots, v_2\}$. From the first equation we obtain $c_{v_{n-1},v_n} = 1 - b_{v_n,v_n}$, from the second equation $c_{v_{n-2},v_{n-1}} = 1 - b_{v_n,v_{n-1}}/(1 - b_{v_n,v_n})$ and so on until we obtain c_{v_1,v_2} from the penultimate equation in (4.44). Because we assumed $\mathrm{pa}(v_1) = \{i\}$ we have $b_{v_1,v_1} = 1 - c_{i,v_1}$ and thus

$$b_{v_n,v_1} = c_{p(v_1,v_n)} b_{v_1,v_1} = c_{v_1,v_2} c_{v_2,v_3} \cdots c_{v_{n-1},v_n} \left(1 - c_{i,v_1}\right),$$

from where we identify c_{i,v_1}. Since $v_n \in U$, all coefficients $b_{v_n,j}$ for $j \in V$ are contained in the max-linear coefficient matrix $b_{U\times V}$. The procedure just described thus allows us to compute all edge coefficients $c_{i,v_1}, c_{v_1,v_2}, \ldots, c_{v_{n-1},v_n}$.

Because the path $p(v_1, v_n)$ satisfies the properties in Lemma 4.6.3.1, and in view of the same lemma, plus the fact that the only parent of $v_1 = v$ is i, it follows that the path $\{(i, v_1)\} \cup p(v_1, v_n)$ is the unique path between i and v_n. Hence we have also

$$b_{v_n,i} = c_{p(i,v_n)} b_{ii} = c_{i,v_1} c_{v_1,v_2} c_{v_2,v_3} \cdots c_{v_{n-1},v_n} b_{ii}. \tag{4.45}$$

As $v_n \in U$, the value of $b_{v_n,i}$ is known from $b_{U\times V}$. The edge coefficients on the right-hand side were expressed in terms of $b_{U\times V}$ in the previous paragraph. From there, we obtain the value of b_{ii}, which will be used next.

Now consider the node $\bar{v}$, renamed to $\bar{v}_1$, which was an arbitrary child of i. For $\bar{v} = \bar{v}_1$ too, we can find a node $\bar{v}_m \in U$ and a sequence of nodes $\{\bar{v}_2, \ldots, \bar{v}_m\}$ according to Lemma 4.6.3.1 satisfying the properties stated there. For all $r \in \{2, \ldots, m\}$, the node $\bar{v}_r$ has a unique parent, $\bar{v}_{r-1}$. We have the following equalities, using again by

$b_{\ell\ell} = 1 - c_{k\ell}$ for a node ℓ with unique parent k

$$\begin{aligned}
b_{\bar{v}_m,\bar{v}_m} &= 1 - c_{\bar{v}_{m-1},\bar{v}_m}, \\
b_{\bar{v}_m,\bar{v}_{m-1}} &= c_{p(\bar{v}_{m-1},\bar{v}_m)} b_{\bar{v}_{m-1},\bar{v}_{m-1}} = c_{\bar{v}_{m-1},\bar{v}_m} \left(1 - c_{\bar{v}_{m-2},\bar{v}_{m-1}}\right), \\
&\vdots \\
b_{\bar{v}_m,\bar{v}_2} &= c_{p(\bar{v}_2,\bar{v}_m)} b_{\bar{v}_2,\bar{v}_2} = c_{\bar{v}_2,\bar{v}_3} \cdots c_{\bar{v}_{m-1},\bar{v}_m} \left(1 - c_{\bar{v}_1,\bar{v}_2}\right), \\
b_{\bar{v}_m,i} &= c_{p(i,\bar{v}_m)} b_{ii} = c_{i,\bar{v}_1} c_{\bar{v}_1,\bar{v}_2} c_{\bar{v}_2,\bar{v}_3} \cdots c_{\bar{v}_{m-1},\bar{v}_m} b_{ii}.
\end{aligned}$$

In the last equality we used that the path $\{(i, \bar{v}_1)\} \cup p(\bar{v}_1, \bar{v}_m)$ is the unique shortest one between i and $\bar{v}_m$, because of Lemma 4.6.3.1 and because $p(\bar{v}_1, \bar{v}_m)$ satisfies the properties 1-2 of the same lemma. Since $\bar{v}_m \in U$, the values of the left-hand sides in the previous equations are contained in the given matrix $b_{U\times V}$. From the first equality we obtain $c_{\bar{v}_{m-1},\bar{v}_m}$, from the second one $c_{\bar{v}_{m-2},\bar{v}_{m-1}}$ and so on until we identify $c_{\bar{v}_1,\bar{v}_2}$ from the penultimate equality, i.e., all edge parameters linked to $p(\bar{v}_1, \bar{v}_m)$. In the last equation above we replace b_{ii} with the expression derived from (4.45) and we obtain the parameter $c_{i,\bar{v}_1}$.

If some of the children of i are in U and some others are in $\bar{U}$, we apply a combination of the techniques used in the two cases described above – when two children are in U or when two children are in $\bar{U}$. This concludes the proof of the sufficiency (if) part. □

Proof of necessity (only if) part in Proposition 4.4.2. Let $\bar{U} \subset V$ be such that at least one of the two conditions (I1) and (I2) is not satisfied and let $\theta = (c_e)_{e\in E} \in \mathring{\Theta}_*$. We will show that there exists another parameter $\theta' = (c'_e)_{e\in E} \in \mathring{\Theta}_*$ such that $\theta' \neq \theta$ but the distribution of $X_{V\setminus u}$ under θ' is the same as the one under θ.

We consider two cases: case (1) (I2) does not hold, i.e., there exists $u \in \bar{U}$ which is not the source of any tournament in $\mathcal{T}$; case (2) (I2) holds but not (I1), i.e., every $u \in \bar{U}$ is the source of some tournament in $\mathcal{T}$ but there exists $u \in \bar{U}$ with less than two children.

Case (1): there exists $u \in \bar{U}$ which is not the source of any tournament in $\mathcal{T}$. Then u belongs to only a single tournament, say $\tau = (V_\tau, E_\tau)$; indeed, a node that belongs to two different tournaments must be the source of at least one of them, because otherwise it would have parents from two different tournaments, yielding a forbidden v-structure.

Let X_{V_τ} be a max-linear model restricted to a single tournament, τ. The coefficients associated to X_{V_τ}, denote by b^τ_{vi} for $v, i \in V_\tau$, are determined by the weights of the edges $e \in E_\tau$. By Lemma 4.6.3.4, we can modify the edge weights c_e for $e \in E_\tau$ in such a way that the max-linear coefficients b^τ_{vi} for $v \in V_\tau \setminus u$ and $i \in V_\tau$ remain unaffected. Let $\tilde{c}_e$ for $e \in E_\tau$ denote such a modified vector of edge weights. Define $\theta' = (c'_e)_{e\in E} \in \mathring{\Theta}_*$ by

$$c'_e = \begin{cases} \tilde{c}_e & \text{if } e \in E_\tau, \\ c_e & \text{if } e \in E \setminus E_\tau. \end{cases}$$

Then $\theta' \in \mathring{\Theta}_*$ too: indeed, $c'_{ii} > 0$ for all $i \in V$ by assumption on θ and the fact the vector $(\tilde{c}_e : e \in E_\tau)$ satisfies $\tilde{c}_{ii} > 0$ for $i \in V_\tau$. By construction, the distribution of $X_{V_\tau\setminus u}$ is the same under θ' as under θ.

We show that the distribution of $X_{V\setminus u}$ under θ' is the same as the one under θ. We proceed by induction on the number of tournaments.

If $\mathcal{T}$ consists of a single tournament, then $\mathcal{T} = \tau$ and there is nothing more to show.

So suppose $\mathcal{T}$ consists of $m \geq 2$ tournaments. The skeleton graph of $\mathcal{T}$ is a block graph and thus a decomposable graph. By the running intersection property, we can order the tournaments $\tau_1, \ldots, \tau_m$ with node sets $V_1, \ldots, V_m$ in such a way that $\tau_1 = \tau$, the tournament containing u, and such that $V_m \cap (V_1 \cup \ldots \cup V_{m-1})$ is a a singleton, say $\{s\}$. Then $s \neq u$ since u belongs to only a single tournament.

Write $W = V_1 \cup \ldots \cup V_{m-1} \setminus u$. The joint distribution of $X_{V \setminus u}$ can be factorized into two parts: first, the distribution of X_W and second, the conditional distribution of $X_{V_m \setminus s}$ given X_W. It is sufficient to show that both parts remain the same when θ is replaced by θ'.

- By the induction hypothesis, the distribution of X_W is the same under θ' as under θ.
- By the global Markov property (Proposition 4.3.2), the conditional distribution of $X_{V_m \setminus s}$ given X_W is the same as the conditional distribution of $X_{V_m \setminus s}$ given X_s. But the latter is determined by the joint distribution of X_{V_m}, which, in turn, only depends on the weights of the edges e in τ_m. By construction, these edge weights are the same under θ as under θ'. It follows that the conditional distribution of $X_{V_m \setminus u}$ given X_W is the same under θ' as under θ.

We conclude that the distribution of $X_{V \setminus u}$ is the same under θ' as under θ. Since $\theta \neq \theta'$, the parameter is not identifiable.

Case (2): any $u \in \bar{U}$ is the source of some tournament in $\mathcal{T}$ but there exists $u \in \bar{U}$ with less than two children. Any $u \in \bar{U}$ must have at least one child (a node without children cannot be the source of a tournament). But then there exists $u \in \bar{U}$ with exactly one child: $\mathrm{ch}(u) = \{w\}$. The tournament of which u is the source can only consist of the nodes u and w and the edge (u, w). Now there are two subcases, depending on whether u has any parents or not.

Case (2).a: u has no parents. Then u is the source node of $\mathcal{T}$ with a single child w. Removing the node u yields the ttt $\mathcal{T}_{\setminus u} := (V \setminus u, E \setminus \{(u, w)\})$ with single source w. The random vector $X_{V \setminus u}$ follows the recursive max-linear model (4.2) with respect to $\mathcal{T}_{\setminus u}$. Its distribution is determined by the coefficients c_e for $e \in E \setminus \{(u, w)\}$. The value of c_{uw} can thus be chosen arbitrarily in $(0, 1)$ without affecting the distribution of $X_{V \setminus u}$.

Case (2).b: u has parents. Any ancestor of w different from u must be an ancestor of u too, since otherwise there would be a v-structure at w; therefore,

$$\mathrm{An}(w) = \mathrm{an}(u) \cup \{u, w\}.$$

Let $\lambda > 0$ and close enough to 1 (as specified below). Define $\theta' = (c'_e)_{e \in E}$ by modifying the weights of edges adjacent to u: specifically,

$$\begin{aligned} c'_{ju} &= \lambda c_{ju}, \qquad j \in \mathrm{pa}(u); \\ c'_{uw} &= \lambda^{-1} c_{uw}; \\ c'_e &= c_e, \qquad e \in E \setminus [\{(j, u) : j \in \mathrm{pa}(u)\} \cup \{(u, w)\}]. \end{aligned}$$

In words, θ' coincides with θ for edges e that do not involve u, and $\theta' = \theta$ if and only if $\lambda = 1$. Since the parameter space $\mathring{\Theta}_*$ is open, θ' belongs to $\mathring{\Theta}_*$ for λ sufficiently close to 1. We claim that the distribution of $X_{V \setminus u}$ is invariant under λ. Hence, for λ different from but sufficiently close to 1, we have found a parameter $\theta' \neq \theta$ producing the same distribution of $X_{V \setminus u}$ as θ.

Under θ', the random vector $X_{V \setminus u}$ follows the max-linear model

$$X_v = \bigvee_{i \in \mathrm{An}(v)} b'_{vi} Z_i, \qquad v \in V \setminus u,$$

where $(Z_i)_{i\in V}$ is a vector of independent unit-Fréchet random variables and where the coefficients b'_{vi} are given by equations (4.4), (5.9) and (4.6) with c_e replaced by c'_e.

First, suppose $v \in V \setminus u$ is not a descendant of u. Then for any $i \in \mathrm{An}(v)$, the coefficient b'_{vi} is a function of edge weights c'_e for edges $e \in E$ different from (u, w) and from (j, u) for $j \in \mathrm{pa}(u)$. It follows that $c'_e = c_e$ for such edges and thus $b'_{vi} = b_{vi}$ for $v \in V \setminus \mathrm{Desc}(u)$ and $i \in \mathrm{An}(v)$.

Second, suppose $v \in \mathrm{desc}(u)$. Then necessarily $v \in \mathrm{Desc}(w)$ too and for any $i \in \mathrm{An}(u)$, the path $p(i, v)$ passes by (or arrives in) w. Furthermore, any ancestor of v not in $\mathrm{An}(w)$ is a descendant of w:

$$\mathrm{An}(v) = \mathrm{an}(u) \cup \{u, w\} \cup [\mathrm{desc}(w) \cap \mathrm{An}(v)]. \tag{4.46}$$

It follows that

$$X_v = \bigvee_{i\in\mathrm{an}(u)} b'_{vi}Z_i \vee \left(b'_{vu}Z_u \vee b'_{vw}Z_w\right) \vee \bigvee_{i\in\mathrm{desc}(w)\cap\mathrm{An}(v)} b'_{vi}Z_i,$$

where the last term on the right-hand side is to be omitted if $v = w$. We treat each of the three terms on the right-hand side separately.

- For $i \in \mathrm{an}(u)$, we have

  $$b'_{vi} = c'_{ii}c'_{p(i,v)} = c'_{ii}c'_{p(i,u)}c'_{uw}c'_{p(w,v)}$$

  where $c'_{p(w,v)} = 1$ if $v = w$. The coefficients c'_{ii} and $c'_{p(w,v)}$ only involve weights c'_e for edges $e \in E$ different from (u, w) and (j, u) for $j \in \mathrm{pa}(u)$; it follows that $c'_{ii} = c_{ii}$ and $c'_{p(w,v)} = c_{p(w,v)}$. Further, given $i \in \mathrm{an}(u)$ there exists $j \in \mathrm{pa}(u)$ such that $p(i, u)$ passes by j right before reaching u (with $i = j$ if $i \in \mathrm{pa}(u)$), and then

  $$c'_{p(i,u)}c'_{uw} = c'_{p(i,j)}c'_{ju}c'_{uw} = c'_{p(i,j)}(\lambda c_{ju})(\lambda^{-1}c_{uw}).$$

  Since $p(i, j)$ does not involve edges meeting u, we find that $c'_{p(i,j)} = c_{p(i,j)}$, so that the above expression does not depend on λ. We conclude that $b'_{vi} = b_{vi}$ for $i \in \mathrm{an}(u)$.

- If $v \in \mathrm{desc}(w)$ and $i \in \mathrm{desc}(w) \cap \mathrm{An}(v)$, the coefficient b'_{vi} is

  $$b'_{vi} = c'_{ii}c'_{p(i,v)}.$$

  The path $p(i, v)$ does not involve edges touching u, so $c_{p(i,v)'} = c_{p(i,v)}$. By Lemma 4.6.3.3, the coefficient $c'_{ii} = b'_{ii}$ is a function of the edge weights c'_e for edges e in the tournament shared by i and its parents. Since $i \in \mathrm{desc}(w)$ and since w is the only child of u, none of these edges touches u, and thus $c'_e = c_e$ for all such edges. It follows that $c'_{ii} = c_{ii}$ too. We conclude that $b'_{vi} = b_{vi}$ for $v \in \mathrm{desc}(w)$ and $i \in \mathrm{desc}(w) \cap \mathrm{An}(v)$.

- The random variable $b'_{vu}Z_u \vee b'_{vw}Z_w$ is independent of all other variables Z_i for $i \in V \setminus \{u, w\}$ and its distribution is equal to $(b'_{vu} + b'_{vw})Z$ for Z a unit-Fréchet variable. We will show that $b'_{vu} + b'_{vw}$ does not depend on λ. Since $1 = \sum_{i\in\mathrm{An}(v)} b'_{vi}$, the partition (4.46) yields

  $$b'_{vu} + b'_{vw} = 1 - \sum_{i\in\mathrm{an}(u)} b'_{vi} - \sum_{i\in\mathrm{desc}(w)\cap\mathrm{An}(v)} b'_{vi}.$$

  (The last sum on the right-hand side is zero if v is not a descendant of w.) In the two previous bullet points, we have already shown that the coefficients b'_{vi} for i in $\mathrm{an}(u)$ or $\mathrm{desc}(w) \cap \mathrm{An}(v)$ do not depend on λ. By the stated identity, the sum $b'_{vu} + b'_{vw}$ then does not depend on λ either.

We have thus shown that if $\bar{U}$ does not satisfy (I1)–(I2), then we can find $u \in \bar{U}$ such that the distribution of $X_{V \setminus u}$ is the same under θ' as under θ. As $\theta' \neq \theta$ by construction, this means that the parameter θ is not identifiable from the distribution of $X_{V \setminus u}$. But as $U = V \setminus \bar{U} \subseteq V \setminus \{u\}$, the parameter θ is not identifiable from the distribution of X_U either. This confirms the necessity of (I1)–(I2) for the identifiability of θ from the distribution of X_U. □

Sum-linear graphical models with heavy tailed factors on trees of transitive tournaments

5

5.1 Introduction

Here we discuss some properties of a linear recursive structural equation model (SEM), which we call *sum-linear* model in analogy to the max-linear which is much more popular in the extreme value literature. Recall the notation for a directed acyclic graph (DAG) $\mathcal{G}$ with a node set V and edge set E, and a random vector with elements associated to every node of the graph, $X = (X_v, v \in V)$. A SEM is defined as

$$X_v = f(X_{\mathrm{pa}(v)}, \varepsilon_v), \qquad v \in V, \tag{5.1}$$

where $(\varepsilon_v, v \in V)$ is a vector of independent noise variables and f is some function. Such a model on a directed acyclic graph is interpreted as defining causal relations between parent variables and their children (Spirtes et al., 1993; Kiiveri et al., 1984; Spirtes, 1994).

The max-linear Bayesian models are introduced in Gissibl and Klüppelberg (2018) and since then, they are the focus of a large part of the literature on graphical models and extremes. They are a special type of linear recursive SEM as their conditional independence relations differ significantly with respect to a classical SEM (Klüppelberg and Lauritzen, 2019).

The sum-linear Bayesian models appear for the first time in the extreme value context in Gnecco et al. (2021). Subsequent literature on topics of causality between extremes involving these models is Pasche et al. (2021), Zhao et al. (2021), and Budhathoki et al. (2022). For instance, Budhathoki et al. (2022) talk about finding *root causes* for extreme values on a known DAG by quantifying the contribution of each noise variable. Zhao et al. (2021) instead, interpret causal learning as a problem of learning the structure of the DAG from a dataset.

From an extreme value perspective, the two models have been studied in Einmahl et al. (2012) where it is shown that they are in the domain of attraction of the same class of extreme value distributions with discrete angular measure. Here we study the same properties that have been studied in Chapter 4 on max-linear models. We show thus, that what regards sum- or max-linear models on a tree of transitive tournaments, they both share the same probabilistic characteristics. This is not surprising in view of the results of Einmahl et al. (2012) mentioned above. However there are some obvious differences between the two models, such as the parameterization and the conditional independence relations.

First we present the sum-linear model, its parameterization and some algebraic properties that render it so similar to the max-linear model with critical shortest paths.

Next we study the markovianity of the model and the factorization of the limiting vector of the conditional tails. There is a section on parameter identifiability where we show that the necessary and sufficient condition is identical to the one for the max-linear model. We conclude outlining the main differences and similarities of the two models.

5.2 Model definition

In a similar way to Gnecco et al. (2021) we assume that $X = (X_v, v \in v)$ is a sum-linear SEM with respect to a ttt, the latter as in Definition 4.2.1 in Chapter 4. Recall that for a directed edge $(v, i) \in E$ we have an edge weight c_{vi}, and for every node $v \in V$ we have a node coefficient d_{vv}. An element of X is given by

$$X_v = \sum_{i \in \mathrm{pa}(v)} c_{iv} X_i + d_{vv} Z_v, \qquad v \in V, \tag{5.2}$$

where $(Z_i, i \in V)$ is a vector of independent unit-Fréchet random variables. We define $\pi(i, j)$ as the set of all directed paths from node i to node j. An element of $\pi(i, j)$ is a directed path, say p, from i to j, which represents a set of edges. If the path p passes through nodes $\{i = v_1, v_2, \ldots, v_n = j\}$ then $p = \{(v_1, v_2), (v_2, v_3), \ldots, (v_{n-1}, v_n)\}$. We keep the notation $p(i, j)$ for the unique shortest directed path between nodes i, j. As in Chapter 4, we have the product of edge weights along some directed path $p \in \pi(i, j)$

$$c_p = \prod_{(a,b) \in p} c_{ab}.$$

Let $c_{\pi(i,j)}$ be the sum of all such path products, i.e.,

$$c_{\pi(u,v)} := \sum_{p \in \pi(u,v)} c_p, \tag{5.3}$$

with the convention $c_{\pi(i,i)} = 1$.

To ensure each path has a positive share in $c_{\pi(i,j)}$ we assume positive edge weights, i.e. $\theta := (c_e, e \in E) > 0$. Then for given pair of nodes i, j, every path product is positive: $c_p > 0$ for all $p \in \pi(i, j)$ thus $c_{\pi(i,j)} > 0$ too. This assumption avoids cases such as $c_p = 0$ or even $c_{\pi(i,j)} = 0$. The case $c_p = 0$ would be interpreted that influence from i cannot go to j through path p, and if $c_{\pi(i,j)} = 0$, then no influence can go from i to j at all, although there are directed paths between them. Let C be the weighted adjacency matrix of the ttt $\mathcal{T}$, defined as $\{C\}_{ji} = c_{ij}$ if $(i, j) \in E$ and zero otherwise. The matrix C is square $|V| \times |V|$ and the diagonal elements are zero. Let $D = \{D\}_{ij} = d_{ij}$ be a diagonal matrix with the node coefficients $\{d_{vv}\}$ from (5.2) on the diagonal. Hence the sum-linear model $X = (X_v, v \in V)$ with heavy tailed errors $Z = (Z_v, v \in V)$ is given as

$$X = CX + DZ. \tag{5.4}$$

According to Remark 2.3 in Gissibl and Klüppelberg (2018) the node labeling on $\mathcal{T}$ can be such that if $i \in \mathrm{pa}(j)$, then $i < j$ and then the matrix C, for which we have $\{C\}_{ji} = c_{ij}$, is lower triangular with zeros on the main diagonal.

Lemma 5.2.1. *Let X be as in* (5.4) *with C a lower diagonal matrix of edge weights. Then*

$$X = BZ, \tag{5.5}$$

or also $X_v = \sum_{i \in V} b_{vi} Z_i$ for every $v \in V$ with

$$b_{vi} = \begin{cases} 0 & \text{if } i \notin \mathrm{An}(v), \\ d_{vv} & \text{if } i = v, \\ d_{ii} c_{\pi(i,v)} & \text{if } i \in \mathrm{an}(v), \end{cases} \tag{5.6}$$

for every $v \in V$.

Proof. From (5.4) we have $(I_{|V|} - C)X = DZ$ and hence $X = (I_{|V|} - C)^{-1}DZ$ as $I_{|V|}$ is an identity matrix of size $|V|$ and $I_{|V|} - C$ as a lower diagonal matrix is invertible. From the equality

$$\left(I_{|V|} + C + C^2 + \cdots + C^k\right)\left(I_{|V|} - C\right) = I_{|V|} - C^{k+1}$$

and the fact that, for $|V|$-square lower triangular matrix with zero diagonal, powers of $k \geq |V|, k = 0, 1, 2, \ldots$ are zero matrices, we obtain

$$\left(I_{|V|} - C\right)^{-1} = \left(I_{|V|} + C + C^2 + \cdots + C^{|V|-1}\right).$$

Then we have

$$B = \left(I_{|V|} - C\right)^{-1}D = \left(I_{|V|} + C + C^2 + \cdots + C^{|V|-1}\right)D \tag{5.7}$$

Since C is a weighted adjacency matrix, an element $\{C^k\}_{ij}$ represents the sum of the path products along all paths from j to i of length k, i.e.,

$$\{C^k\}_{ij} = \sum_{p \in \pi(j,i), |p|=k} c_p. \tag{5.8}$$

Because C is lower diagonal with zero diagonal and D is diagonal, the matrix B is also lower diagonal with $\{d_{vv}\}$ on the main diagonal. This shows the second statement in (5.6). For $v > i$ we have

$$b_{vi} = (\{C\}_{vi} + \cdots + \{C^{|V|-1}\}_{vi})d_{ii}$$

The interpretation of $\{C^k\}_{vi}$ in (5.8) means that the sum in the parenthesis of the above equation is $c_{\pi(i,v)}$ and this shows the third statement in (5.6). From the equation above it follows that if $\mathrm{an}(v)$ is empty then the sum is zero as all terms $\{C^k\}_{vi}$ are zero too. This shows the first statement of (5.6). □

We will refer sometimes to B as the coefficient matrix of the sum-linear model. We require $d_{ii} = b_{ii}$ for all $i \in V$ to be also positive meaning that the noise variable on node i has positive influence on X_i too. The result in (5.5) and (5.6) is the one given in Gnecco et al. (2021, Equation 5).

Next we impose the following standardization on the row sums of the matrix B:

$$\sum_{i \in V} b_{vi} = \sum_{i \in \mathrm{An}(v)} b_{vi} = 1, \qquad \forall v \in V. \tag{5.9}$$

If the ancestor set is empty, the sum is zero and $b_{vv} = 1$. The constraint in (5.9) means that X_v is in the domain of attraction of the unit-Fréchet distribution. This is because the ratio of the tail probabilities of the sum and the maximum exceeding x tends to 1 (Einmahl, Krajina, and Segers, 2012, Page 15):

$$\lim_{x \to \infty} \frac{\mathbb{P}\left(\sum_{i \in V} b_{vi} Z_i > x\right)}{\mathbb{P}\left(\max_{i \in V} b_{vi} Z_i > x\right)} = 1.$$

Using matrix algebra we can state the condition in (5.9) as

$$B1_{|V|} = 1_{|V|}, \tag{5.10}$$

with $1_{|V|}$ being a vector of ones of length $|V|$.

The constraints of positive edge weights, $c_e > 0, \forall e \in E$; positive node weights, $c_{vv} > 0, \forall v \in V$ and (5.9) lead to the following parameter space for θ

$$\mathring{\Theta}_\Sigma := \left\{\theta = (c_e, e \in E) \in (0,1)^E : \forall v \in V, c_{vv} > 0\right\}. \tag{5.11}$$

This is the same definition as the one of $\mathring{\Theta}$ in (4.7) but the spaces are different as the coefficients $c_{vv}, v \in V$, have different definitions for max- and sum-linear models. We will write B_θ for the matrix in (5.5) to emphasize the dependence on the edge parameters.

Definition 5.2.1 (Sum-linear recursive structural equation model on a ttt)**.** *Let $X = (X_v, v \in V)$ be a sum-linear recursive SEM with respect to a ttt, $\mathcal{T}$. The random vector X has a representation as in* (5.5)*, with unit Fréchet independent factors $\{Z_v\}$; matrix $B := B_\theta$ given by* (5.6) *and satisfying* (5.10)*; and $\theta = (c_e, e \in E) \in \mathring{\Theta}_\Sigma$ in* (5.11)*.*

The distribution of X is in the max-domain of attraction of the same class of extreme value distributions as the max-linear model (Einmahl, Krajina, and Segers, 2012). Hence the extreme value distribution, the stdf, and the angular measure are the same up to the coefficients $\{b_{ij}\}_{i,j\in V}$. A brief overview of the relations between these objects is given in Section 4.2.3 in Chapter 4. For completeness we repeat the angular measure of X, given by

$$H_\theta = \sum_{i \in V} m_i \delta_{a_i},$$

where $m_i = \sum_{v \in V} b_{vi}$ is the mass of an atom

$$a_i = \frac{1}{m_i}(b_{vi}, v \in V). \tag{5.12}$$

5.3 Algebraic properties of the model coefficients

There are two key properties of the sum-linear model that make it very close to the max-linear model at least in the studied framework — Bayesian networks on trees of transitive tournaments. The first one relates to factorization of $c_{\pi(i,j)}$ along the unique shortest path and the second one relates to the limiting tail distribution. Here we state the first property.

Lemma 5.3.1. *Let X be sum-linear model as in Definition 5.2.1. Let $\{u = v_1, \dots, v_n = v\}$ be the set of nodes on the unique shortest directed path between a pair of nodes $u, v \in V$. Then we have*

$$c_{\pi(u,v)} = \prod_{i=1}^{n-1} c_{\pi(v_i, v_{i+1})}. \tag{5.13}$$

Proof. We proceed by induction on the number of nodes in the unique shortest path $p(u, v)$. If there are two nodes in the path, these must be adjacent, hence equation (5.13) is trivially satisfied. Let there be n nodes in the unique shortest path, i.e., $\{v_1 = u, \dots, v_n = v\}$, and we assume that (5.13) holds. We add one more node to obtain the unique shortest path on $n + 1$ nodes $\{u = v_1, \dots, v_n, v_{n+1} = v\}$. We aim to show

$$c_{\pi(u,v)} = \prod_{i=1}^{n} c_{\pi(v_i, v_{i+1})}. \tag{5.14}$$

Consider the two sets $\pi(v_1, v_n)$ and $\pi(v_n, v_{n+1})$, and union of them, denoted by $\pi(v_1, v_n) \otimes \pi(v_n, v_{n+1})$. An element of this new set is $p_i \cup \bar{p}_j$ for $p_i \in \pi(u, v_n)$ and

$\bar{p}_j \in \pi(v_n, v_{n+1})$:

$$\pi(v_1, v_n) \otimes \pi(v_n, v_{n+1}) = \begin{Bmatrix} p_1 \\ \vdots \\ p_m \end{Bmatrix} \otimes \begin{Bmatrix} \bar{p}_1 \\ \vdots \\ \bar{p}_k \end{Bmatrix} = \begin{Bmatrix} p_1 \cup \bar{p}_1 & \cdots & p_1 \cup \bar{p}_k \\ \vdots & \dots & \vdots \\ p_m \cup \bar{p}_1 & \cdots & p_m \cup \bar{p}_k \end{Bmatrix} \tag{5.15}$$

The aim is to show

$$\pi(u, v_{n+1}) = \pi(u, v_n) \otimes \pi(v_n, v_{n+1}). \tag{5.16}$$

Since every path from the set on the right hand-side of (5.16) is a path from u to $v = v_{n+1}$ we have $\pi(u, v_n) \otimes \pi(v_n, v_{n+1}) \subset \pi(u, v_{n+1})$. By definition, all paths in $\pi(u, v_n) \otimes \pi(v_n, v_{n+1})$ pass through v_n. But by Lemma 5.8.1, every path $p \in \pi(u, v_{n+1})$ contains the nodes on the shortest path, in particular it contains v_n. Thus we have $\pi(u, v_{n+1}) \subset \pi(u, v_n) \otimes \pi(v_n, v_{n+1})$. So we showed (5.16). Then we have

$$\begin{aligned} c_{\pi(u,v_{n+1})} &= \sum_{p \in \pi(u,v_{n+1})} c_p = \sum_{p \in \pi(u,v_n) \otimes \pi(v_n,v_{n+1})} c_p = \sum_{p_i \in \pi(u,v_n)} \sum_{\bar{p}_j \in \pi(v_n,v_{n+1})} c_{p_i} c_{\bar{p}_j} \\ &= \left(\sum_{p_i \in \pi(u,v_n)} c_{p_i} \right) \left(\sum_{\bar{p}_j \in \pi(v_n,v_{n+1})} c_{\bar{p}_j} \right) = c_{\pi(u,v_n)} c_{\pi(v_n,v_{n+1})}. \end{aligned}$$

The third equality can be understood in the light of (5.15). After that, the equalities follow from rearranging and the definition of $c_{\pi(i,j)}$. Together with the induction hypothesis we obtain (5.14). □

Remark. This creates the link with the coefficients $\{b_{ij} > 0\}$ for the max-linear model in the previous chapter. In the max-linear context, $c_{p(u,v)}$ is the product of the edge weights along the unique shortest path $p(u, v)$. Because of the criticality assumption we have $b_{ij} = c_{p(j,i)} b_{jj}$. For the sum-linear model, we have $b_{ij} = c_{\pi(j,i)} b_{jj}$, but the term $c_{\pi(j,i)}$ is factorizable along the unique shortest path $p(j, i)$.

We also have due to the constraint in (5.10) the following result according to which the coefficient $b_{vv} = d_{vv}$ in (5.6), case $i = v$, involves only the weights on edges incident to parents of v.

Lemma 5.3.2. *Let X be sum-linear model as in Definition 5.2.1. We have*

$$b_{vv} = 1 - \sum_{u \in \mathrm{pa}(v)} c_{uv}, \qquad v \in V. \tag{5.17}$$

Proof. From the constraint in (5.10) and the form of B in (5.7) we get

$$(I_{|V|} - C)^{-1} D 1_{|V|} = 1_{|V|}.$$

Multiplying both sides of the equation by $(I_{|V|} - C)$ we get $D1_{|V|} = (I_{|V|} - C)1_{|V|}$. The left hand-side represents a vector of the coefficients $\{b_{vv}\} = \{d_{vv}\}$ and the right hand-side is a vector where every element is given by (5.17). □

5.4 Markovianity with respect to the skeleton of $\mathcal{T}$

In Proposition 4.3.2 we have given a necessary and sufficient condition for the max-linear model to factorize with respect to the skeleton of the ttt, $\mathcal{T}$. The necessity has been proven using the theory on *completeness* of entirely new concept of separation based on so called $*$-connecting paths. Completeness of graph separation criteria has not been established so far in general. All we know is that if two sets are not separated

by a third one in the graph, be it a non-directed or a directed one, for a set of probability measures of measure zero the variables on the two sets are independent conditionally on the variables in the separator set (Koller and Friedman, 2009, Theorem 3.5 and 4.3).

We give only a sufficient condition under which the sum-linear model satisfies the global Markov property with respect to the undirected graph of $\mathcal{T}$. We believe this one of the underlying causes for the factorization of the tail limiting distribution into independent variables along the unique shortest undirected paths in $\mathcal{T}$.

To show markovianity for a max-linear model it was necessary to use the theory of conditional independence developed recently in Améndola et al. (2022). In this respect the sum-linear graphical model is considerably closer to the classical theory of conditional independence as in Lauritzen (1996) or Koller and Friedman (2009, Chapters 3-4). This is due to the fact that the sum-linear model has positive and continuous probability density.

Lemma 5.4.1 (Continuous density function)**.** *Let X be a sum-linear model as in Definition 5.2.1. The probability density function is continuous over $(x_v, v \in V) \in (0,\infty)^V$.*

Proof. From (5.5) we have $f_X(x) = f_Z(B^{-1}x)\det(B^{-1})$ which is a continuous function as $Z = (Z_v, v \in V)$ is jointly independent and hence f_Z is the product of unit Fréchet probability density functions. Also B from (5.7) is invertible as it is given by the product of two invertible matrices. □

Some notions from traditional conditional independence follow next.

Let P be a probability distribution with positive and continuous density with respect to a measure on the product space $\times_{v\in V}\mathcal{X}_v$ where $\mathcal{X}_v \subseteq \mathbb{R}$ is the range of values of X_v. The distribution P satisfies the local directed Markov property (Koller and Friedman, 2009, Definition 3.1) with respect to $\mathcal{T}$ if for every $v \in V$ we have

$$X_v \perp\!\!\!\perp X_{V\setminus \mathrm{Desc}(v)} \mid X_{\mathrm{pa}(v)}. \tag{5.18}$$

According to Koller and Friedman (2009, Theorem 3.1) if P as the probability distribution of X contains the conditional independence relations arising from (5.18), then P factorizes according to $\mathcal{T}$ by satisfying the equation

$$f_V(x_v, v \in V) = \prod_{v\in V} f_{v|\mathrm{pa}(v)}(x_v \mid x_{\mathrm{pa}(v)}), \tag{5.19}$$

where f_V is the density function associated to the probability measure P and $f_{v|\mathrm{pa}(v)}$ are conditional marginal density functions.

Theorem 1.4.1 in Pearl (2000) states that every SEM as in (5.1) with respect to a DAG and with jointly independent errors $\{\epsilon_v\}$ satisfies the conditional independence relations in (5.18). Hence our sum-linear model too.

We recall the global Markov property for non-directed graphs. If for any A, B, S which are non-empty and disjoint subsets of the node set V, we have that S separates A and B in the graph, then X_A is conditionally independent from X_B given X_S.

The so called *moralized* graph of $\mathcal{T}$, denoted by $\mathcal{T}^m$, is obtained by connecting every pair of parents with common child and removing directions (Lauritzen, 1996, Section 2.1.1). The skeleton (non-directed version) of $\mathcal{T}$ is denoted by T. A ttt $\mathcal{T}$ with unique source has the property $\mathcal{T}^m = T$ (see Lemma 5.8.1). However for a ttt with at least two sources, the two graphs are different. Then by Lemma 4.6.1 there is at least one v-structure. Hence the moral graph has edges between the parents of every v-structure.

Finally, Lemma 3.21 in Lauritzen (1996) says that if a distribution factorizes according to a DAG, then it satisfies the global Markov property with respect to the moral graph.

The following proposition follows from the three theorems cited above.

Proposition 5.4.1 (Markovianity)**.** *Let X be a sum-linear model as in Definition 5.2.1. The random vector X satisfies the global Markov property with respect to the skeleton, T, if $\mathcal{T}$ has a unique source.*

Proof. First suppose $\mathcal{T}$ has a unique source. Then by Lemma 5.8.1 we have $\mathcal{T}^m = T$. The claim follows from Pearl (2000, Theorem 1.4.1), Koller and Friedman (2009, Theorem 3.1) and Lauritzen (1996, Theorem 3.21). □

For a sum-linear model with innovations $\{Z_v\}$ that have finite variances necessity can be shown too. However, in our case these are unit Fréchet variables. It would be not surprising if the necessity can be shown using the method employed in Améndola et al. (2022, Section 6).

5.5 Characterization of the tail limit

We are interested in the limit distribution of the tails of a sum-linear model as a high threshold is exceeded at a particular node. In this section we will present a necessary and sufficient condition for this limit to be composed of independent multiplicative increments along the unique shortest trail (undirected path) between two variables. This behavior is observed for Markov trees (Segers, 2020b), Markov block graphs (Asenova and Segers, 2021) and Markov max-linear Bayesian ttt as in Chapter 4.

We have presented the factorizable structure of $c_{\pi(u,v)}$ in Lemma 5.3.1 as one of the two key properties that render the sum-linear model very similar to the max-linear one. The other one refers to the limit of the vector, scaled by a particular variable given that this variable exceeds a high threshold.

Proposition 5.5.1. *Let $X = (X_v, v \in V)$ be sum-linear recursive SEM on some index set V . For every $u \in V$, as $t \to \infty$*

$$\begin{aligned} \mathcal{L}(X_v/X_u, v \in V \mid X_u > t) \xrightarrow{d} \mathcal{L}(A^{(u)}) &= \mathcal{L}(A_{uv}, v \in V) \\ &= \sum_{j \in V} b_{uj} \delta_{(b_{vj}/b_{uj}, v \in V)}. \end{aligned} \tag{5.20}$$

Proof. Let F_X be the joint multivariate distribution function of the sum-linear model $X = (X_v, v \in V)$. By Einmahl et al. (2012, Lemma 6.1) we know that F_X is in the max-domain of attraction of a distribution G, written also $F_X \in D(G)$. The condition $F_X \in D(G)$ implies by Rootzén and Tajvidi (2006, Theorem 2.2(i)) and Beirlant et al. (2004, Section 8.3) that there is a random vector, Y, with distribution given by

$$\frac{\ln G(\min(1, x_v), v \in V) - \ln G(x)}{\ln G(1, \ldots, 1)}, \qquad x \in (0, \infty]^V,$$

such that the following convergence in distribution holds

$$(X_v/t, v \in V) \mid \max_{v \in V} X_v > t \xrightarrow{d} Y. \tag{5.21}$$

The convergence in (5.21) implies that for every $u \in V$ there is a random vector $Y^{(u)} := (Y_v, v \in V) \mid Y_u > 1$ such that

$$(X_v/t, v \in V) \mid X_u > t \xrightarrow{d} (Y_v, v \in V) \mid Y_u > 1 =: Y^{(u)}.$$

By Segers (2020b, Theorem 2.a,c)) the convergence above is equivalent to

$$(X_v/X_u, v \in V) \mid X_u > t \xrightarrow{d} (\xi_{uv}, v \in V), \tag{5.22}$$

for some $\xi_u = (\xi_{uv}, v \in V)$ such that $Y^{(u)} = \zeta\xi_u$ with ζ being unit Pareto and independent of ξ_u. If X' is a max-linear with parameter matrix $\{b_{ij}\}_{i,j\in V}$, from Segers (2020b, Example 3) we have

$$(X'_v/X'_u, v \in V) \mid X'_u > t \to (A_{uv}, v \in V) \sim \sum_{j\in V} b_{uj}\delta_{(b_{vj}/b_{uj}, v\in V)}. \tag{5.23}$$

Because X' as max-linear and X as sum-linear models belong to the same domain of attraction, the one of the distribution G,we must have that the two limits in (5.22) and (5.23) are the same, hence $\xi_u = (A_{uv}, v \in V)$. □

Remark. The convergence of the distribution in (5.20) holds for any sum-linear model with representation as in (5.5), not necessarily a graphical model.

A unique shortest trail (undirected path) exists for every pair of nodes according to property (P3) of a ttt discussed in Section 4.2. Recall the definitions from Section 4.3 for $t(u,v)$, $t_u(u,v)$, the set of edges E_u from (4.15), and the node $w_{u,\tau}$ for fixed u and every $\tau \in \mathbb{T}$. The following proposition says that the limit of $(A_{uv}, v \in V)$ is based on the joint distribution of a vector of random edge weights $(M_e, e \in E_u)$. The following proposition is the counterpart of Proposition 4.3.1 from the previous chapter for sum-linear models.

Proposition 5.5.2 (Factorization of the tail limit)**.** *Let $(X_v, v \in V)$ be a sum-linear model as in Definition 5.2.1. Fix $u \in V$. Let E_u be as in* (4.15) *and let $(M_e, e \in E_u)$ be a random vector composed of independent subvectors $M^{(u,\tau)} = \big(M_{w_{u,\tau},j} : j \in V_\tau, (w_{u,\tau}, j) \in E_u\big)$ for every transitive tournament $\tau \in \mathbb{T}$. Let $M^{(u,\tau)}, \tau \in \mathbb{T}$, have the marginal limiting distribution as in Lemma 5.8.2 and be mutually independent.*

The following statements are equivalent:

(i) $\mathcal{T}$ has a unique source.

(ii) For every $u \in V$, we have, as $t \to \infty$, the weak convergence

$$\mathcal{L}(X_v/X_u, v \in V \mid X_u > t) \xrightarrow{d} \mathcal{L}(A^{(u)}) = \mathcal{L}(A_{uv}, v \in V) \tag{5.24}$$

with

$$A_{uv} = \prod_{e\in t_u(u,v)} M_e, \qquad v \in V. \tag{5.25}$$

(iii) There exists $u \in V$ such that the limit in (5.24) *and* (5.25) *holds.*

5.6 Unobserved variables and the identifiability criterion

Consider again the case when some of $X_v, v \in V$ are unobserved. The set of nodes with observed variables is U and the set with non-observed ones $\bar{U}$. The angular measure of the subvector $(X_v, v \in U)$ is $H_{\theta,U}$ and is given by

$$H_{\theta,U} = \sum_{r=1}^{s} \mu_r \delta_{(\omega_{vr}, v\in U)} \tag{5.26}$$

for a collection of atoms $\{\omega_r\}_{r=1}^s = \{(\omega_{vr}, v \in U)\}_{r=1}^s$ and their masses $\{\mu_r\}_{r=1}^s$. An atom ω_r is equal to $(\beta_{vr}, v \in U)/\mu_r$ where $(\beta_{vr}, v \in U)$ is a vector of coefficients whose value is related to the one of the vectors $\{(b_{vi}, v \in U)\}$.

When there are unobserved variables, it is possible that two different edge parameter vectors $\theta_1, \theta_2 \in \mathring{\Theta}_\Sigma$ induce the same distribution of $(X_v, v \in U)$ and accordingly the same angular measure $H_{\theta,U}$. We call this the identifiability problem due to unobserved variables and it turns out that a criterion for identifiability can be provided for a ttt with unique source.

We first need to assure ourselves that there is no identifiability problem in the distribution of the complete vector.

Lemma 5.6.1 (Parameter identifiability). *Let X be a sum-linear model as in Definition 5.2.1 with respect to a ttt, $\mathcal{T}$. Then for the angular measure H_θ we have*

(i) $m_i > 0$ *for all* $i \in V$*;*

(ii) for any atom $a_i = (a_{vi})_{v\in V}$*, we have* $a_{vi} > 0$ *if and only if* $v \in \mathrm{Desc}(i)$*. Specifically, all* $|V|$ *vectors* a_i *are different;*

(iii) for $\theta_1 \neq \theta_2 \in \mathring{\Theta}_\Sigma$ *we have* $H_{\theta_1} \neq H_{\theta_2}$*.*

Proof. When there are no latent variables we have for every node $j \in V$, $m_j = \sum_{i\in\mathrm{Desc}(j)} b_{ij} > 0$ and $\mathrm{Desc}(j)$ contains at least the node j for which we have $b_{jj} > 0$ by the assumptions on the parameter space $\mathring{\Theta}_\Sigma$. This shows that we have $|V|$ well defined atoms and their respective positive masses.

Also, we cannot have $a_i = a_j$. The argument is the same as in Lemma 4.4.1 which is about identifiability of the parameters in the max-linear model. This is because $b_{vi} > 0$ for all $v \in \mathrm{Desc}(i)$, and respectively $b_{vj} > 0$ for all $v \in \mathrm{Desc}(j)$. Since $\mathrm{Desc}(i) \neq \mathrm{Desc}(j)$ for every $i \neq j \in V$, the vectors a_i, a_j have different zero patterns so all $|V|$ atoms are distinct from each other and the coefficient vectors $\{(b_{vi}, v \in V)\}_{i=1,\dots,|V|}$ can be uniquely associated to a node of the graph and hence to an expressions in terms of $(c_e, e \in E)$ according to (5.7). This shows (ii).

Given that we can reconstruct the matrix B from the set of atoms $\{a_i\}$ and masses $\{m_i\}$ we can solve from (5.7) for C. This will prove the third statement. Since we have

$$B = (I_{|V|} - C)^{-1}D = (I_{|V|} - C)^{-1}\,\mathbf{diag}(B),$$

with $\mathbf{diag}(B)$ being a diagonal matrix with main diagonal the elements $\{b_{vv}\}_{v\in V}$. Then

$$C = (\mathbf{diag}(B) - B)B^{-1}.$$

Note that B is invertible because from (5.7) it is given by the product of two invertible matrices. □

When there are latent variables the first two points in the lemma above may not hold for the angular measure $H_{\theta,U}$. This is a consequence of the fact that it is possible to have $\mathrm{Desc}(i) \cap U = \emptyset$ for some $i \in V$ or $\mathrm{Desc}(i) \cap U = \mathrm{Desc}(j) \cap U$ for $i \neq j \in V$.

Recall the conditions of identifiability for the max-linear model:

(I1) any $u \in \bar{U}$ has at least two children,

(I2) any $u \in \bar{U}$ is the source of some tournament in $\mathcal{T}$.

Exactly the same conditions hold for the sum-linear model too.

Proposition 5.6.1 (Parameters' identifiability in case of latent variables)**.** *Let X be a* sum-linear *model as in Definition 5.2.1 and let $\mathcal{T} = (V, E)$ be a ttt with single source. Let U be the set of nodes whose variables are unobserved. The parameter θ is uniquely identifiable from the distribution of $(X_u : u \in U)$ if and only if conditions (I1) and (I2) are satisfied.*

5.7 Conclusion

We have dedicated the present chapter on the sum-linear model, and have shown that the properties of interest are exactly identical to those of the max-linear model under some constraints on the parameter spaces.

In particular, the condition for markovianity with respect to the skeleton and the parameter identifiability criterion are the same for both models. Factorization of the tail limits is possible for both models under the same condition. The differences are

in the parameter space for $\theta = (c_e, e \in E)$ and the map $\theta \mapsto \{b_{ij}\}_{i,j} \in V$. However the common feature is that $\{b_{ij}\}$ factorize into terms, each of which contains only the edge weights within a certain tournament. Note however that this one depends on the graph, the ttt is a special graph with tree-like structure.

We can hope that the similarities between the two models might be the case for other properties known for the max-linear model, but not studied on the sum-linear model, or the other way around. For instance graph structure learning methods under the constraint that the graph is a tree of transitive tournaments with unique source might be the same for both models. This is hypothesis for further research because the different parameter spaces and coefficients $\{b_{ij}\}_{i,j\in V}$ can be important factors. Although not as popular as the max-linear models in extreme value literature the sum-linear Bayesian networks are used in several papers so far, such as Budhathoki et al. (2022), Zhao et al. (2021), and Pasche et al. (2021).

5.8 Supplement

5.8.1 Additional properties of a ttt

Lemma 5.8.1. *Let $\mathcal{T} = (V, E)$ be a ttt as in Definition 4.2.1.*

1. *Let $\mathcal{T}$ have a unique source. The skeleton of $\mathcal{T}$, T, is the same as the moral graph $\mathcal{T}^m$, hence $\mathcal{T}^m = T$.*
2. *Let $\mathcal{T}$ have a unique source. Let $A, B, S \subset V$ be nonempty, disjoint subsets. Separation of A and B by S in the skeleton T of $\mathcal{T}$ implies that A and B are d-separated by S in $\mathcal{T}$,i.e.*

$$A \perp_T B \mid S \implies d\text{-}sep_{\mathcal{T}}(A, B \mid S).$$

3. *Every path $p \in \pi(a, b)$ contains the unique shortest path, $p(a, b)$.*

Proof. 1. We need to show that in the graph T we don't need to add anymore edges between two nodes with common child in $\mathcal{T}$ and hence considering the skeleton of $\mathcal{T}$ is enough to obtain $\mathcal{T}^m$. Consider two nodes, a, b with common child in $\mathcal{T}$. Since there may not be v-structures by Lemma 4.6.1, the parents and the child must belong to the same tournament, then they are connected.

2. Assume that for nonempty disjoint sets $A, B, S \subset V$, the sets A and B are separated in T by S. Then all non-directed paths from A to B pass through S. Hence S must contain at least one separator node for every path from A to B. According to Lemma 4.6.1-2, none of the nodes may have parents from different tournaments. This means that all paths from A to B are not active according to Definition 3.6 of Koller and Friedman (2009) and hence that A and B are d-separated given S in $\mathcal{T}$ according to Definition 3.7 of the same book.

3. Let the unique shortest path be on the following node sequence $\{a = v_1, \ldots, v_n = b\}$. For a triple of consecutive nodes on this path v_{i-1}, v_i, v_{i+1} we must have that v_{i-1} and v_{i+1} belong to different tournaments as otherwise there will be an edge between the two and hence a shorter path– a contradiction to the assumption that the shortest path involves v_i.

Let there be a path $p \in \pi(a, b)$ for which a certain node from the unique shortest path, say v_r, does not belong to p. Instead, let the path p pass through nodes $\{v_{r-1}, \ldots, u, \ldots, v_{r+1}\}$.

The nodes v_r, u must belong to the same tournament, as otherwise there is a non directed cycle involving several tournaments, i.e., on the set of nodes $\{a, \ldots, v_r, \ldots, b\} \cup \{b, \ldots, u, \ldots, a\}$, but this is impossible according to property (P2). If u is in the

same tournament as v_r it can be either in the tournament determined by the edge $(v_{r-1}, v_r) \in E$, say τ_{r-1}, or in a tournament determined by the edge $(v_r, v_{r+1}) \in E$, say τ_{r+1}.

Assume u belongs to τ_{r-1}. The only possible way for u to have a directed path going to v_{r+1} without violating properties (P1) and (P2) of a ttt is that this path passes through v_r, hence every path from a to b containing u passes through v_r too. Similarly if u belong to τ_{r+1}: the only possible way for u to have a directed path coming from v_{r-1} without violating properties (P1) and (P2) of a ttt is that this path passes through v_r. □

5.8.2 Proofs for section 5.5

Lemma 5.8.2. *Let $(X_v, v \in V)$ be sum-linear graphical model as in Definition 5.2.1. Let τ be a given transitive tournament on nodes V_τ. Then for $u \in V_\tau$*

$$\mathcal{L}\left(\frac{X_v}{X_u}, v \in V_\tau \mid X_u > t\right) \xrightarrow{d} \mathcal{L}(M^{(u,\tau)}) = \mathcal{L}(M_{uv}, v \in V_\tau) = \sum_{j \in \mathrm{An}(u)} b_{uj} \delta_{\left\{\frac{c_{\pi(j,v)}}{c_{\pi(j,u)}}, v \in V_\tau\right\}}. \tag{5.27}$$

The vector $M^{(u,\tau)} = (M_{uv}, v \in V_\tau)$ has dependent variables and the distribution of a single element is as follows.

1. *The distribution of M_{uv} when $(u, v) \in E$.*
   (a) *If u is the source node of τ, the distribution is given by $\mathcal{L}(M_{uv}) = \delta_{\{c_{\pi(u,v)}\}}$.*
   (b) *If u is not the source node of τ, the distribution is given by*

$$\mathcal{L}(M_{uv}) = \sum_{j \in \mathrm{An}(u)} b_{uj} \delta_{\left\{\frac{c_{\pi(j,v)}}{c_{\pi(j,u)}}\right\}}.$$

2. *The distribution of M_{uv} when $(v, u) \in E$.*
   (a) *If v is the source node of τ, the distribution is given by*

$$\mathcal{L}(M_{uv}) = c_{\pi(v,u)} \delta_{\{1/c_{\pi(v,u)}\}} + (1 - c_{\pi(v,u)}) \delta_{\{0\}}.$$

   (b) *If v is not the source node of τ, the distribution is given by*

$$\mathcal{L}(M_{uv}) = \sum_{j \in \mathrm{An}(v)} b_{uj} \delta_{\left\{\frac{c_{\pi(j,v)}}{c_{\pi(j,u)}}\right\}} + \sum_{j \in \mathrm{An}(u) \setminus \mathrm{An}(v)} b_{uj} \delta_{\{0\}}.$$

Proof. The proof largely follows the one to the corresponding lemma in the chapter on max-linear models. The only difference is the in products over unique shortest path: in the max-linear model $c_{p(i,j)}$ is replaced by the sum over all path products, $c_{\pi(i,j)}$ in the sum-linear model. Also, an edge coefficient in the role of a path product over the unique shortest path between two adjacent nodes, c_{ij}, in the max-linear model, is replaced by the sum over all paths, $c_{\pi(i,j)}$. We elaborate the full proof, but details that are common for both sum- and max-linear models, might be omitted and we refer to Lemma 4.6.3.

From Proposition 5.5.1 we have the limit

$$\mathcal{L}(M^{(u,\tau)}) = \sum_{j \in V} b_{uj} \delta_{\left\{\frac{b_{vj}}{b_{uj}}, v \in V_\tau\right\}}.$$

Adapting this representation to exclude $b_{uj} = 0$ for $j \notin \mathrm{An}(u)$ and using $b_{ij} = c_{\pi(j,i)} b_{jj}$ for $j \in \mathrm{An}(i)$ we obtain

$$\sum_{j \in \mathrm{An}(u)} b_{uj} \delta_{\left\{\frac{c_{\pi(j,v)} b_{jj}}{c_{\pi(j,u)} b_{jj}}, v \in V_\tau\right\}}. \tag{5.28}$$

Recall that by definition $c_{\pi(i,i)} = 1$ and $c_{\pi(i,j)} = 0$ if $i \notin \mathrm{An}(j)$. In what follows Lemma 5.3.1 will be used all the time.

Next we show that $(M_{uv}, v \in V_\tau)$ are mutually dependent. When u is the source of τ then for all $j \in \mathrm{An}(u)$ the atom

$$\left(\frac{c_{\pi(j,v)}}{c_{\pi(j,u)}}, v \in V_\tau\right) = \left(\frac{c_{\pi(j,u)}c_{\pi(u,v)}}{c_{\pi(j,u)}}, v \in V_\tau\right) = (1; c_{\pi(u,v)}, v \in V_\tau \setminus u)$$

gets probability $\sum_{j\in\mathrm{An}(u)} b_{uj} = 1$. Hence $(M_{uv}, v \in V_\tau)$ are at the same time perfectly dependent and independent.

Assume u is not the source node. Let for brevity $V_\tau = \{1, 2, \ldots, m\}$: the nodes are labeled according to their order of out-degrees within τ: the source node of τ has $m-1$ (largest) out-degree and is labeled by 1, the node with out-degree $m-2$ is labeled as 2, etc. Note that for two nodes $i, i+1 \in V_\tau$ we have $c_{\pi(i,i+1)} = c_{i,i+1}$, because there is only one path from i to $i+1$, namely $(i, i+1) \in E_\tau$.

Consider u being the node 2. We have, thanks to the no-cycle property within a tournament $\mathrm{An}(2) = \mathrm{An}(1) \cup \{2\}$. For all $j \in \mathrm{An}(1)$ we have

$$\left(\frac{c_{\pi(j,v)}}{c_{\pi(j,2)}}, v = 1, \ldots, m\right) = \left(\frac{c_{\pi(j,1)}}{c_{\pi(j,1)}c_{12}}; 1; \frac{c_{\pi(j,1)}c_{\pi(1,v)}}{c_{\pi(j,1)}c_{12}}, v = 3, \ldots, m\right), \tag{5.29}$$

which is an atom of $(M_{2v}, v = 1, \ldots, m)$ with mass $\sum_{j\in\mathrm{An}(1)} b_{2j}$. This means that for the marginal distribution of M_{21} we have $\mathbb{P}(M_{21} = 1/c_{12}) \geq \sum_{j\in\mathrm{An}(1)} b_{2j}$. For $j = 2$ we have an atom $(0, 1, c_{\pi(2,3)}, \ldots, c_{\pi(2,m)})$ with mass b_{22}. This means that for the marginal probabilities of the variables $(M_{23}, \ldots, M_{2m})$ we have $\mathbb{P}(M_{2v} = c_{\pi(2,v)}) \geq b_{22}$, for all $v = 3, \ldots, m$. Take a vector of the following coordinates

$$(1/c_{12}, 1, c_{23}, c_{\pi(2,4)}, \ldots, c_{\pi(2,m)}).$$

Note that this vector cannot be the same as the one in (5.29). For $v = 3$ we cannot have $c_{23} = c_{\pi(1,3)}/c_{12} = (c_{13} + c_{12}c_{23})/c_{12}$, because this would require $c_{13} = 0$. However, edge weights must be all positive by assumption. The joint probability of the vector of coordinates $(1/c_{12}, 1, c_{23}, c_{\pi(2,4)}, \ldots, c_{\pi(2,m)})$ is

$$\mathbb{P}(M_{21} = 1/c_{12}, M_{22} = 1, M_{23} = c_{23}, M_{24} = c_{\pi(2,4)}, \ldots, M_{2m} = c_{\pi(2,m)}) = 0.$$

However the product of marginal probabilities is positive:

$$\mathbb{P}(M_{21} = 1/c_{12})\,\mathbb{P}(M_{22} = 1)\prod_{v=3}^{m}\mathbb{P}(M_{2v} = c_{\pi(2,v)}) \geq \sum_{j\in\mathrm{An}(1)} b_{2j} \times b_{22}^{m-1} > 0.$$

Now let $u \geq 3$. Take the vector of coordinates in (5.28) corresponding to $j = 1$ which is equal to $(1, c_{12}, c_{\pi(1,3)}, \ldots, c_{\pi(1,m)})/c_{\pi(1,u)}$ and has probability at least b_{u1}. Consider also the vector of coordinates for $j = u$ which is $(0, \ldots, 0, 1; c_{\pi(u,v)}, v = u+1, \ldots, m)$ with mass at least b_{uu}. Replace the first coordinate by $1/c_{\pi(1,u)}$. The vector obtained in this way has joint probability zero. For every $j \in \mathrm{pa}(u)$ we have $b_{vj}/b_{uj} = 0$ when v is not child of j or equivalently, given the order in the node labeling, when $v < j$. So for fixed $u \geq 3$, for $j = 1$ the vector $(b_{vj}/b_{uj}, v = 1, \ldots, m)$ has no zeros. For $j = 2$ the vector $(b_{vj}/b_{uj}, v = 1, \ldots, m)$ has one zero, namely $(0; b_{vj}/b_{uj}, v = 2, \ldots, m)$, for $j = 3$ the vector $(b_{vj}/b_{uj}, v = 1, \ldots, m)$ has two zeros, namely $(0, 0; b_{vj}/b_{uj}, v = 3, \ldots, m)$ and so on until $j = u$ with the corresponding vector $(b_{vj}/b_{uj}, v = 1, \ldots, m) = (0, \ldots, 0; b_{vj}/b_{uj}, v = u, \ldots, m)$. By replacing the first coordinate by a non-zero value in this vector we get an impossible value for the random vector $(M_{uv}, v = 1, \ldots, m)$ or a value with probability zero. Considering the

univariate marginal distributions of $(M_{uv}, v = 1, \ldots, m)$ we obtain for the product of marginal probabilities a positive value, namely:

$$\mathbb{P}(M_{u1} = 1/c_{\pi(1,u)}) \left[\prod_{v=2}^{u-1} \mathbb{P}(M_{uv} = 0)\right] \mathbb{P}(M_{uu} = 1) \prod_{v=u+1}^{m} \mathbb{P}(M_{uv} = c_{\pi(u,v)}) \geq b_{u1} \times b_{uu}^{m-1} > 0$$

This shows that for any $u \in V_\tau$ the vector $(M_{u1}, \ldots, M_{um})$ has jointly dependent elements.

Next we show the distribution of a single element $M_{uv}, v \in V_\tau \setminus u$.

1. Consider first when u is the source node in τ. The same reasoning holds as in the corresponding lemma for the max-linear model. Hence for $j \in \mathrm{An}(u)$

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{\pi(j,u)} c_{\pi(u,v)} b_{jj}}{c_{\pi(j,u)} b_{jj}} = c_{\pi(u,v)},$$

and since $\sum_{j \in \mathrm{An}(u)} b_{uj} = 1$ we obtain the desired result under 1.(a).

When u is not the source node in τ not all shortest paths to v pass through u hence for $j \in \mathrm{An}(u)$ we have

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{\pi(j,v)} b_{jj}}{c_{\pi(j,u)} b_{jj}} = \frac{c_{\pi(j,v)}}{c_{\pi(j,u)}} > 0,$$

with mass b_{uj}. Hence the result in 1.(b). Note that zero is not possible value as we still have $\mathrm{An}(u) \subset \mathrm{An}(v)$. Also $c_{\pi(u,u)} = 1$ by convention.

2. Let us have now $(v, u) \in E$. When v is a source node in τ, we have, for $j \in \mathrm{An}(v)$

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{\pi(j,v)} b_{jj}}{c_{\pi(j,u)} b_{jj}} = \frac{c_{\pi(j,v)}}{c_{\pi(j,v)} c_{\pi(v,u)}} = \frac{1}{c_{\pi(v,u)}} > 0,$$

which is an atom with probability

$$\sum_{j \in \mathrm{An}(v)} b_{uj} = \sum_{j \in \mathrm{An}(v)} c_{\pi(j,v)} c_{\pi(v,u)} b_{jj} = c_{\pi(v,u)} \sum_{j \in \mathrm{An}(v)} c_{\pi(j,v)} b_{jj} = c_{\pi(v,u)} \sum_{j \in \mathrm{An}(v)} b_{vj} = c_{\pi(v,u)}.$$

The probability of the zero atom is

$$\sum_{j \in \mathrm{An}(u) \setminus \mathrm{An}(v)} b_{uj} = \sum_{j \in \mathrm{An}(u)} b_{uj} - \sum_{j \in \mathrm{An}(v)} b_{uj} = 1 - c_{\pi(v,u)}.$$

This shows 2.(a).

When v is not a source node of τ we have for $j \in \mathrm{An}(v)$

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{\pi(j,v)} b_{jj}}{c_{\pi(j,u)} b_{jj}} = \frac{c_{\pi(j,v)}}{c_{\pi(j,u)}} > 0,$$

an atom with mass b_{uj} and zero atom with probability $\sum_{j \in \mathrm{An}(u) \setminus \mathrm{An}(v)} b_{uj}$. This shows 2.(b). □

Lemma 5.8.3. *Let $(X_v, v \in V)$ be sum-linear graphical model as in Definition 5.2.1. Let $\mathcal{T}$ have unique source. For any $u \in V$ we have*

$$\mathcal{L}\left(\frac{X_v}{X_u}, v \in V \mid X_u > t\right) \xrightarrow{d} \mathcal{L}(A_{uv}, v \in V) = \sum_{j \in \mathrm{An}(u)} b_{uj} \delta_{\left\{\frac{c_{\pi(j,v)}}{c_{\pi(j,u)}}, v \in V\right\}}. \tag{5.30}$$

The distribution of A_{uv} depends on the three types of possible trails according to Lemma 4.6.2-2. In what follows we assume that u, v are not adjacent. For the case $(u, v) \in E$ see Lemma 5.8.2.

1. *Distribution of A_{uv} on a path $\{u = v_1, r = v_2, \ldots, v = v_n\}$ with $u, r \in \tau$, one of the tournaments of $\mathcal{T}$.*
   (a) *If u is a source node in τ then $\mathcal{L}(A_{uv}) = \delta_{\{c_{\pi(u,v)}\}}$.*
   (b) *If u is not a source node in τ we have*

$$\mathcal{L}(A_{uv}) = \sum_{j \in \mathrm{An}(u)} b_{uj} \delta_{\left\{\frac{c_{\pi(j,r)}}{c_{\pi(j,u)}} c_{\pi(r,v)}\right\}}.$$

2. *Distribution of A_{uv} on a path $\{v = v_1, r = v_2, \ldots, u = v_n\}$ with $v, r \in \tau$.*
   (a) *If v is a source node in τ then*

$$\mathcal{L}(A_{uv}) = c_{\pi(v,u)} \delta_{\left\{\frac{1}{c_{\pi(v,u)}}\right\}} + (1 - c_{\pi(v,u)}) \delta_{\{0\}}.$$

   (b) *If v is not a source node in τ then*

$$\mathcal{L}(A_{uv}) = \sum_{j \in \mathrm{An}(v)} c_{\pi(r,u)} b_{rj} \delta_{\left\{\frac{c_{\pi(j,v)}}{c_{\pi(j,r)} c_{\pi(r,u)}}\right\}} + \sum_{j \in \mathrm{An}(u) \setminus \mathrm{An}(v)} b_{uj} \delta_{\{0\}}.$$

3. *The distribution of A_{uv} on a trail composed of two paths $p(r, u)$ and $p(r, v)$. Let the trail be on nodes $\{u, \ldots, m, r, n, \ldots, v\}$. Let also τ_m, τ_n be two tournaments with $r, m \in \tau_m$ and $r, n \in \tau_n$.*
   (a) *If r is source in both τ_m and τ_n, then*

$$\mathcal{L}(A_{uv}) = c_{\pi(r,u)} \delta_{\left\{\frac{c_{\pi(r,v)}}{c_{\pi(r,u)}}\right\}} + (1 - c_{\pi(r,u)}) \delta_{\{0\}}.$$

   (b) *If r is source in τ_m, but not in τ_n, then*

$$\mathcal{L}(A_{uv}) = \sum_{j \in \mathrm{An}(r)} c_{\pi(r,u)} b_{rj} \delta_{\left\{\frac{c_{\pi(j,n)} c_{\pi(n,v)}}{c_{\pi(j,r)} c_{\pi(r,u)}}\right\}} + \sum_{j \in \mathrm{An}(u) \setminus \mathrm{An}(r)} b_{uj} \delta_{\{0\}}.$$

   (c) *If r is source in τ_n, but not in τ_m, then*

$$\mathcal{L}(A_{uv}) = \sum_{j \in \mathrm{An}(r)} c_{\pi(m,u)} b_{mj} \delta_{\left\{\frac{c_{\pi(j,r)} c_{\pi(r,v)}}{c_{\pi(j,m)} c_{\pi(m,u)}}\right\}} + \sum_{j \in \mathrm{An}(u) \setminus \mathrm{An}(r)} b_{uj} \delta_{\{0\}}.$$

Proof. Again, the proof is the same as the one for the max-linear model of Lemma 4.6.4. Because of Lemma 5.3.1 we only need to change $c_{p(i,j)}$ by $c_{\pi(i,j)}$. We, however, elaborate only the first four cases, omitting details that refer to properties of the graph and that can be found in Lemma 4.6.4. We repeat some of the pictures from Lemma 4.6.4 for the ease of the reader.

From Proposition 5.5.1 we have the limit

$$\mathcal{L}\left(\frac{X_v}{X_u}, v \in V \mid X_u > t\right) \xrightarrow{d} \sum_{j \in V} b_{uj} \delta_{\left\{\frac{b_{vj}}{b_{uj}}, v \in V\right\}}.$$

Adapting this representation to eliminate non defined atoms with zero probabilities when $b_{uj} = 0, j \notin \mathrm{An}(u)$, and using $b_{ij} = c_{\pi(j,i)} b_{jj}$ for $j \in \mathrm{An}(i)$ we obtain

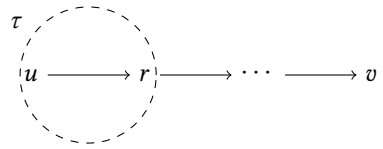
$$\sum_{j \in \mathrm{An}(u)} b_{uj} \delta_{\left\{\frac{c_{\pi(j,v)}}{c_{\pi(j,u)}}, v \in V\right\}}.$$

Recall that $c_{\pi(i,i)} = 1$ and $c_{\pi(i,j)} = 0$ if $i \notin \mathrm{An}(j)$. For a single $v \in V \setminus u$ we have the marginal distribution

$$\mathcal{L}(A_{uv}) = \sum_{j \in \mathrm{An}(u)} b_{uj} \delta_{\left\{\frac{b_{vj}}{b_{uj}}\right\}}. \tag{5.31}$$

As for the max-linear model the distribution of A_{uv} depends in a deterministic way on properties of the ttt which are stated in Lemma 4.6.2.

First we deal with 1.(a). The case is illustrated by the graph below. All paths from $\mathrm{An}(u)$ to v pass through u because u is source in τ.

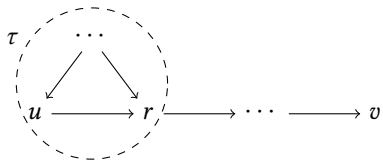


Hence for all $j \in \mathrm{An}(u)$ we have

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{\pi(j,v)} b_{jj}}{c_{\pi(j,u)} b_{jj}} = \frac{c_{\pi(j,u)} c_{\pi(u,v)}}{c_{\pi(j,u)}} = c_{\pi(u,v)} > 0,$$

with mass $\sum_{j \in \mathrm{An}(u)} b_{uj} = 1$.

Next we show 1.(b). The unique shortest path from $j \in \mathrm{An}(u)$ to v pass through r, hence according to Lemma 5.3.1 $c_{\pi(j,v)} = c_{\pi(j,r)} c_{\pi(r,v)}$.

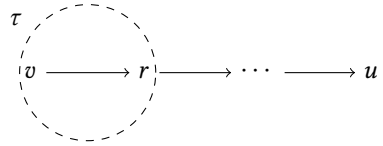


We have for $j \in \mathrm{An}(u)$

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{\pi(j,v)} b_{jj}}{c_{\pi(j,u)} b_{jj}} = \frac{c_{\pi(j,r)}}{c_{\pi(j,u)}} c_{\pi(r,v)} > 0,$$

with mass b_{uj}, hence the expression in 1.(b).

Next we show 2.(a). When the directed path is from v to u zero is a possible value of A_{uv}. All shortest paths from $j \in \mathrm{An}(v)$ to u pass through v as v is source in τ.



For $j \in \mathrm{An}(v)$ the non-zero atom is given by

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{\pi(j,v)} b_{jj}}{c_{\pi(j,u)} b_{jj}} = \frac{c_{\pi(j,v)}}{c_{\pi(j,v)} c_{\pi(v,u)}} = \frac{1}{c_{\pi(v,u)}} > 0, \qquad j \in \mathrm{An}(v),$$

with mass

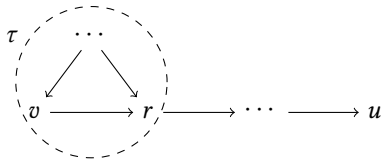
$$\begin{aligned} \sum_{j \in \mathrm{An}(v)} b_{uj} &= \sum_{j \in \mathrm{An}(v)} c_{\pi(j,v)} c_{\pi(v,u)} b_{jj} = c_{\pi(v,u)} \sum_{j \in \mathrm{An}(v)} c_{\pi(j,v)} b_{jj} \\ &= c_{\pi(v,u)} \sum_{j \in \mathrm{An}(v)} b_{vj} = c_{\pi(v,u)}. \end{aligned}$$

For the zero atom we have probability

$$\sum_{j\in \mathrm{An}(u)\setminus \mathrm{An}(v)} b_{uj} = \sum_{j\in \mathrm{An}(u)} b_{uj} - \sum_{j\in \mathrm{An}(v)} b_{uj} = 1 - c_{\pi(v,u)}.$$

This shows 2.(a).

To show 2.(b) we note that the unique shortest path from $j \in \mathrm{An}(v)$ to u pass through r, hence by Lemma 5.3.1 $c_{\pi(j,u)} = c_{\pi(j,r)}c_{\pi(r,u)}$.



Hence for $j \in \mathrm{An}(v)$ we have

$$\frac{b_{vj}}{b_{uj}} = \frac{c_{\pi(j,v)}b_{jj}}{c_{\pi(j,u)}b_{jj}} = \frac{c_{\pi(j,v)}}{c_{\pi(j,r)}c_{\pi(r,u)}} > 0,$$

which is an atom with mass $b_{uj} = c_{\pi(j,r)}c_{\pi(r,u)}b_{jj} = c_{\pi(r,u)}b_{rj}$. The zero atom comes from the fact that $b_{vj} = 0$ for all $j \in \mathrm{An}(u) \setminus \mathrm{An}(v)$, and it has probability $\sum_{j\in \mathrm{An}(u)\setminus \mathrm{An}(v)} b_{uj}$. This shows the distribution under 2.(b).

For the distributions in cases 3.(a)-(c) we refer to Lemma 4.6.4 where we replace $c_{p(i,j)}$ by $c_{\pi(i,j)}$ as we did in the four cases above. Again this is possible thanks to Lemma 5.3.1. □

Proof of Proposition 5.5.2

Proof. First we prove that (i) implies (ii). As in the proofs of the previous two lemmas the proof laregely utilizes the corresponding proof for the max-linear model. We present the complete proof.

Assume $\mathcal{T}$ has a unique source. We have to prove that for any $u \in V$ an element from the limiting vector in (5.24) is given by (5.25).

In Lemma 5.8.3 we have seen seven possible cases for the distribution of A_{uv} depending on deterministic properties of the trail between u and v. Below we consider each of these seven cases again.

1. Let the unique shortest trail between u and v be a path on node sequence $\{u = v_1, r = v_2, \ldots, v_n = v\}$. Let τ be the tournament containing u, r.

Case 1.(a) Let u be source in τ. From Lemma 5.8.3-1.(a) we have $P(A_{uv} = c_{\pi(u,v)}) = 1$. Consider the variables $(M_e, e \in p(u,v))$ which are by construction independent between each other because they belong to different tournaments. Note that in this case all nodes $v_1, \ldots, v_{n-1}$ are source nodes in the tournament containing that node and the next one in the sequence. This follows from Lemma 4.6.2-1. Then according to Lemma 5.8.2 1.(a) for every $M_{ij}, (i,j) \in p(u,v)$ we have $\mathbb{P}(M_{ij} = c_{\pi(i,j)}) = 1$ and hence

$$\mathbb{P}\Big(\prod_{(i,j)\in p(u,v)} M_{ij} = c_{\pi(u,v)}\Big) = \prod_{(i,j)\in p(u,v)} \mathbb{P}(M_{ij} = c_{\pi(i,j)}) = 1,$$

which shows $A_{uv} = \prod_{(i,j)\in p(u,v)} M_{ij}$.

Case 1.(b). If u is not the source in τ, the distribution of M_{ur} is as in Lemma 5.8.2-1.(b). As in the case 1.(a) all nodes $r = v_2, v_3, \ldots, v_{n-1}$ are source nodes in the tournament containing that node and the next one in the sequence. The variables $M_{ij}, (i,j) \in p(r,v)$ are degenerate at $c_{\pi(i,j)}$, hence their product is degenerate at $\prod_{(i,j)\in p(r,v)} c_{\pi(i,j)}$, which by Lemma 5.3.1 equals $c_{\pi(r,v)}$. As the case 1.(a) above the variables $(M_{ij}, (i,j) \in p(u,v))$ are by construction independent between each other

because they are indexed by edges which belong to different tournaments. Then we have

$$\begin{aligned}\mathcal{L}\Big(\prod_{(i,j)\in p(u,v)} M_{ij}\Big) &= \mathcal{L}\Big(M_{ur}\prod_{(i,j)\in p(r,v)} M_{ij}\Big)\\ &= \left(\sum_{j\in \mathrm{An}(u)} b_{uj}\delta_{\left\{\frac{c_{\pi(j,r)}}{c_{\pi(j,u)}}\right\}}\right)\otimes \delta_{\{c_{\pi(r,v)}\}} = \sum_{j\in \mathrm{An}(u)} b_{uj}\delta_{\left\{\frac{c_{\pi(j,r)}}{c_{\pi(j,u)}}c_{\pi(r,v)}\right\}}.\end{aligned} \tag{5.32}$$

The sign $\otimes$ has the same meaning as in the proof of the corresponding proposition for the max-linear model. The last one expression in (5.32) is the distribution of A_{uv} in Lemma 5.8.3-1.(b).

2. Let the unique shortest trail between u and v be a path from v to u on the node sequence $\{v = v_1, r = v_2, \ldots, v_n = u\}$. Let τ be the tournament containing v, r.

Case 2.(a). Let v be source in τ. Consider the random variables M_{v_{i+1},v_i} for $i = 1, \ldots, n-1$ whose distributions are as in Lemma 5.8.2-2.(a). Since this is the unique shortest trail from v to u, all edges on it belong to different tournaments and the vector $(M_{v_{i+1},v_i}, i = 1, \ldots, n-1)$ contains independent variables by definition. Then

$$\begin{aligned}\mathbb{P}\Big(\prod_{i=1}^{n-1} M_{v_{i+1},v_i} = \frac{1}{c_{\pi(v,u)}}\Big) &= \prod_{i=1}^{n-1}\mathbb{P}\left(M_{v_{i+1},v_i} = \frac{1}{c_{\pi(v_i,v_{i+1})}}\right)\\ &= \prod_{i=1}^{n-1} c_{\pi(v_i,v_{i+1})} = c_{\pi(v,u)}.\end{aligned} \tag{5.33}$$

In the second and last equality we used Lemma 5.3.1 according to which $c_{\pi(v,u)} = \prod_{i=1}^{n-1} c_{\pi(v_i,v_{i+1})}$. For the zero atom we have

$$\begin{aligned}\mathbb{P}\Big(\prod_{i=1}^{n-1} M_{v_{i+1},v_i} = 0\Big) &= 1 - \prod_{i=1}^{n-1}\mathbb{P}(M_{v_{i+1},v_i} > 0)\\ &= 1 - \prod_{i=1}^{n-1}\mathbb{P}\left(M_{v_{i+1},v_i} = \frac{1}{c_{\pi(v_i,v_{i+1})}}\right) = 1 - c_{\pi(v,u)}.\end{aligned} \tag{5.34}$$

The expressions in (5.33) and (5.34) represent indeed the distribution of A_{uv} in Lemma 5.8.3-2.(a).

Case 2.(b). If v is not source in τ consider a random variable M_{rv} with distribution as in Lemma 5.8.2-2.(b) and a random variable A_{ur} constructed as in case 2.(a) here above, i.e., as the product $\prod_{i=2}^{n-1} M_{v_{i+1},v_i}$. By construction M_{rv} is independent from A_{ur} with the same argument as above. We have

$$\begin{aligned}\mathcal{L}(A_{ur}M_{rv}) = &\left(c_{\pi(r,u)}\delta_{\left\{\frac{1}{c_{\pi(r,u)}}\right\}} + (1 - c_{\pi(r,u)})\delta_{\{0\}}\right)\\ &\otimes\left(\sum_{j\in \mathrm{An}(v)} b_{rj}\delta_{\left\{\frac{c_{\pi(j,v)}}{c_{\pi(j,r)}}\right\}} + \sum_{j\in \mathrm{An}(r)\setminus \mathrm{An}(v)} b_{rj}\delta_{\{0\}}\right),\end{aligned}$$

which gives non-zero atoms $c_{\pi(j,v)}/(c_{\pi(j,r)}c_{\pi(r,u)})$, $j \in \mathrm{An}(v)$ with masses $b_{rj}c_{\pi(r,u)}$, $j \in \mathrm{An}(v)$. To show the probability of the zero atom, consider

$$\begin{aligned}\mathbb{P}(A_{ur}M_{rv} = 0) &= 1 - \mathbb{P}(A_{ur} > 0)\,\mathbb{P}(M_{rv} > 0) = 1 - c_{\pi(r,u)}\sum_{j\in \mathrm{An}(v)} b_{rj}\\ &= \sum_{j\in \mathrm{An}(u)} b_{uj} - \sum_{j\in \mathrm{An}(v)} c_{\pi(j,r)}b_{jj}c_{\pi(r,u)}\end{aligned}$$

$$= \sum_{j\in \mathrm{An}(u)} b_{uj} - \sum_{j\in \mathrm{An}(v)} b_{uj} = \sum_{j\in \mathrm{An}(u)\setminus \mathrm{An}(v)} b_{uj},$$

which is what we need to confirm $A_{uv} = A_{ur}M_{rv}$ where A_{uv} is as in Lemma 5.8.3-2.(b).

3. In the three cases that follow let the unique shortest trail from u to v be given by two paths $p(r,u)$ and $p(r,v)$. Let the trail be on nodes $\{u,\dots,m,r,n,\dots,v\}$. Let also τ_m, τ_n be two tournaments with $r, m \in \tau_m$ and $r, n \in \tau_n$.

Case 3.(a) Let r be source in both τ_m and τ_n. Consider random variables A_{rv} as in Lemma 5.8.3-1.(a) and A_{ur} as in Lemma 5.8.3-2.(a). In the present proof we have shown in cases 1.(a) and 2.(a) that A_{rv} and A_{ur} are factorizable in independent multiplicative increments. By construction, A_{rv} and A_{ur} are independent from each other, because the multiplicative increments are independent. We have

$$\mathbb{P}\left(A_{ur}A_{rv} = \frac{c_{\pi(r,v)}}{c_{\pi(r,u)}}\right) = \mathbb{P}\left(A_{ur} = \frac{1}{c_{\pi(r,u)}}\right)\mathbb{P}(A_{rv} = c_{\pi(r,v)}) = c_{\pi(r,u)}.$$

For the probability of the zero atom we have

$$\mathbb{P}(A_{ur}A_{rv} = 0) = P(A_{ur} = 0) = (1 - c_{\pi(r,u)}).$$

The two displays above represent indeed the distribution of A_{uv} in Lemma 5.8.3-3.(a).

Case 3.(b) Let r be source in τ_m, but not source in τ_n. Consider three random variables A_{ur}, M_{rn}, A_{nv} with distributions as in Lemma 5.8.3-2.(a), Lemma 5.8.2-1.(b) and Lemma 5.8.3-1.(a) respectively. For A_{ur} and A_{nv} we have shown in cases 2.(a) and 1.(a) in this proof that they are factorizable in independent multiplicative increments. By construction M_{rn} is independent from the increments in A_{ur} and A_{nv}. Then

$$\begin{aligned}\mathcal{L}(A_{ur}M_{rn}A_{nv}) &= \left(c_{\pi(r,u)}\delta_{\left\{\frac{1}{c_{\pi(r,u)}}\right\}} + (1 - c_{\pi(r,u)})\delta_{\{0\}}\right) \\ &\quad \otimes \left(\sum_{j\in \mathrm{An}(r)} b_{rj}\delta_{\left\{\frac{c_{\pi(j,n)}}{c_{\pi(j,r)}}\right\}}\right) \otimes \delta_{\{c_{\pi(n,v)}\}} \\ &= \sum_{j\in \mathrm{An}(r)} b_{rj}c_{\pi(r,u)}\delta_{\left\{\frac{c_{\pi(j,n)}c_{\pi(n,v)}}{c_{\pi(j,r)}c_{\pi(r,u)}}\right\}} + (1 - c_{\pi(r,u)})\delta_{\{0\}}.\end{aligned}$$

Note that

$$\begin{aligned}\sum_{j\in \mathrm{An}(u)\setminus \mathrm{An}(r)} b_{uj} &= \sum_{j\in \mathrm{An}(u)} b_{uj} - \sum_{j\in \mathrm{An}(r)} b_{uj} = 1 - \sum_{j\in \mathrm{An}(r)} c_{\pi(j,r)}c_{\pi(r,u)}b_{jj} \\ &= 1 - c_{\pi(r,u)} \sum_{j\in \mathrm{An}(r)} b_{rj} = 1 - c_{\pi(r,u)}.\end{aligned}$$

In the second equality above we used $c_{\pi(j,u)} = c_{\pi(j,r)}c_{\pi(r,u)}$ for $j \in \mathrm{An}(r)$ according to Lemma 5.3.1, because the unique shortest path from $j \in \mathrm{An}(r)$ to u passes through r because r is source of τ_m. This shows that the distribution of $A_{ur}M_{rn}A_{nv}$ is the one of A_{uv} in Lemma 5.8.3-3.(b).

Case 3.(c) Let r be source in τ_n, but not in τ_m. Consider variables A_{um}, M_{mr},and A_{rv} with distributions as in Lemma 5.8.3-2.(a), Lemma 5.8.2-2.(b) and Lemma 5.8.3-1.(a) respectively. The variables A_{um} and A_{rv} have been shown to factorize in independent increments in cases 2.(a) and 1.(a) of this proof respectively, hence they are independent from each other too. By construction M_{mr} is independent from A_{um} and A_{rv}. Then we have

$$\mathcal{L}(A_{um}M_{mr}A_{rv}) = \left(c_{\pi(m,u)}\delta_{\left\{\frac{1}{c_{\pi(m,u)}}\right\}} + (1 - c_{\pi(m,u)})\delta_{\{0\}}\right)$$

$$\otimes\left(\sum_{j\in\mathrm{An}(r)} b_{mj}\delta_{\left\{\frac{c_{\pi(j,r)}}{c_{\pi(j,m)}}\right\}} + \sum_{j\in\mathrm{An}(m)\setminus\mathrm{An}(r)} b_{mj}\delta_{\{0\}}\right)\otimes\delta_{\{c_{\pi(r,v)}\}}.$$

The non-zero atoms are $c_{\pi(j,r)}c_{\pi(r,v)}/(c_{\pi(j,m)}c_{\pi(m,u)})$ for $j\in\mathrm{An}(r)$ with masses $c_{\pi(m,u)}b_{mj} = c_{\pi(j,m)}c_{\pi(m,u)}b_{jj}$ By Lemma 5.3.1 since the unique shortest path from $j\in\mathrm{An}(r)$ to node u passes through node m we have $c_{\pi(j,m)}c_{\pi(m,u)}b_{jj} = c_{\pi(j,u)}b_{jj} = b_{uj}$. The probability of the zero atom is given by

$$\begin{aligned}
&\mathbb{P}(A_{um}M_{mr}A_{rv}=0) = 1-\mathbb{P}(A_{um}>0)\,\mathbb{P}(M_{mr}>0)\\
&= 1-c_{\pi(m,u)}\sum_{j\in\mathrm{An}(r)} b_{mj} = 1-\sum_{j\in\mathrm{An}(r)} c_{\pi(j,m)}c_{\pi(m,u)}b_{jj}\\
&= \sum_{j\in\mathrm{An}(u)} b_{uj} - \sum_{j\in\mathrm{An}(r)} b_{uj} = \sum_{j\in\mathrm{An}(u)\setminus\mathrm{An}(r)} b_{uj}.
\end{aligned}$$

Hence the distribution of $A_{um}M_{mr}A_{rv}$ is the one of A_{uv} in Lemma 5.8.3-3.(c). This completes the proof that the statement in (i) implies (ii).

The statement in (iii) holds trivially from (ii).

Next we prove that (iii) implies (i) by contraposition: we assume that $\mathcal{T}$ has at least two sources and we will show that it is not possible to obtain the factorization in (5.25). Consider again as in the proof of Proposition 4.3.1 for the max-linear model the v-structure on nodes 1, 2, 3 and the two trails $t(u,1)$ and $t(u,2)$.

First we take the case when, w.l.o.g., the v-structure belongs to $t(u,2)$ but not to $t(u,1)$. Let the trail from 1 to u be on nodes $\{v_1 = 1, v_2, \ldots, v_n = u\}$. With the same arguments as for the max-linear model we conclude that A_{u2} is degenerate at zero. As for the variables $(M_{v_{i+1},v_i}, i = n-1,\ldots,1; M_{13}, M_{32})$ which by construction are independent as they belong to different tournaments and distributed according to Lemma 5.8.2, we see that none of these is degenerate at zero. Hence their product cannot be degenerate at zero too.

Next we consider the second case, when w.l.o.g. $(1,3)\in t(u,1)$ and $(2,3)\in t(u,2)$. Let the trail from node 3 to u be on nodes $\{v_1 = 3, v_2, \ldots, v_n = u\}$. First we consider the case when we have at least one $i = 1,\ldots,n-1$ for which $(v_{i+1}, v_i)\in E$, i.e., we have at least one edge with direction from u to 3. For the same reasons as in the proof for the max-linear model we conclude $\mathcal{L}(A_{u1}, A_{u2}) = \delta_{\{0,0\}}$. But then again the factorisation (5.24)–(5.25) cannot hold, as according to Lemma 5.8.2 the increments are never degenerate at zero.

Now let the trail from node 3 to u be actually a path. Let also nodes 1 and 2 be sources with respect to the tournaments shared with node 3, say $1,3\in V_{\tau_1}$ and $2,3\in V_{\tau_2}$. It is always possible to choose 1 and 2 in such a way they are the sources of τ_1 and τ_2. This is because node 3 obviously is not a source in τ_1 and τ_2, so the sources of these must point to 3. We can decompose $\mathrm{An}(u)$ into three disjoint sets, $\mathrm{An}(1)$, $\mathrm{An}(2)$ and the rest, $\mathrm{An}(u)\setminus\{\mathrm{An}(1)\cup\mathrm{An}(2)\}$. For the distribution of (A_{u1}, A_{u2}) we have

$$\begin{aligned}
\mathcal{L}(A_{u1},A_{u2}) &= \sum_{j\in\mathrm{An}(u)} b_{uj}\delta_{\left\{\frac{b_{1j}}{b_{uj}},\frac{b_{2j}}{b_{uj}}\right\}}\\
&= \sum_{j\in\mathrm{An}(1)} b_{uj}\delta_{\left\{\frac{b_{1j}}{b_{uj}},\frac{b_{2j}}{b_{uj}}\right\}} + \sum_{j\in\mathrm{An}(2)} b_{uj}\delta_{\left\{\frac{b_{1j}}{b_{uj}},\frac{b_{2j}}{b_{uj}}\right\}}\\
&\quad + \sum_{j\in\mathrm{An}(u)\setminus\{\mathrm{An}(1)\cup\mathrm{An}(2)\}} b_{uj}\delta_{\left\{\frac{b_{1j}}{b_{uj}},\frac{b_{2j}}{b_{uj}}\right\}}.
\end{aligned}$$

For the atoms in the first summation we have

$$\frac{b_{1j}}{b_{uj}} = \frac{c_{\pi(j,1)}b_{jj}}{c_{\pi(j,u)}b_{jj}} = \frac{c_{\pi(j,1)}}{c_{\pi(j,1)}c_{\pi(1,3)}c_{\pi(3,u)}} = \frac{1}{c_{\pi(1,3)}c_{\pi(3,u)}}$$

and $b_{2j}/b_{uj} = 0$ as $b_{2j} = 0$ for all $j \in \mathrm{An}(1)$. Hence we have an atom that does not depend on $j \in \mathrm{An}(1)$, i.e., $\big(1/(c_{\pi(1,3)}c_{\pi(3,u)}), 0\big)$ and its mass is

$$\sum_{j\in\mathrm{An}(1)} b_{uj} = \sum_{j\in\mathrm{An}(1)} c_{\pi(j,1)}c_{\pi(1,3)}c_{\pi(3,u)}b_{jj} = c_{\pi(1,3)}c_{\pi(3,u)} = c_{\pi(1,u)}.$$

In a similar way, from the second summation in the last display we have an atom $\big(0, 1/(c_{\pi(2,3)}c_{\pi(3,u)})\big)$ with mass $c_{\pi(2,3)}c_{\pi(3,u)} = c_{\pi(2,u)}$. In the third summation term the atom is $(0,0)$ as $b_{1j} = b_{2j} = 0$ for all $j \in \mathrm{An}(u) \setminus \{\mathrm{An}(1) \cup \mathrm{An}(2)\}$ and its mass is $1 - c_{\pi(1,3)}c_{\pi(3,u)} - c_{\pi(2,3)}c_{\pi(3,u)} = 1 - c_{\pi(3,u)}(c_{\pi(1,3)} + c_{\pi(2,3)})$. Consider now the multiplicative increments $(M_{31}; M_{32}, M_{v_{i+1},v_i} i = 1, \ldots, n-1)$ which are mutually independent since they belong to different tournaments. Because node 1 is a source node in the tournament τ_1 the distribution of M_{31} is $c_{\pi(1,3)}\delta_{\{1/c_{\pi(1,3)}\}} + (1 - c_{\pi(1,3)})\delta_{\{0\}}$ by Lemma 5.8.2-2.(a). Similarly for M_{32}. Similarly to (4.30) for the max-linear model in the previous chapter

$$\begin{aligned}&\mathbb{P}\left(M_{31}\prod_{i=1}^{n-1} M_{v_{i+1},v_i} = 0, M_{32}\prod_{i=1}^{n-1} M_{v_{i+1},v_i} = 0\right)\\ &= 1 - \prod_{i=1}^{n-1}\mathbb{P}(M_{v_{i+1},v_i} > 0)(c_{13} + c_{23} - c_{13}c_{23}).\end{aligned} \tag{5.35}$$

When all nodes in $\{v_1 = 3, v_2, \ldots, v_{n-1}\}$ are source nodes with respect to the tournament involving the next node in the sequence, then $\mathbb{P}(M_{v_{i+1},v_i} > 0) = c_{\pi(v_i,v_{i+1})}$ for $i = 1, \ldots, n-1$, and the probability in (5.35) equals $1 - c_{\pi(3,u)}(c_{\pi(1,3)} + c_{\pi(2,3)} - c_{\pi(1,3)}c_{\pi(2,3)})$, which is different than $\mathbb{P}(A_{u1} = 0, A_{u2} = 0) = 1 - c_{\pi(3,u)}(c_{\pi(1,3)} + c_{\pi(2,3)})$. In the second sub-case, i.e., if at least one node from $\{v_1 = 3, v_2, \ldots, v_{n-1}\}$ is not source with respect to the tournament involving the next node in the sequence then the possible values for $M_{31}\prod_{i=1}^{n-1} M_{v_{i+1},v_i}$ are not only $\{0, 1/c_{\pi(1,u)}\}$, which are the only possible values of A_{u1} as we showed in the previous paragraph. Let $i \in \{1, \ldots, n-1\}$ be such that node v_i is not the source node in the tournament shared with v_{i+1}, say τ_i. Recall the distribution of M_{v_{i+1},v_i} from Lemma 5.8.2-2.(b):

$$\mathcal{L}(M_{v_{i+1},v_i}) = \sum_{j\in\mathrm{An}(v_i)} b_{v_{i+1},j}\delta_{\{b_{v_i,j}/b_{v_{i+1},j}\}} + \sum_{j\in\mathrm{An}(v_{i+1})\setminus\mathrm{An}(v_i)} \delta_{\{0\}}.$$

Take for instance a node, say s, such that it is the only parent of v_i in the tournament τ_i. Then

$$\frac{b_{v_i,s}}{b_{v_{i+1},s}} = \frac{c_{sv_i}b_{ss}}{c_{\pi(s,v_{i+1})}b_{ss}}$$

is a possible value of M_{v_{i+1},v_i} with positive probability, namely at least $b_{v_{i+1},s}$. Another possible positive value is for $j = v_i \in \mathrm{An}(v_i)$, namely

$$\frac{b_{v_i,v_i}}{b_{v_{i+1},v_i}} = \frac{1}{c_{\pi(v_i,v_{i+1})}}$$

with probability at least b_{v_{i+1},v_i}. Note that we can not have

$$\frac{c_{sv_i}}{c_{\pi(s,v_{i+1})}} = \frac{1}{c_{\pi(v_i,v_{i+1})}}. \tag{5.36}$$

If $c_{\pi(s,v_{i+1})} = c_{sv_{i+1}} + c_{sv_i}c_{\pi(v_i,v_{i+1})} + x$ for some $x > 0$ which is a sum of path products then (5.36) means that we have $c_{sv_{i+1}} = -x < 0$, which is impossible given the parameter space $\mathring{\Theta}_\Sigma$. This means that the product $M_{31}\prod_{i=1}^{n-1} M_{v_{i+1},v_i}$ has at least two different positive values - one involving $\frac{c_{sv_i}}{c_{\pi(s,v_{i+1})}}$ and another $1/c_{\pi(v_i,v_{i+1})}$. However A_{u1} has only one possible positive value. This completes the proof that (iii) implies (i). □

5.8.3 Proofs for section 5.6

Lemma 5.8.4 (Identifiability on a tournament)**.** *Let X be a sum-linear model as in Definition 5.2.1 with respect to a single transitive tournament, i.e., $\mathcal{T} = \tau = (V, E)$. If a node $v \in \bar{U}$ has at least one parent then we can find two different $\theta, \theta' \in \mathring{\Theta}_\Sigma$ for which $H_{\theta,U} = H_{\theta',U}$.*

Proof. The proof is the same as for Lemma 4.6.3.4, because the map $f : \mathring{\Theta}_\Sigma \to \mathbb{R}^{|V|\times|V|}$ given by $b = f(\theta)$ is continuous and by Lemma 5.6.1 is injective too. □

Proof for Proposition 5.6.1

Proof. First we prove the sufficiency (if) part. We assume (I1) and (I2). We follow the same steps as in the proof for the max-linear model: the first one is to show that $H_{\theta,U}$ has $|V|$ distinct atoms $\{\omega_r\}$ that can be uniquely matched to the atoms $\{a_i\}_{i\in V}$. The second step is to show how to extract the parameter vector $\theta = (c_e, e \in E)$ from the atoms $\{a_i\}_{i\in V}$ and their masses $\{m_i\}_{i\in V}$.

Step 1. A mass $m_{i,U}$ can equal zero if $\mathrm{Desc}(i) \cap U = \emptyset$. The last expression means that all descendants of and including i are unobserved. But since the graph is finite, i must have at least one descendant without any children. However, a node in $\bar{U}$ without children is excluded by (I1). This means that all $|V|$ atoms are well defined and have positive masses.

Next we look at the possibility to have two atoms $a_{i,U}$ and $a_{j,U}$ with the same zero coordinates, i.e. $\{v \in U : b_{vi} > 0\} = \{v \in U : b_{vj} > 0\}$. This happens when for $i, j \in V, i \neq j$ it holds $\mathrm{Desc}(i) \cap U = \mathrm{Desc}(j) \cap U$. By Lemma 4.6.3.2-5 there cannot be any other node $k \in V \setminus \{i, j\}$ with the same descendants in U. In the representation

$$H_{\theta,U} = \sum_{v\in V} m_{v,U}\delta_{a_{v,U}} = \sum_{r=1}^{s} \mu_r \delta_{\omega_r}$$

there are thus two exactly two atoms, ω and ω', say, with the same indices of non-zero coordinates as $a_{i,U} = (b_{vi}/m_{i,U})_{v\in V}$ and $a_{j,U} = (b_{vj}/m_{j,U})_{v\in V}$. The question is then how to know whether $\omega = a_{i,U}$ and $\omega' = a_{j,U}$ or vice versa, $\omega = a_{j,U}$ and $\omega' = a_{i,U}$. Let μ and μ' be the masses of ω and ω', respectively, and consider the vectors $\beta = \mu\omega$ and $\beta' = \mu'\omega'$. An equivalent question is then how to match β and β' to the two max-linear coefficient vectors $(b_{vi})_{v\in U}$ and $(b_{vj})_{v\in U}$.

We have nodes i, j such that $\mathrm{Desc}(i) \cap U = \mathrm{Desc}(j) \cap U$. By Lemma 4.6.3.2-2 we have established that in this case there is a node u such that the triple of nodes $\{i, j, u\}$ forms a triangle with edge set $\{(i, j), (i, u), (j, u)\} \subset E$ or $\{(j, i), (i, u), (j, u)\} \subset E$. We assume it is $\{(i, j), (i, u), (j, u)\} \subset E$. We have to choose the node u such that $\{i, j\}$ are the only parents of u. In this way the expression of b_{ui} is determined of only two paths $\{i, u\}$ and $\{i, j, u\}$. Hence choose u such that $\mathrm{pa}(u) = \{i, j\}$. This is possible because u can not have parents from other tournaments than the one determined by the triangle i, j, u, say τ. Also take u such that j has one out-degree more than u.

We know that the vector which is associated to node i equals $b_{vi} = b_{ii}c_{\pi(i,v)}$ for all $v \in U$. The vector associated to node j equals $b_{vj} = b_{jj}c_{\pi(j,v)}$ for all $v \in U$. If $j, u \in U$ we have for $v = u$

$$b_{ui} = (c_{iu} + c_{ij}c_{ju})b_{ii}, \qquad b_{uj} = c_{ju}b_{jj},$$

and for $v = j$

$$b_{ji} = c_{ij}b_{ii}, \qquad b_{jj} = b_{jj}.$$

We should have $c_{iu} > 0$ hence

$$c_{iu} = \frac{b_{ui}}{b_{ii}} - c_{ij}c_{ju} = \frac{b_{ui}}{b_{ii}} - \frac{b_{ji}}{b_{ii}}\frac{b_{uj}}{b_{jj}} > 0 \Longrightarrow \frac{b_{ui}}{b_{ji}} > \frac{b_{uj}}{b_{jj}}.$$

Note that $b_{ii} = c_{ii} > 0$ hence we don't change the sign of the inequality.

In order to make the correct assignment of two vectors $\beta = (\beta_v)_{v\in U}$ and $\beta' = (\beta'_v)_{v\in U}$ we need to check the inequality above. If we have $\beta_u/\beta_j > \beta'_u/\beta'_j$ we assign the vector β to node i and β' to node j. If we have $\beta'_u/\beta'_j > \beta_u/\beta_j$ we assign β' to node i and β to node j.

When $j, u \in \bar{U}$ we need to use by Lemma 4.6.3.1 the unique paths from j, u to some nodes $j', u' \in U$ respectively. Because these are unique paths we have $c_{\pi(j,j')} = c_{p(j,j')}$ where $p(j, j')$ is the set of edges along the unique shortest path between nodes j, j', and $c_{p(j,j')} = \prod_{e\in p(j,j')} c_e$ is as in the chapter on max-linear models. For $v = u'$

$$b_{u'i} = (c_{iu} + c_{ij}c_{ju})c_{p(u,u')}b_{ii}, \qquad b_{u'j} = c_{ju}c_{p(u,u')}b_{jj},$$

and for $v = j'$

$$b_{j'i} = c_{ij}c_{p(j,j')}b_{ii}, \qquad b_{j'j} = c_{p(j,j')}b_{jj}.$$

Then for $c_{iu} > 0$ we should have the inequality $b_{u'i}/b_{j'i} > b_{u'j}/b_{j'j}$. We make the assignment as above. This shows that each of the collections $\{\omega_r\}$ and $\{a_{i,U}\}$ has $|V|$ distinct atoms and each element in $\{\omega_r\}$ can be matched uniquely to an element of $\{a_{i,U}\}$.

When $j \in U, u \in \bar{U}$ or $j \in \bar{U}, u \in U$ we combine the two methods described above.

Step 2. When there are latent variables, and conditions (I1) and (I2) are satisfied, the available coefficients, as shown in step 1 above, are $\{b_{ij}\}_{i\in U, j\in V}$.

Let $v \in \bar{U}$ and let $K = \{k_i\}$ be a (sub)set of children of v within a given tournament, τ, such that v has one out-degree more than k_1, the node k_1 has one out-degree more than k_2 and so on. We will show that we can identify all the edge weights on the sub-tournament induced by the node set $K \cup \{v\}$, say, $\mathcal{K}$. Note that $\mathcal{K}$ is a transitive tournament itself where v is the source node and it satisfies Harary and Moser (1966, Corollary 5a).

We will proceed by induction on the number of nodes in K.

Case 2.A) First we will assume $K \in U$. The initial step is to identify all parameters in $\mathcal{K}$ if $K = \{i, j\}$, i.e. $|K| = 2$. So i, j are children of v in $\mathcal{K}$ and they are selected such that (up to interchanging i and j) v has one out-degree more than i and i one degree more than j. This means that we have $(v, i), (v, j), (i, j) \in E$ and that $\pi(v, i) = \{(v, i)\}$ and $\pi(i, j) = \{(i, j)\}$.

If v has no parents in τ we have $b_{ii} = 1 - c_{vi}$ and we hane $c_{vi} = 1 - b_{ii}$. If v has parents in τ then it must take part in at least one more tournament according to (I2). Consider the node u in one of these tournaments such that $\pi(v, u) = \{(v, u)\}$. If $u \in U$ we have $b_{uu} = 1 - c_{vu}$, then $c_{vu} = 1 - b_{uu}$ and then from $b_{uv} = c_{vu}b_{vv}$ we obtain $b_{vv} = b_{uv}/(1-b_{uu})$.If $u \in \bar{U}$ then there exists a node $u' \in U$ which according to Lemma 4.6.3.1 has a unique directed path from u, say $\{u = u_1, \ldots, u_n = u'\}$. By the same lemma, every node in $\{u_2, \ldots, u_n = u'\}$ has a unique parent, which is the previous node in the sequence, so by Lemma 5.3.2 we have $b_{u_r,u_r} = 1 - c_{u_{r-1},u_r}$. Therefore we can obtain all the parameters $c_{u_{r-1},u_r}, r = n, \ldots, 2$ using the expressions $b_{u',u_r} = c_{p(u_r,u')}b_{u_r,u_r} = c_{p(u_r,u')}(1-c_{u_{r-1},u_r}), r = n, \ldots, 2$. Next, from $b_{u_n,u_1} = c_{p(u_1,u_n)}b_{u_1,u_1} = c_{p(u_1,u_n)}(1-c_{v,u})$ we obtain the value of $c_{v,u}$ and from $b_{u_n,v} = c_{p(v,u_n)}b_{vv}$ we obtain the value of b_{vv}. Once we have b_{vv} from $b_{iv} = c_{vi}b_{vv}$ we compute the value of c_{vi}.

Remember that the scope is to be able to compute all weights on edges (v, i), (v, j), and (i, j). Next, take $b_{ji} = c_{ij}b_{ii}$ from which we obtain c_{ij} because $i \in U$ means that

we know b_{ii}. Then $b_{jv} = (c_{vj} + c_{vi}c_{ij})b_{vv}$ because there are only two directed paths from v to j, one direct and one passing through node i. Given that we identified c_{vi}, c_{ij} and b_{vv} we obtain tha value c_{vj} from b_{jv}. This completes the initial step of the induction.

The induction assumption is that for $|K| = n - 1$, in the tournament induced by the nodes $K \cup v$, all edge parameters have been obtained.

So let's add one more node, say z, to the tournament induced by the $n-1$ children of v and v itself. We link z to the nodes in $K \cup v$ by edges $(k, z), k \in K \cup v$. We assume $z \in U$. The graph induced by $K \cup v \cup z$ is a tournament satisfying Harary and Moser (1966, Corollary 5a).

Let the node y be such that $\pi(y, z) = \{(y, z)\}$. Then because $y \in K \subset U$ we know the value of b_{yy}, and $b_{zy} = c_{yz}b_{yy}$ and we obtain the value c_{yz}.

Then consider the node, say, x which is such that $\pi(x, y) = \{(x, y)\}$. From node x there are two paths to z, one (x, z) and the other $\{(x, y), (y, z)\}$. Hence $b_{zx} = (c_{xz} + c_{xy}c_{yz})b_{xx}$ which allows us to identify c_{xz}, because we have found c_{yz}, the coefficient c_{xy} is obtained by the induction hypothesis, and b_{xx} is available because $x \in K \subset U$ by assumption.

Next consider node $w \in K$ such that $\pi(w, x) = \{(w, x)\}$. The directed paths from w to z are (w, z) and those going through nodes $\{w, x, z\}$, $\{w, y, z\}$, and $\{w, x, y, z\}$. Hence we have $b_{zw} = (c_{wz} + c_{wx}c_{xz} + c_{wy}c_{yz} + c_{wx}c_{xy}c_{yz})b_{ww}$. The parameters c_{xy}, c_{wx}, c_{wy} are available because of the induction assumption and the parameters c_{xz}, c_{yz} have just been computed. The coefficient b_{ww} is available because of the assumption $w \in K \subset U$. Then we can obtain the value c_{wz}. In this way we continue to identify each of the remaining edge coefficients $(i, z), i \in K \setminus \{w, x, y\}$ until we identify c_{vz}. Note that to identify b_{vv} we proceed as for the initial step of the induction.

Case 2.B) Now assume $K \cup z \subset \bar{U}$. The difference with respect to the previous case is that now the coefficients $b_{ii}, i \in K$ are not available. But to identify these we use the unique paths that the nodes in $\bar{U}$ have to nodes in U according to Lemma 4.6.3.1. Details follow.

First we will show the initial induction step, i.e., when $K = \{i_1, j_1\}$. We take i_1 such that v has one out-degree more than i_1, node j_1 has one out-degree more than i_1. This means that we have $(v, i_1), (v, j_1), (i_1, j_1) \in E$ and that $\pi(v, i_1) = \{(v, i_1)\}$ and $\pi(i_1, j_1) = \{(i_1, j_1)\}$.

By Lemma 4.6.3.1 we can find nodes $i_n, j_m \in U$ such that the paths $(i_1, \ldots, i_n)$ and $(j_1, \ldots, j_m)$ satisfy the properties therein. Then we can compute all coefficients $(c_{i_{r-1},i_r}, r = n, \ldots, 2)$ from $b_{i_n,i_r} = c_{p(i_r,i_n)}b_{i_r,i_r} = c_{p(i_r,i_n)}(1 - c_{i_{r-1},i_r})$, which means that we obtain the values of all coefficients on the path product $c_{p(i_1,i_n)}$. From $b_{i_n,i_1} = c_{p(i_1,i_n)}b_{i_1,i_1}$ we obtain the value b_{i_1,i_1}. Note that if v has no parents in τ, then i_1 has no other parents than v and then $b_{i_1,i_1} = 1 - c_{vi_1}$. This means that we compute $c_{vi_1} = 1 - b_{i_1,i_1}$. If v has parents in τ we make the same reasoning as in the case 2.A) above which shows that we are able to compute the value b_{vv}. Then because $\pi(v, i_1) = \{(v, i_1)\}$ and $\pi(i_1, i_n) = \{p(i_1, i_n)\}$ we have $b_{i_n,v} = c_{vi_1}c_{p(i_1,i_n)}b_{vv}$ and from here we compute the value of c_{vi_1}.

Next we compute the parameters $(c_{j_{r-1},j_r}, r = m, \ldots, 2)$ from

$$b_{j_m,j_r} = c_{p(j_r,j_m)}b_{j_r,j_r} = c_{p(j_r,j_m)}(1 - c_{j_{r-1},j_r}).$$

Because for the path sets we have $\pi(i_1, j_1) = \{(i_1, j_1)\}$ and $\pi(j_1, j_m) = \{p(j_1, j_m)\}$, and the fact that j_2 cannot have other parent than j_1 according to Lemma 4.6.3.1 in the previous chapter we have $b_{j_m,i_1} = c_{\pi(i_1,j_m)}b_{i_1,i_1} = c_{i_1,j_1}c_{p(j_1,j_m)}b_{i_1,i_1}$ and from here we obtain the value c_{i_1,j_1}. Note that from the previous paragraph we have obtained b_{i_1,i_1}. Then because there are two paths from v to j_1, we have $b_{j_m,v} = (c_{vj_1} + c_{vi_1}c_{i_1,j_1})c_{p(j_1,j_m)}b_{vv}$ and the only unknown parameter is c_{v,j_1}, hence we are

able to express it in terms of the other known parameters in the equation. In this way we showed how to compute the values of $c_{vi_1}, c_{vj_1}, c_{i_1,j_1}$. This completes the initial induction step.

The induction assumption is that for $|K| = n - 1$, in the tournament induced by the nodes $K \cup v$, all edge parameters are computed.

Again, we add a node $z_1 \in \bar{U}$ to K such that the graph $\mathcal{K}$ contains n nodes, those in K and z_1. We connect z_1 to all other nodes by an arc toward z_1. We have to show that we can identify $c_{az_1}, a \in K \cup v$. By Lemma 4.6.3.1 we can find node $z_n \in U$ such that the path $(z_1, \ldots, z_n)$ satisfy the properties therein. We compute the parameters $(c_{z_{r-1},z_r}, r = n, \ldots, 2)$ from $b_{z_n,z_r} = c_{p(z_r,z_n)} b_{z_r,z_r} = c_{p(z_r,z_n)}(1 - c_{z_{r-1},z_r})$ which means that we know the product $c_{p(z_1,z_n)}$. Then from $b_{z_n,z_1} = c_{p(z_1,z_n)} b_{z_1,z_1}$ we obtain the coefficient b_{z_1,z_1}.

Let $y_1, x_1, w_1 \in K \subset \bar{U}$ be children of v such that $\pi(y_1, z_1) = \{(y_1, z_1)\}$, $\pi(x_1, y_1) = \{(x_1, y_1)\}$ and $\pi(w_1, y_1) = \{(w_1, y_1)\}$. For each of the nodes y_1, x_1, w_1 we have by Lemma 4.6.3.1 nodes $y_q, x_m, w_l \in U$ such that there are unique paths $p(y_1, y_q)$, $p(x_1, x_m)$, and $p(w_1, w_l)$. Similarly as for the unique path $p(z_1, z_n)$ we are able to compute the path products $c_{p(y_1,y_q)}$, $c_{p(x_1,x_m)}$, $c_{p(w_1,w_l)}$ and the coefficients b_{y_1,y_1}, b_{x_1,x_1}, b_{w_1,w_1}.

Because there is a unique path from y_1 to z_n and it passes through z_1 we have $b_{z_n,y_1} = c_{y_1,z_1} c_{p(z_1,z_n)} b_{y_1,y_1}$ where the only unknown is c_{y_1,z_1}, so we are able to obtain its value from this equation. Because there are two paths from x_1 to z_1, one direct and one going through y_1, we have $b_{z_n,x_1} = (c_{x_1,z_1} + c_{x_1,y_1} c_{y_1,z_1}) c_{p(z_1,z_n)} b_{x_1,x_1}$ where c_{x_1,y_1} is known by the induction hypothesis and $c_{y_1,z_1}, c_{p(z_1,z_n)}, b_{x_1,x_1}$ have been found previously. Then we obtain the value of c_{x_1,z_1}. We continue in the same way backwards until we identify c_{vz_1}. Note that to identify b_{vv} we can proceed as in the initial step of the induction. Below we show this once again for completeness.

For the last coefficient we need the value of b_{vv}. If v has no parents in τ consider the following. Let $i_1 \in \bar{U}$ be its child such that $\mathrm{pa}(i_1) = \{v\}$. Because $i_1 \in \bar{U}$, we have according to Lemma 4.6.3.1 a node $i_r \in U$ with $\{p(i_1, i_r)\} = \pi(i_1, i_r)$. We can identify all edge weights $(c_{i_1 i_2}, c_{i_2 i_3}, \ldots, c_{i_{r-1}, i_r})$ from $b_{i_r i_2}, \ldots, b_{i_r i_r}$. Note that $b_{i_1 i_1} = 1 - c_{vi_1}$ hence from $b_{i_r i_1} = c_{p(i_1,i_r)} b_{i_1 i_1}$ we obtain the value of c_{vi_1}. From $b_{i_r v} = c_{vi_1} c_{p(i_1,i_r)} b_{vv}$ we obtain the value of b_{vv}.

If v has parents in τ we have that v must satisfy (I2), hence participate in at least one another tournament, τ. There must exist a node $v' \in U$ such that $\pi(v, v') = \{p(v, v')\}$ according to Lemma 4.6.3.1. We can identify all edge weights on the path $p(v, v')$ and then from $b_{v'v} = c_{p(v,v')} b_{vv}$ we obtain b_{vv}.

The principle is illustrated in Fig. 5.1.

For the case when any subset of K is unobserved we can combine the methods in case 2.A) and in case 2.B).

Next we prove necessity (only if). We will show that when at least one of the two conditions in (I1) and (I2) are not satisfied, there are two parameters $\theta \neq \theta' \in \mathring{\Theta}_\Sigma$ such that $H_{\theta,U} = H_{\theta',U}$. We will look at two cases: (1) condition (I2) is not satisfied; (2) condition (I1) is not satisfied and condition (I2) is satisfied.

Case (1). The proof is analogous to the one for the max-linear model. Let $v \in \bar{U}$ be such that condition (I2) is not satisfied, i.e., there is a $v \in \bar{U}$ which is not a source of any tournament. Then v belongs to only one tournament, say τ. We will proceed by induction on the number of tournaments. For a single tournament, i.e., τ, the lack of identifiability follows directly from Lemma 5.8.4. Let $\mathcal{T}$ have m tournaments. In the corresponding proof for the max-linear model it was argued that we can find an order on the tournaments, $\tau_1, \tau_2, \ldots, \tau_m$ such that $V_m \cap \{V_1, \ldots, V_{m-1}\}$ is a singleton, say s. Let τ be one of $\tau_1, \tau_2, \ldots, \tau_{m-1}$. Note that $v \neq s$ as v should belong to a single tournament. Then by the same argument as in the proof of Proposition 4.4.2 we argue

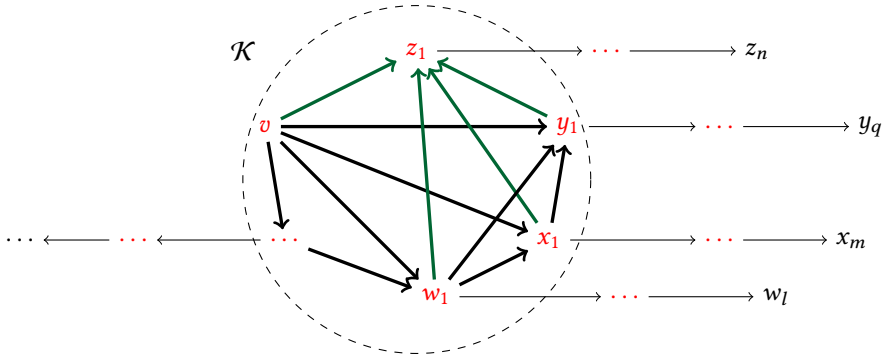


Figure 5.1: Illustration of the proof method: this is the subgraph induced by the nodes $K \cup v \cup z_1$, $\mathcal{K}$, which is itself a transitive tournament. All the nodes in this graph could be unobserved, therefore colored in red. The parameters on the thick black edges are available by the induction assumption. The nodes in black carry observable variables. The nodes in red carry non-observable. The scope is to identify the coefficients on the thick green edges, which all point to the node z_1. We first compute the value of c_{y_1,z_1} as explained in the proof. Next, we want to compute the value of c_{x_1,z_1}. The non-direct path from x_1 to z_1 given by $c_{x_1,y_1}c_{y_1,z_1}$ is available, because c_{x_1,y_1} is available from the induction assumption and c_{y_1,z_1} has just been computed. Then we can identify the direct one, i.e., c_{x_1,z_1} as explained in the proof. After that, all non-direct paths from w_1 to z_1 are available and we can identify the direct path c_{w_1,z_1}. In this way working backwards to the node v until we identify c_{v,z_1} which completes the proof.

that the distribution of $X_{V\setminus v}$ as a sum-linear model in Definition 5.2.1 is identical under two different parameters $\theta, \theta' \in \mathring{\Theta}_\Sigma$. We only need to use the markovianity of X from Proposition 5.4.1.

Case (2). Let $v \in \bar{U}$ be such that condition (I1) is not satisfied and condition (I2) is satisfied. Because v has to be a source in at least one tournament, but at the same time it should have less than two children, then v has actually one child, say w, and the tournament composed of nodes $\{v, w\}$ and edges (v, w) is the one in which v is source.

If v has parents, these must belong to only one tournament, say $\tau = (V_\tau, E_\tau)$ is the tournament that v shares with its parents.

Note that v is the sink (node without children) in τ. This means that edge weights $c_{iv}, i \in \mathrm{pa}(v)$ are contained only in coefficient products on directed paths passing through or ending at v. These paths are from $k \in \mathrm{An}(v)$ to $l \in \mathrm{Desc}(v)$. The coefficients in the max-linear model X that depend on these path products are accordingly

$$b_{lk}, \qquad k \in \mathrm{An}(v), l \in \mathrm{Desc}(v). \tag{5.37}$$

Similar reasoning holds for w: because w is a sink in the tournament shared with v, only the coefficients

$$b_{lk}, \qquad k \in \mathrm{An}(w), l \in \mathrm{Desc}(w) \tag{5.38}$$

depend on c_{vw}.

When $v \in \bar{U}$ from the definition of $H_{\theta,U}$ in (5.26) or also (4.20) and (4.21), the coefficients $b_{vi}, i \in \mathrm{An}(v)$ are no longer available. Adjusting the set of coefficients in (5.37) and (5.38), by removing $b_{vi}, i \in \mathrm{An}(v)$, gives

$$\begin{aligned} &b_{lk}, \qquad k \in \mathrm{An}(v), l \in \mathrm{desc}(v) \\ &b_{lk}, \qquad k \in \mathrm{An}(w), l \in \mathrm{Desc}(w). \end{aligned} \tag{5.39}$$

Besides this, the atoms $a_{v,U} \propto (b_{iv}, i \in \text{Desc}(v) \cap U)$ and $a_{w,U} \propto (b_{iw}, i \in \text{Desc}(w) \cap U)$ are equal, hence the measure $H_{\theta,U}$ has an atom, say ω, which is equal to $a_{v,U} = a_{w,U}$. We explain this now. We have $\text{Desc}(v) \cap U = \text{Desc}(w) \cap U$ and for every $i \in \text{Desc}(w) \cap U$ we have $b_{iv} = c_{\pi(v,i)} b_{vv} = c_{vw} c_{\pi(w,i)} b_{vv}$ and $b_{iw} = c_{\pi(w,i)} b_{ww} = c_{\pi(w,i)}(1 - c_{vw})$. Hence the two vectors are proportional

$$(b_{iv}, i \in \text{Desc}(w) \cap U) = \frac{b_{vv} c_{vw}}{1 - c_{vw}} (b_{iw}, i \in \text{Desc}(w) \cap U).$$

This means that the two atoms $a_{w,U}$, because of their definition as in in (5.12), are equal

$$a_{v,U} = a_{w,U} = \left(\frac{c_{\pi(w,i)}}{\sum_{i \in \text{Desc}(w) \cap U} c_{\pi(w,i)}}, i \in \text{Desc}(w) \cap U \right), \tag{5.40}$$

and as we see its elements do not depend on c_{vw}. Hence $H_{\theta,U}$ contains an atom $\omega = a_{v,U} = a_{w,U}$ given by (5.40). The mass of it, say μ, is given by

$$m_{v,U} + m_{w,U} = (c_{vw} b_{vv} + 1 - c_{vw}) \sum_{i \in \text{Desc}(w) \cap U} c_{\pi(w,i)}. \tag{5.41}$$

Because of the atom ω given by (5.40) the coefficients $b_{\cdot v}$ and $b_{\cdot w}$ are not anymore available in $H_{\theta,U}$. Adjusting the set of coefficients in (5.39)

$$\begin{aligned} &b_{lk}, \qquad k \in \text{an}(v), \quad l \in \text{desc}(v) \\ &b_{lk}, \qquad k \in \text{an}(w) \setminus v, \quad l \in \text{Desc}(w) \end{aligned} \tag{5.42}$$

Note that $\text{an}(v) = \text{an}(w) \setminus v$ and $\text{desc}(v) = \text{Desc}(w)$. Therefore these coefficients in $H_{\theta,U}$ that depend on at least one of the edge weights $c_{iv}, i \in \text{pa}(v)$ depend on c_{vw} too and they are

$$b_{lk}, \qquad k \in \text{an}(v), \quad l \in \text{Desc}(w). \tag{5.43}$$

If $\text{an}(v) = \emptyset$, then $\text{pa}(v) = \emptyset$ too and from (5.43) there are no coefficients b_{lk} that depend on c_{vw} since $k \in \text{an}(v) = \emptyset$ and hence c_{vw} can take arbitrary value in $(0, 1)$ without changing $H_{\theta,U}$. If $\text{an}(v)$ is non empty we will look at two cases: A) when $k \in \text{pa}(v)$ and B) when $k \in \text{an}(v) \setminus \text{pa}(v)$.

Let's first consider two parameters, $\theta = (c_e, e \in E) \in \mathring{\Theta}_\Sigma$ and θ', with

$$\theta' = \begin{cases} \lambda c_{iv}, & \text{for all } i \in \text{pa}(v), \\ \lambda^{-1} c_{vw}, & \\ c_e, & e \in E \setminus \{(v,w), (i,v), i \in \text{pa}(v)\}. \end{cases} \tag{5.44}$$

for some λ such that $\theta' \in \mathring{\Theta}_\Sigma$ too. Since $\mathring{\Theta}_\Sigma$ is open space we can always find such λ. Then we have coefficients b'_{lk} for $k \in \text{an}(v), l \in \text{Desc}(w)$ which depend on at least one of the parameters $\lambda c_{iv}, i \in \text{pa}(v)$ and on $\lambda^{-1} c_{vw}$.

Case 2.A) If $k \in \text{pa}(v)$ we have the following expression for the sum of all product paths

$$c_{\pi(k,v)} = \sum_{i \in \text{pa}(v) \cap \text{Ch}(k)} c_{\pi(k,i)} c_{iv}. \tag{5.45}$$

Then

$$\begin{aligned} b'_{lk} &= c_{\pi(k,l)} b'_{kk} = c_{\pi(k,v)} \lambda^{-1} c_{vw} c_{\pi(w,l)} b'_{kk} \\ &= \left(\sum_{i \in \text{pa}(v) \cap \text{Ch}(k)} c_{\pi(k,i)} \lambda c_{iv} \right) \lambda^{-1} c_{vw} c_{\pi(w,l)} b_{kk} \\ &= c_{\pi(k,v)} c_{vw} c_{\pi(w,l)} b_{kk} = b_{lk}. \end{aligned}$$

Note that $b'_{kk} = b_{kk} = 1 - \sum_{j \in \mathrm{pa}(k)} c_{jk}$ by (5.44) and Lemma 4.6.3.3.

Case 2.B) If k is not a parent of v but another ancestor, then we can write $c_{\pi(k,v)} = c_{\pi(k,s)} c_{\pi(s,v)}$ where s is the source of τ (the tournament v shares with its parents). Then we have a similar expression for $c_{\pi(s,v)}$ as in (5.45). Hence

$$\begin{aligned} b'_{lk} &= c_{\pi(k,s)} \left(\sum_{i \in \mathrm{pa}(v) \cap \mathrm{Ch}(s)} c_{\pi(s,i)} \lambda c_{iv} \right) \lambda^{-1} c_{vw} c_{\pi(w,l)} b'_{kk} \\ &= c_{\pi(k,s)} c_{\pi(s,v)} c_{vw} c_{\pi(w,l)} b_{kk} = b_{lk}. \end{aligned}$$

In both cases two different parameters θ, θ' generate the same distribution $H_{\theta,U} = H_{\theta',U}$.

□

Estimation and analysis with gremes

6

The package has been published on CRAN in December 2021.

6.1 Introduction

The R package `gremes` provides tools for estimation of the tail dependence parameters in graphical models parameterized by a family of Hüsler–Reiss distributions. The only supported graphs are *trees* and *block graphs*. The estimation methods are variations of method of moments (Engelke, Malinowski, Kabluchko, and Schlather, 2014; Asenova, Mazo, and Segers, 2021), maximum likelihood (Asenova, Mazo, and Segers, 2021; Engelke and Hitz, 2020) and a method based on extremal coefficients (Einmahl, Kiriliouk, and Segers, 2018).

The package is provided with rich explanations and illustrations of the available functionalities, their use and relation to the statistical model. The accompanying vignettes are written with the scope to explain at a moderate level of detail the statistical model and the estimation methods, aiming to be self-contained and avoid referring the reader to the corresponding articles. The following vignettes make part of the literature available with the package.

- Vignette *Detailed contents* contains a detailed guide into the documentation of the package and a summary of the main functionalities. We recommend a potential user to have a look at this document first.
- Vignette *Hüsler–Reiss distributions* presents the particular parameterizations used in the Hüsler–Reiss distributions.
- Vignette *Subsets and coordinates* presents classes and methods used to create subsets on the vertex set. The choice of subsets can be based on the principle of a neighborhood around a particular node. Some of the coordinates are set to zero in the extremal coefficient estimator. The choice of subsets and non-zero coordinates may be a non-trivial task when there are latent variables. Details are not provided in this chapter.
- Vignettes *Estimation - Note* 1-6 present detailed description of the estimation methods.
- Vignettes *Code - Note* 1-6 illustrate the use of the estimation tools (the methods, classes and functions) of the package.
- Vignette *Additional functionalities* provides explanation and illustrates the functions related to additional functionalities such as generating random sample from a model, computing extremal coefficients, tail dependence coefficients, confidence intervals for one of the estimators.

The package is developed in an *object-oriented* style. The classes are `S3` classes. There are two main types of objects.

- An object containing the graph and the dataset is created using classes `Network`, `Tree`, `BlockGraph`, and subclasses of these.
- An object containing the graph and the edge weights is created with classes `HRMnetwork`, `HRMtree`, `HRMBG`, and subclasses of these.

The first type of objects represents the non-parametric view on the problem - all we know is the graph and the data. The second type of objects represents the Hüsler–Reiss parametric model: every clique is parameterized by a Hüsler–Reiss distribution with parameters - the edge weights within this clique. Hence all that characterizes the parametric model is the graph and the edge weights.

Consider for instance the method `extrCoeff` which is written both for classes `Tree` and `HRMtree`. If we pass an object of class `Tree` to the method `extrCoeff`, the command will return non-parametric estimates of the extremal coefficients. If the object passed is of class `HRMtree`, parametric extremal coefficients will be returned.

The main goal of the package is estimation, therefore the method `estimate` is the key functionality of the tools provided in the package. Estimation in `gremes` happens by using the method `estimate` on an object from one of the following classes:

- `MME`, `MLE`, `MLE1`, `MLE2`, `EKS`, `EKS_part`, `EngHitz`, `MMEave`, `MLEave` in which case it estimates the edge weights on a tree.
- `HRMBG` in which case it estimates the edge weights on a block graph.

We recall once again the models and the object of estimation.

6.2 Theoretical setting - once again

6.2.1 The model on trees

Let $T = (V, E)$ be a tree with node set V and edge set E. Consider a $|V|$-variate random vector $X = (X_v, v \in V)$ for which it holds: X satisfies the global Markov property with respect to the tree T; every bivariate distribution between two adjacent variables uses a bivariate Hüsler–Reiss copula with parameter θ_e for some $e \in E$; standardized to unit-Pareto univariate marginal distributions. The particular parameterization of the Hüsler–Reiss distribution is as in Chapter 2, i.e., for two adjacent nodes u, v and $z_u, z_v \in (0, \infty)^2$

$$H_{\theta_e}(z_u, z_v) = \exp\left\{-\frac{1}{z_u}\Phi\left(\frac{\theta_e}{2} + \frac{\ln z_v/z_u}{\theta_e}\right) - \frac{1}{z_v}\Phi\left(\frac{\theta_e}{2} + \frac{\ln z_u/z_v}{\theta_e}\right)\right\},$$

Based on different asymptotic results of X, we can have different estimators of the parameters $\theta_e, e \in E$, which we collect in a vector $\theta \in (0, \infty)^E$.

(L1) Consider the log-differences conditional on a high threshold exceeding at a particular node

$$(\ln X_v - \ln X_u, v \in V) \mid X_u > t, \qquad t \to \infty.$$

It can be shown that the limiting distribution of the vector above is multivariate Gaussian distribution with mean vector $\mu_u(\theta)$ and covariance matrix $\Sigma_u(\theta)$ which depend on the edge weights and on the particular node u. This is a result

from Chapter 2 or also Asenova, Mazo, and Segers (2021) and references therein. The MME and the CLE both aim at estimating Σ_u and accordingly θ. The estimators are implemented in methods `estimate.MME`, `estimate.MLE`, `estimate.MLE1` and `estimate.MLE2`.

(L2) Consider the limiting distribution of the scaled componentwise maxima if we dispose of a random sample of size n of X, i.e., $\{X_{v,i}\}_{i\in 1,\ldots,n; v\in V}$

$$\left(\frac{1}{n}\max_{i=1,\ldots,n} X_{v,i}, v \in V\right), \qquad n \to \infty.$$

The limit shown in Chapter 2 or also Asenova, Mazo, and Segers (2021) is a max-stable Hüsler–Reiss copula with unit Fréchet margins and with parameter matrix Λ given by

$$\big(\Lambda(\theta)\big)_{ij} = \lambda_{ij}^2(\theta) = \frac{1}{4}\sum_{e\in p(i,j)} \theta_e^2, \qquad i,j \in V,\ i \neq j, e \in E. \tag{6.1}$$

The notation $p(i,j)$ means the unique shortest path between i, j. The extremal coefficient estimator, introduced in Einmahl, Kiriliouk, and Segers (2018) and used in Asenova, Mazo, and Segers (2021) is based on bi- and tri-variate extremal coefficients derived from this max-stable distribution. The estimator is implemented in methods `estimate.EKS`, `estimate.EKS_part`.

(L3) Consider the scaled random vector, given that the maximum exceeds a high threshold:

$$(X_v, v \in V)/t \mid \max_{v\in V} X_v > t, \qquad t \to \infty.$$

The limit is a so called Hüsler-Reiss Pareto (Engelke and Hitz, 2020) distribution with the same matrix Λ as in (6.1). Proof of this limit is provided in Asenova and Segers (2021), see also Engelke and Hitz (2020). For Hüsler–Reiss Pareto distributions with respect to trees, Engelke and Hitz (2020) presents a cliquewise estimator which is has been implemented in `gremes` and called 'Engelke and Hitz' estimator. We have adapted it to make it suitable when there are latent variables. More details are presented in the next section. The estimator is implemented in method `estimate.EngHitz`.

(L4) Consider the differences with respect to the mean of the log-transformed variables

$$(\ln X_v - \overline{\ln X}, v \in V) \mid \overline{\ln X} > t, \qquad t \to \infty,$$

where $\overline{\ln X} = (1/|V|)\sum_{v\in V} \ln X_v$. The limit of this vector is also a multivariate Gaussian distribution with mean and covariance matrix, say $\bar{\Sigma}$, that contain the matrix $\Lambda(\theta)$ in (6.1). This asymptotic result is shown in an unpublished note Segers (2019) and details are not provided here. The package offers method of moments and maximum composite likelihood estimates, The estimators are implemented in methods `estimate.MMEave`, `estimate.MLEave`.

To illustrate the idea, recall the Seine network from Chapter 2 on seven nodes as in Fig. 6.1, which represent measurement locations in cities, $V = \{\text{Paris}, 2, \text{Meaux}, \text{Melun}, 5, \text{Nemours}, \text{Sens}\}$, and edge weights $(\theta_1, \ldots, \theta_6)$. Let (X_{Paris}, X_2) have bivariate Hüsler–Reiss copula with parameter θ_1. We make the analogous assumptions for all adjacent pair of variables. Then for X the four asymptotic results above hold. To estimate tail dependence measures in X, we need estimates of $\theta_1, \ldots, \theta_6$. We need to choose between one of the following methods - for method

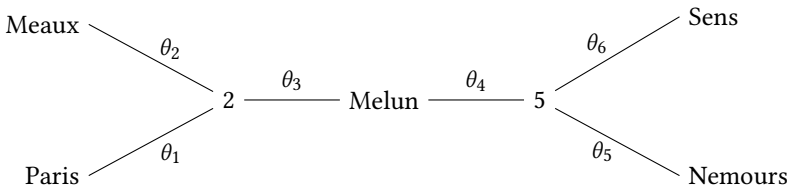


Figure 6.1: Schematic representation of Seine network on seven nodes.

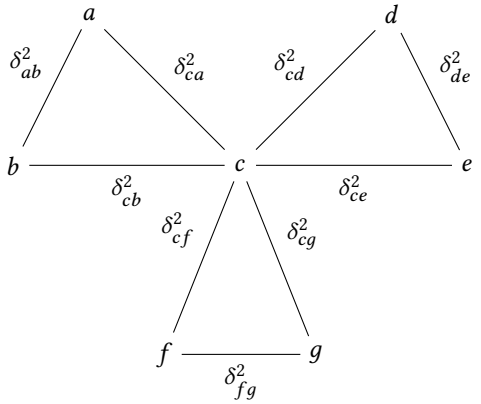


Figure 6.2: A block graph on three cliques: $\{a, b, c\}, \{c, d, e\}, \{c, f, g\}$.

of moments estimates `estimate.MME`; for maximum composite likelihood estimation `estimate.MLE`, `estimate.MLE1`, `estimate.MLE2`; for method based on extremal coefficients `estimate.EKS` and `estimate.EKS_part`; cliquewise estimation `estimate.EngHitz`; for estimation based on different conditioning event `estimate.MMEave` and `estimate.MLEave`.

6.2.2 The model on block graphs

Let $G = (V, E)$ be a block graph with node set V and edge set E. Let X be a random vector on V with the following characteristics: X satisfies the global Markov property with respect to G; every distribution of variables belonging to the same maximal clique (or block), say $C \subset V$, is a multivariate Hüsler–Reiss distribution with parameter matrix $\{\delta^2_{ij}\}, i, j \in C$; the univariate marginal distributions are unit Pareto.

Consider the log-differences conditional on a high threshold being exceeded at a particular node

$$(\ln X_v - \ln X_u, v \in V) \mid X_u > t, \qquad t \to \infty.$$

It can be shown that the limiting distribution is multivariate Gaussian distribution with mean vector $\mu_u(\delta)$ and covariance matrix $\Sigma_u(\delta)$ where $\delta = (\delta^2_e, e \in E)$ (Asenova and Segers, 2021) and references therein. Based on this asymptotic result we come up with method of moments type estimator for Σ_u and subsequently of δ.

To illustrate the idea, consider the following block graph in Fig. 6.2 with only three cliques (or blocks) and nine edge parameters: three per each clique. We suppose that the subvector (X_a, X_b, X_c) has a Hüsler–Reiss copula with parameter matrix Δ_1 which is symmetric, with zero diagonal and non-zero parameters $(\delta^2_{ab}, \delta^2_{ca}, \delta^2_{cb})$. Similarly for the other two subvectors (X_c, X_e, X_d) and (X_c, X_f, X_g). To estimate tail dependence in X we need to estimate the edge weights (or edge parameters), $\delta^2_{ij}, (i, j) \in E$. We apply the method `estimate.HRMBG`.

Similar results as in (L2), (L3), (L4) for trees exist for block graphs. However, the package does not implement an estimator based on these results.

6.3 Estimation in models on trees

6.3.1 Method of moment estimator

This method is based on the asymptotic results in (L1). The MME is described in detail in Section 4.1 in Asenova, Mazo, and Segers (2021). For completeness we repeat most of the notations here. The idea is to find $(\theta_e, e \in E)$ which minimizes the distance between the empirical and the theoretical covariance matrices:

$$\hat{\theta}^{\mathrm{MM}}_{n,k} = \underset{\theta\in(0,\infty)^E}{\arg\min} \sum_{u\in U} \|\hat{\Sigma}_{W_u,u} - \Sigma_{W_u,u}(\theta)\|_F^2,$$

where

- n is the number of all observations in the sample;
- k is the number of the upper order statistics used in the estimation;
- u is the node for which we condition on the event $\{X_u > t\}$;
- $\|\cdot\|_F$ is the Frobenius norm;
- $U \subseteq V$ is the set of observable variables;
- W_u is a subset on the node set depending on u. Typically a neighborhood of u or the nodes that are flow connected to u or the intersection of both. Note that the induced graph on W_u must be connected. *A good practice is to compose the sets such that within each subset all parameters are uniquely identifiable.* This means that every node in W with latent variable should be connected to at least three other nodes in the same set W.
- $\hat{\Sigma}_{W_u,u}$ is the non-parametric covariance matrix;
- $\Sigma_{W_u,u}(\theta)$ is the parametric covariance matrix;
- For fixed u and W_u the parametric matrix $\Sigma_{W_u,u}$ is given by

$$\big(\Sigma_{W,u}(\Lambda)\big)_{ij} = 2(\lambda^2_{iu} + \lambda^2_{ju} - \lambda^2_{ij}), \qquad i,j \in W \setminus u$$

  with

$$\big(\Lambda(\theta)\big)_{ij} = \lambda^2_{ij}(\theta) = \frac{1}{4} \sum_{e\in p(i,j)} \theta_e^2, \qquad i,j \in V,\ i \neq j, e \in E;$$

- If the sample of the original variables is $\xi_{v,i}, v \in U, i = 1,\dots,n$ consider the transformation using the empirical cumulative distribution function $\hat{F}_{v,n}(x) = \Big[\sum_{i=1}^n \mathbb{1}(\xi_{v,i} \le x)\Big]/(n+1)$

$$\hat{X}_{v,i} = \frac{1}{1 - \hat{F}_{v,n}(\xi_{v,i})}, \qquad v \in U, \quad i = 1,\dots,n;$$

- Fix u and W_u. For given $k \in \{1,\dots n\}$ consider the set of indices

$$I_u = \{i = 1,\dots,n : \hat{X}_{u,i} > n/k\}.$$

- For every $v \in W_u \setminus u$ and $i \in I_u$ compose the differences

$$\Delta_{uv,i} = \ln \hat{X}_{v,i} - \ln \hat{X}_{u,i};$$

- The vector of means of these differences is given by

$$\hat{\mu}_{W_u,u} = \frac{1}{|I_u|} \sum_{i\in I_u} (\Delta_{uv,i}, v \in W_u \setminus u);$$

- The non-parametric covariance matrix $\hat{\Sigma}_{W_u,u}$ is given by

$$\hat{\Sigma}_{W_u,u} = \frac{1}{|I_u|}\sum_{i\in I_u}(\Delta_{uv,i} - \hat{\mu}_{W_u,u}, v \in W_u \setminus u)(\Delta_{uv,i} - \hat{\mu}_{W_u,u}, v \in W_u \setminus u)^\top .$$

We will illustrate the method on Seine dataset, which is also part of the package. Detailed description of the content of the dataset is provided in the supplement section of Chapter 2. We load the dataset from the package, generate the graph and name the nodes. Assigning names to nodes is crucial. The names of the nodes should correspond to the names of the columns in the dataset.

```
data("SeineData", package = "gremes") # loads object Seine as
    database
seg<- graph(c(1,2, 2,3, 2,4, 4,5, 5,6, 5,7), directed = FALSE)
name_stat<- c("Paris", "2", "Meaux", "Melun", "5", "Nemours",
    "Sens")
seg<- set.vertex.attribute(seg, "name", V(seg), name_stat)
```

In the first step, subsets are used for local estimation of the edge weights. In the second step, the estimates are combined through a minimal distance procedure to obtain unique estimates.

```
subs<- Neighborhood()
subs<- subset(subs, 2, seg, U_bar=c("2", "5"))
```

Estimate using method of moment estimator and 20% of the upper order statistics.

```
mme<- MME(seg)
mme<- suppressMessages(
            estimate(mme, Seine, subs, k_ratio = 0.2)
)
```

There are messages which have been suppressed. They inform about certain things along the estimation process, but as long as they do not stop the estimation they are not errors. The estimates are squares of the parameters, hence take the square root.

```
sqrt(mme$depParams)
#>       e1       e2       e3       e4       e5       e6
#> 0.3833844 0.9491690 0.5914040 0.6287704 1.0034604 0.6162232
```

6.3.2 Composite likelihood estimation

The methods here are also based on the asymptotic result in (L1). We have three versions of the composite likelihood method. Consider the same notation as in the previous subsection.

Composite likelihood method - Version 1 The estimator of $(\theta_e, e \in E)$ is obtained in a two-step procedure:
Step 1: For each $u \in U$ obtain

$$\hat{\theta}_{W_u,k,n} = \underset{\theta_{W_u}\in(0,\infty)^{W_u\setminus u}}{\arg\max}\; L\Big(\mu_{W_u\setminus u}(\theta), \Sigma_{W_u\setminus u}(\theta); \{\Delta_{uv,i} : v \in W_u \setminus u, i \in I_u\}\Big).$$

Step 2: Then solve

$$\hat{\theta}^{MLE1}_{k,n} = \underset{\theta\in[0,\infty)^E}{\arg\min}\sum_{u\in U}\sum_{e\in E}(\hat{\theta}_{e,W_u} - \theta_e)^2,$$

where $\hat{\theta}_{e,W_u}$ is the estimate of θ_e from the vector of estimates $\hat{\theta}_{W_u}$ found in step 1. The advantage of the first version is that the computational time is shorter with respect to the second version.

Composite likelihood method - Version 2 Consider the likelihood function of a random sample $y_i, i = 1, \ldots, k$ of multivariate normal distribution with mean vector μ and covariance matrix Σ, where y_i is of dimension d

$$\begin{aligned} L(\mu, \Sigma; y_1, \ldots, y_k) &= \prod_{i=1}^{k} \phi_d(y_i - \mu; \Sigma) \\ &= (2\pi)^{-kd/2} (\det \Sigma^{-1})^{k/2} \exp\Big(-\frac{1}{2} \sum_{i=1}^{k} (y_i - \mu)^T \Sigma^{-1} (y_i - \mu) \Big). \end{aligned}$$

The method of composite likelihoods consists of optimizing a function that collects the likelihood functions across all the sets $W_u, u \in U$. So let for all $u \in U$ the subsets W_u be given. Consider the composite likelihood function

$$\begin{aligned} &L\big(\theta; \{\Delta_{uv,i} : v \in W_u \setminus u,\ i \in I_u, u \in U\}\big) \\ &\quad = \prod_{u \in U} L\big(\theta_{W_u}; \{\Delta_{uv,i} : v \in W_u \setminus u, i \in I_u\}\big) \\ &\quad = \prod_{u \in U} \prod_{i \in I_u} \phi\Big(\{\Delta_{uv,i} : v \in W_u \setminus u, i \in I_u\} - \mu_{W_u,u}(\theta); \Sigma_{W_u,u}(\theta) \Big). \end{aligned}$$

The estimator is given by

$$\hat{\theta}^{MLE2}_{k,n} = \underset{\theta \in (0,\infty)^E}{\arg\max}\, L\big(\theta; \{\Delta_{uv,i} : v \in W_u \setminus u, i \in I_u, u \in U\}\big).$$

The assumption under this definition is that for any $u, v \in U$ we have the independence condition $\Delta_{W_u \setminus u} \perp\!\!\!\perp \Delta_{W_v \setminus v}$, which is clearly not true for overlapping vertex sets W_u and W_v. However, this simplifies the joint likelihood function, and simulation results show that the estimator has comparable qualities to the one under the method of moments or the one based on extremal coefficients.

Covariance selection model Let $G(W_u)$ is the subgraph induced on the node set W_u. This graph must be connected. The maximum likelihood estimator of $\theta \in (0, \infty)^E$ is obtained in a two-step procedure similar to the MM estimator, but using the MLE of $\Sigma_{\cdot \setminus u}(\theta)$ instead.

As a first step a maximum likelihood approach is used to obtain an estimator of $\Sigma_{\cdot \setminus u}(\theta)$ and then a least squares procedure is used to estimate θ in the second step. The first step is an implementation of the Iterative Proportional Scaling algorithm from the `R` package `gRim` Højsgaard (2017). For description of the IPS we refer to Lauritzen (1996).
Step 1: For every $u \in V$ obtain

$$\widehat{\Sigma^{-1}}_{W_u \setminus u} = \underset{(\Sigma^{-1})_{W_u \setminus u}}{\arg\max}\, L\Big(\hat{\mu}_{W_u \setminus u}, (\Sigma^{-1})_{W_u \setminus u}; \{\Delta_{uv,i} : v \in W_u \setminus u, i \in I_u\} \Big).$$

Step 2: Solve the problem:

$$\hat{\theta}^{MLE}_{k,n} = \underset{\theta \in [0,\infty)^E}{\arg\min} \sum_{W_u \subseteq V, u \in V} \Big\| \widehat{\Sigma^{-1}}_{W_u \setminus u} - \Sigma_{W_u \setminus u}(\theta) \Big\|_F^2.$$

We illustrate the three estimators using the dataset on Seine. First, the nodes for which we do not observe realizations are extracted, the subsets are created, and we check if these can induce identifiability problem:

```
tobj<- Tree(seg, Seine)
Uc<- getNoDataNodes(tobj) #obtain the nodes with latent variables

subs<- Neighborhood()
subs<- subset(subs, 3, seg, Uc) # neighborhood of order three

# verify if the identifiability criterion is satisfied for every
    subgraph induced by a subset
is_identifiable(subs, tobj)
#> The nodes with latent variables { 5 } in set { Paris 2 Meaux
    Melun 5 } have degree less than three.
#>             The subgraph contains edge parameters that are
    non-identifiable.
#>
#> The nodes with latent variables { 5 } in set { Meaux 2 Paris
    Melun 5 } have degree less than three.
#>             The subgraph contains edge parameters that are
    non-identifiable.
#>
#> The nodes with latent variables { 2 } in set { Nemours 5 Melun
    Sens 2 } have degree less than three.
#>             The subgraph contains edge parameters that are
    non-identifiable.
#>
#> The nodes with latent variables { 2 } in set { Sens 5 Melun
    Nemours 2 } have degree less than three.
#>             The subgraph contains edge parameters that are
    non-identifiable.
#>

# change the order of the neighborhood and verify the
    identifiability again
subs<- subset(subs, 2, seg, Uc) # neighborhood of order two
is_identifiable(subs, tobj)
```

Since no message is produced after calling `is_identifiable`, the subsets are appropriate for estimation. They are created on the principle of neighborhood of order two for every observed variable.

Estimate CLE Version 1 We estimate using the first version of the CLE.

```
mle1<- MLE1(seg)
mle1<- suppressMessages(
            estimate(mle1, Seine, subs, k_ratio = 0.2)
)
```

Estimate CLE Version 2 Now we use the second version.

```
mle2<- MLE2(seg)
mle2<- suppressMessages(
            estimate(mle2, Seine, subs, k_ratio = 0.2)
)
```

The estimates from the two versions of the likelihood based estimators are very similar.

```
mle1$depParams
#>      e1      e2      e3      e4      e5      e6
#> 0.3111864 1.0747109 0.6152536 0.6901670 1.0305526 0.6546275
mle2$depParams
#>      e1      e2      e3      e4      e5      e6
#> 0.4097665 1.0571181 0.6065951 0.6859331 1.0104937 0.6784847
```

Estimate the Covariance Selection Method The Covariance Selection Method in Højsgaard (2017) is not applicable when there are latent variables. To illustrate the method using the dataset on Seine we create a graph without nodes with latent variables. This will be a tree on five nodes and four edges, hence we have only five parameters to estimate.

```
seg_short<- graph(c(1,2, 2,3, 2,4, 2,5), directed = FALSE)
name_stat<- c("Paris", "Melun", "Meaux", "Nemours", "Sens")
seg_short<- set.vertex.attribute(seg_short, "name", V(seg_short),
    name_stat)
```

The constructor of class `Tree` checks the nodes in the graph and the columns of the dataset if they match, and if there are latent variables.

```
tobj<- Tree(seg_short, Seine)
#> From validate.Network: Edges have been assigned names
#> From validate.Network: No latent variables
#> From validate.Network: Edges have been assigned names
#> From validate.Network: No latent variables
Uc<- getNoDataNodes(tobj)
```

Create the subsets.

```
subs_short<- Neighborhood()
subs_short<- subset(subs, 2, seg_short, Uc) # neighborhood of
    level two
```

Estimate using Covariance Selection Model.

```
mle<- MLE(seg_short)
mle<- suppressMessages(
            estimate(mle, Seine, subs_short, k_ratio = 0.2)
)
```

The estimates are squares of the parameters, hence take the square root.

```
sqrt(mle$depParams)
#>      e1      e2      e3      e4
#> 0.5010519 0.9133062 1.0333349 0.6907978
```

6.3.3 Extremal coefficients' estimator

The pairwise extremal coefficients' estimator is introduced in Einmahl, Kiriliouk, and Segers (2018), and is based on bivariate or trivariate stable tail dependence functions (stdf). It is described in Section 4.3 in Asenova, Mazo, and Segers (2021). We repeat some formulas for completeness.

For the Hüsler–Reiss distribution with parameter matrix $\Lambda(\theta)$ and for a pair of nodes $J = \{u, v\}$, the bivariate extremal coefficient is given by

$$l_J(1,1) = 2\Phi(\lambda_{uv}(\theta)),$$

with Φ the standard normal cumulative distribution function (cdf).

The non-parametric estimator of the stdf dates back to Drees and Huang (1998) and yields the following estimator for the extremal coefficient $l_J(1,\ldots,1)$ for $J \subseteq V$:

$$\hat{l}_{J;n,k}(1,\ldots,1) = \frac{1}{k}\sum_{i=1}^{n} \mathbb{1}\left(\max_{j\in J} n\hat{F}_{j,n}(\xi_{j,i}) > n + 1/2 - k\right).$$

Let $\mathcal{Q} \subseteq \{J \subseteq U : |J| = 2\}$, i.e., a collection of pairs of nodes associated to observable variables and put $q = |\mathcal{Q}|$, ensuring that $q \geq |E|$, the number of free edge parameters.

ECE Version 1 The pairwise extremal coefficients estimator (ECE) of θ is

$$\hat{\theta}^{\mathrm{ECE}}_{n,k} = \underset{\theta\in(0,\infty)^E}{\arg\min} \sum_{J\in\mathcal{Q}} \left(\hat{l}_{J;n,k}(1,1) - l_J(1,1;\theta)\right)^2.$$

One may include also trivariate extremal coefficients, in which case $l_J(1,1,1)$, $|J| = 3$ is composed of two bivariate normal cdfs.

ECE Version 2 The second version of the ECE which is implemented with object of class `EKS_part` uses the subsets W_u for every $u \in U$. It is similar to the CLE Version 1.

For fixed u and W_u such that $G(W_u)$ is a connected subgraph we apply the EC estimator of θ_{W_u} which is the collection of all edge weights within the subgraph $G(W_u)$. In the first step we solve for every $u \in U$ and given W_u

$$\hat{\theta}_{W_u,n,k} = \underset{\theta\in(0,\infty)^{W_u\setminus u}}{\arg\min} \sum_{J\in\mathcal{Q}_u} \left(\hat{l}_{J;n,k}(1,1) - l_J(1,1;\theta)\right)^2,$$

where $\mathcal{Q}_u$ is the collection of all possible pairs (and possibly triples) of nodes in W_u. In a second step we combine all estimates to obtain $(\theta_e, e \in E)$

$$\hat{\theta}^{\mathrm{ECEp}}_{k,n} = \underset{\theta\in[0,\infty)^E}{\arg\min} \sum_{u\in U}\sum_{e\in E} (\hat{\theta}_{e,W_u} - \theta_e)^2.$$

In some cases the second version may be faster than the first one.

Below we illustrate the application of these methods in `gremes`. We use again the dataset Seine. Consider the objects created earlier, the graph `seg`. Obtain the set of nodes for which we do not observe realizations

```
tobj<- Tree(seg, Seine) # via an object of class Tree
Uc<- getNoDataNodes(tobj)
Uc<- c("2", "5") # alternatively do it manually (not recommended)
```

Create the set of coordinates, in this case only vectors with two non-zero entries. If three variate extremal coefficients are to be computed we have to create an empty object of class `Triples` and pass it to method `evalPoints`.

```
tup<- Tuples()
x<- rep(1,5)
names(x)<- getNodesWithData(tobj)
tup<- evalPoints(tup, tobj, x)
```

Estimate ECE Version 1 Estimation with ECE requires to create an object of class EKS. Again 20% of the upper order statistics is used.

```
eks<- EKS(seg)
eks<-estimate(eks, Seine, tup, k_ratio = 0.2)
```

Estimate ECE Version 2 To apply ECE Version 2, we need to create subsets for local estimation.

```
subs<- Neighborhood()
subs<- subset(subs, 2, seg, Uc) # neighborhood of level two
```

Call the estimation method.

```
eks_part<- EKS_part(seg)
eks_part<- suppressMessages(
               estimate(eks_part, Seine, subs, k_ratio = 0.2, xx =
                   x)
)
```

The estimates are quite similar. For large graphs the second version is faster.

```
eks$depParams
#>      e1      e2      e3      e4      e5      e6
#> 0.6998687 1.2696199 0.2228986 0.5155293 1.1730987 0.9649355
eks_part$depParams
#>      e1      e2      e3      e4      e5      e6
#> 0.6430240 1.2985508 0.4067386 0.6610828 1.1893985 0.9446311
```

6.3.4 Cliquewise likelihood estimator

The estimator in this section is based on Engelke and Hitz (2020). We summarize their idea and introduce a variation of their estimator when there are latent variables.

The cliquewise estimator of Engelke and Hitz (2020) Recall the simple extreme value distribution with exponent measure that will be denoted by Λ only in this subsection in order to be in uniform with the notation in Engelke and Hitz (2020)

$$G(x_v, v \in V) = \exp(-\Lambda(x_v, v \in V)),$$

for every $(x_v, v \in V) \in \mathcal{E} := [0, \infty]^V \setminus \{0\}$. When Λ has positive and continuous density, the latter will be denoted by λ. The density of the multivariate Pareto distribution corresponding to this extreme value distribution is given by (Engelke and Hitz, 2020, equation 7)

$$f(y_v, v \in V) = \lambda(y_v, v \in V)/\Lambda(1_{|V|}), \tag{6.2}$$

where $1_{|V|}$ is a vector of ones of length $|V|$ and $(y_v, v \in V) \in \{z \in \mathcal{E} : \|z\|_\infty\} > 1$. Let further have $\mathcal{C}$ the set of cliques and $\mathcal{D}$ the set of minimal separators. The minimal separators in a block graph (including trees) is a single node. According to Engelke and Hitz (2020, Theorem 1) if Y has a multivariate Pareto distribution with positive and continuous density and satisfies the global or pairwise Markov property with respect to a decomposable graph (in the sense of equation 20 of their paper), the density of Y with expression as in (6.2) factorizes as follows

$$f_Y(y_v, v \in V) = \frac{1}{\Lambda(1_{|V|})} \frac{\prod_{C \in \mathcal{C}} \lambda_C(y_v, v \in C)}{\prod_{D \in \mathcal{D}} \lambda_D(y_v, v \in D)}. \tag{6.3}$$

We have λ_C the density of Λ_C which is the exponent measure associated to the marginal extreme value distribution $G_C(x_v, v \in C)$. Because for a block graph the minimal separators, $D \in \mathcal{D}$ are singletons we have $\lambda_D(y_v, v \in D) = y_D^{-2}$. Replace this in (6.3) and multiply and divide by $\prod_{C\in\mathcal{C}} \lambda_C(1_{|C|})$. We get from (6.3)

$$f_Y(y_v, v \in V) = \frac{1}{Z} \prod_{C\in\mathcal{C}} \frac{\lambda_C(y_v, v \in C)}{\Lambda_C(1_{|C|})},$$

with

$$Z = \Lambda(1_{|V|}) \prod_{D\in\mathcal{D}} y_D^{-2} \frac{1}{\prod_{C\in\mathcal{C}} \Lambda_C(1_{|C|})}.$$

Emphasizing that the density above depends on the edge weights $\theta = (\theta_C, C \in \mathcal{C})$, we get

$$f_Y(y_v, v \in V; \theta_C, C \in \mathcal{C}) = \frac{1}{Z(\theta)} \prod_{C\in\mathcal{C}} \frac{\lambda_C(y_v, v \in C; \theta_C)}{\Lambda_C(1_{|C|}; \theta_C)}.$$

By equation 7 or 35 in Engelke and Hitz (2020) we get

$$f_Y(y_v, v \in V; \theta_C, C \in \mathcal{C}) = \frac{1}{Z(\theta)} \prod_{C\in\mathcal{C}} f_C(y_v, v \in C; \theta_C). \tag{6.4}$$

The factorization into the densities in the last expression is in the origin of the idea to estimate each θ_C from the likelihood function only of the variables in that clique C. The term Z depends on all the parameters and it does not factorize, but the authors say that the results based on full likelihood and cliquewise likelihood estimation show not much difference between the two.

Recall our model for X: X Markov with respect to a block graph, X is parameterized cliquewise by Hüsler–Reiss copulas and it has unit Fréchet margins. In Asenova and Segers (2021) we showed that for such a model the Pareto limit is an extremal graphical model with respect to the same graph. We can thus apply Theorem 1 in Engelke and Hitz (2020), respectively equations (6.3) and (6.4) because the Pareto type limit of our model satisfies the pairwise (the global too) Markov properties with respect to the original graph (in the sense of equation 20 in Engelke and Hitz (2020)). We could implement the cliquewise estimator to data $(X_v/t, v \in V) \mid \max_v X_v > t$ for some large t, because the approximate distribution has density as in (6.4). However all variables must be observed.

Engelke and Hitz estimator with unobserved variables When there are unobserved variables, the density in (6.4) is inappropriate as some cliques will contain unobserved variables and hence there will be unidentifiable coefficients. In what follows we will present our estimator inspired from the idea of Engelke and Hitz (2020) and adapted to allow estimation with unobserved variables. It is this estimator that is implemented in `gremes`.

For non-empty disjoint subsets $A, B, C \subset V$ and $A \cup B \cup C = V$ such that C is separator of A and B, by Proposition 1 in Engelke and Hitz (2020) the density of Y as extremal graphical model factorizes as

$$f_Y(y_v, v \in V) = \frac{\lambda(y_v, v \in V)}{\Lambda(1_{|V|})} = \frac{\lambda_{A\cup C}(y_v, v \in A \cup C)\lambda_{B\cup C}(y_v, v \in B \cup C)}{\Lambda(1_{|V|})\lambda_C(y_v, v \in C)}. \tag{6.5}$$

We choose A, B, C such that C is a singleton, $A \cup C$ induces a subgraph for which the identifiability criterion holds (every node with latent variable must belong to at least

three cliques), and similarly for $B \cup C$. When C contains one node only, say node c, then $\lambda_c(y_c) = y_c^{-2}$. The density of Y_U is given by

$$f_U(y_v, v \in U) = \frac{1}{\Lambda(1_{|V|})y_c^{-2}} \int_{y_{\bar{U}}} \lambda_{A\cup C}(y_{A\cup C})\lambda_{B\cup C}(y_{B\cup C})\mathrm{d}y_{\bar{U}}.$$

If $C \notin \bar{U}$, there are no overlapping variables in $\lambda_{A\cup C}$ and $\lambda_{B\cup C}$, so we can split the integral

$$\frac{1}{\Lambda(1_{|V|})y_c^{-2}} \int_{y_{(A\cup C)\cap\bar{U}}} \lambda_{A\cup C}(y_{A\cup C})\mathrm{d}y_{(A\cup C)\cap\bar{U}} \times \int_{y_{(B\cup C)\cap\bar{U}}} \lambda_{B\cup C}(y_{B\cup C})\mathrm{d}y_{(B\cup C)\cap\bar{U}}.$$

The two integrals are the margins of λ, (see also equation 5 in Engelke and Hitz (2020)), hence we get

$$f_U(y_v, v \in U) = \frac{1}{\Lambda(1_{|V|})y_c^{-2}} \lambda_{(A\cup C)\cap U}(y_{(A\cup C)\cap U})\lambda_{(B\cup C)\cap U}(y_{(B\cup C)\cap U}).$$

Multiplying and dividing by both the measures $\Lambda_{(A\cup C)\cap U}(1_{|(A\cup C)\cap U|})$ and by $\Lambda_{(B\cup C)\cap U}(1_{|(B\cup C)\cap U|})$ we get

$$f_U(y_U) = \frac{1}{Z} \frac{\lambda_{(A\cup C)\cap U}(y_{(A\cup C)\cap U})}{\Lambda_{(A\cup C)\cap U}(1_{|(A\cup C)\cap U|})} \frac{\lambda_{(B\cup C)\cap U}(y_{(B\cup C)\cap U})}{\Lambda_{(B\cup C)\cap U}(1_{|(B\cup C)\cap U|})}$$

with

$$Z = \frac{\Lambda(1_{|V|})y_c^{-2}}{\Lambda_{(B\cup C)\cap U}(1_{|(B\cup C)\cap U|})\Lambda_{(A\cup C)\cap U}(1_{|(A\cup C)\cap U|})}.$$

Using equation 7 or 35 in Engelke and Hitz (2020) we write

$$f_U(y_U) = \frac{1}{Z} f_{(A\cup C)\cap U}(y_{(A\cup C)\cap U}) f_{(B\cup C)\cap U}(y_{(B\cup C)\cap U}).$$

Consider $\theta_{A\cup C}$ and $\theta_{B\cup C}$. These are the weights on the edges in the subgraph induced on the nodes of the corresponding set. We have $\theta = (\theta_{A\cup C}, \theta_{B\cup C})$. As a function of θ, we have

$$f_Y(y_v, v \in V; \theta) = \frac{1}{Z(\theta)} f_{(A\cup C)\cap U}(y_{(A\cup C)\cap U}; \theta_{A\cup C}) \times f_{(B\cup C)\cap U}(y_{(B\cup C)\cap U}; \theta_{B\cup C}). \tag{6.6}$$

The expression in (6.6) suggests to estimate $\theta_{A\cup C}$ and $\theta_{B\cup C}$ separately, ignoring the term $Z(\theta)$ as explained in Engelke and Hitz (2020, Section 5). It is important that A, B, C are chosen so that the identifiability criterion holds for each of the subgraphs induced by the node sets $A \cup C$ and $B \cup C$.

The method can be generalized if we further decompose $A \cup C$ in two sets and a separator - *singleton* between them, for instance $\bar{A}, \bar{B}, \bar{C}$ such that $\bar{A} \cup \bar{B} \cup \bar{C} = A \cup C$ and $\bar{C}$ is the singleton separator between $\bar{A}$ and $\bar{B}$. Further, we should have that on the subgraph induced by $\bar{A} \cup \bar{C}$ the identifiability criterion is satisfied. Similarly for $\bar{B} \cup \bar{C}$. Then using again Engelke and Hitz (2020, Proposition 1 and equation 7) we have

$$f_{A\cup C} = \frac{\lambda_{A\cup C}}{\Lambda_{A\cup C}(1_{|A\cup C|})} \qquad \text{and} \qquad f_{A\cup C} = \frac{\lambda_{\bar{A}\cup\bar{C}}\lambda_{\bar{B}\cup\bar{C}}}{\Lambda_{A\cup C}(1_{|A\cup C|})\lambda_{\bar{C}}}$$

from which it follows

$$\lambda_{A\cup C} = \frac{\lambda_{\bar{A}\cup\bar{C}}\lambda_{\bar{B}\cup\bar{C}}}{\lambda_{\bar{C}}}.$$

We replace $\lambda_{A\cup C}$ in (6.5) with the right hand side of the above expression. We can continue like this to obtain an expression for f_U, given by

$$f_U(y_U;\theta) = \frac{1}{Z(\theta)}\prod_i \frac{\lambda_{B_i\cap U}(y_{B_i\cap U};\theta_{B_i})}{\Lambda_{B_i\cap U}(1;\theta_{B_i})} = \frac{1}{Z(\theta)}\prod_i f_{B_i\cap U}(y_{B_i\cap U};\theta_{B_i}), \qquad (6.7)$$

for some carefully chosen subsets $\{B_i\}$. The idea is to estimate θ_{B_i} separately from $f_{B_i\cap U}$ ignoring $Z(\theta)$. For instance, in Seine network the only possible subsets which will allow estimation of all edge weights is $B_1 = \{\text{Paris}, 2, \text{Meaux}, \text{Melun}\}$ and $B_2 = \{\text{Melun}, 5, \text{Nemours}, \text{Sens}\}$. Instead of using cliques, we use the smallest subgraphs from which the corresponding subset of parameters is still uniquely identifiable.

For implementation of the estimator based on (6.7) we need

- the subsets $\{B_i\}$;
- the set of indices $I_{k,n}$ given by

$$I_{k,n} = \{i = 1,\dots,n : \max_{v\in U}\hat{X}_{v,i} > n/k\}, \qquad k \in \{1,\dots n\};$$

- for every $i \in I_{k,n}$ compose the ratios

$$\hat{Y}_{v,i} = \hat{X}_{v,i}/(n/k),$$

  which gives rise to a sample $(\hat{y}_{v,i}, v \in U, i \in I_{k,n})$.

For given B from the set $\{B_i\}$ an estimate of $\theta_B = (\theta_{ij}, i,j \in B, (i,j) \in E)$ is obtained by maximizing

$$\prod_{j\in I_{k,n}} f_{B\cap U}(\theta_B;\hat{y}_{B\cap U,j}) = \prod_{j\in I_{k,n}} \frac{\psi_{B\cap U}(\theta_B;\hat{y}_{B\cap U,j})}{\Psi_{B\cap U}(\theta_B;1_{|B\cap U|})},$$

where Ψ and ψ are the Hüsler-Reiss Pareto exponent measure and its density respectively as in Section 2 of Engelke and Hitz (2020). We have

$$\Psi_{B\cap U}(\theta_B;1_{|B\cap U|}) = l_{B\cap U}(\theta_B;1_{|B\cap U|}) = l(\theta;\iota_{B\cap U}),$$

where $l_{B\cap U}$ is the stdf of the subvector $X_{B\cap U}$, l is the stdf of the full vector $X = (X_v, v \in V)$, and $\iota_{B\cap U} = (1_{i\in B\cap U}, i \in V)$.

The density of the exponent measure ψ is given for every $u \in B\cap U$ by (see Engelke and Hitz (2020, Section 2))

$$\begin{aligned}&\psi_{B\cap U}(\theta_B;\hat{y}_{B\cap U,j})\\ &= \hat{y}_{u,j}^{-2}\prod_{v\neq u}\hat{y}_{v,j}^{-1}\phi_{|B\cap U|-1}\Big(\ln(\hat{y}_{u,j}/\hat{y}_{v,j}) + \lambda_{uv}^2(\theta_B)/2;\Sigma_{B\cap U,u}(\theta_B)\Big).\end{aligned}$$

The vector of all $\hat{\theta}_B$ for all B in the collection $\{B_i\}$ is what we called the cliquewise likelihood estimator or the Engelke-and-Hitz estimator.

We will illustrate it using the dataset Seine. Create a list of subsets on the node set. These sets should cover all the vertex set, but every two subsets are allowed to have at most one node in common.

```
subs<- list(c("Paris", "2", "Meaux", "Melun"), c("Melun", "5",
    "Nemours", "Sens"))
```

We require that within the induced subgraph of a given subset all edge parameters are identifiable. This is key in this estimator. We can check this requirement with the function `is_identifiable`. To use this function we need however to create an object of class `RootDepSet` and an object of class `Tree`. If it is clear that the criterion is satisfied there is no need from this step and the two additional objects created below.

```
rdsobj<- RootDepSet()
rdsobj<- setRootDepSet(rdsobj, subs, c("Paris", "Melun"))
#> From setRootDepSet.RootDepSet: The order of the subset must
    correspond to the
#>         order of its corresponding root
tobj<- Tree(seg, Seine)
#> From validate.Network: Edges have been assigned names
#> From validate.Network: There are nodes with latent variables
#> From validate.Network: Edges have been assigned names
#> From validate.Network: There are nodes with latent variables
is_identifiable(rdsobj, tobj)
```

We get no messages from the call of the function `is_identifiable`, the identifiability criterion is satisfied. Estimate using the cliquewise likelihood method.

```
ehobj<- EngHitz(seg)
#> From HRMnetwork: Edges have been assigned names
ehobj<- suppressMessages(
estimate(ehobj, Seine, subs, k_ratio = 0.2)
)
ehobj$depParams
#>       e1       e2       e3       e4       e5       e6
#> 0.3653028 1.0398823 0.5751127 0.6949520 1.0905722 0.6667259
```

6.3.5 Conditioning on the mean

This estimator differs from the others because the conditioning event does not depend on a particular node u but it depends on the event that the geometric mean exceeds a high threshold. It is based on the limit result in (L4) and it is applicable when there are latent variables.

In an unpublished note Segers (2019) it is shown that if the random vector $X = (X_1, \ldots, X_d)$ has unit Pareto margins and is in the max-domain of attraction of a Hüsler-Reiss copula with parameter matrix $\Lambda = (\lambda^2)_{ij}$, then it holds

$$\mathcal{L}\Big((Y_v - \bar{Y})_{v=1}^d | \bar{Y} > y\Big) \to \mathcal{N}_d(\bar{\mu}, \bar{\Sigma}),$$

with $Y = (Y_1, \ldots, Y_d) = (\ln X_1, \ldots, \ln X_d)$ and certain matrices $\bar{\Sigma}$ and $\bar{\mu}$. We do not provide further details on this result as it an unpublished work.

Consider a tree $T = (V, E)$ and edge weights $\theta = (\theta_e, e \in E)$. Under the assumption that $X = (X_v, v \in V)$ is in the domain of attraction of a Hüsler-Reiss copula with unit Fréchet margins and structured parameter matrix $\Lambda(\theta)$

$$\big(\Lambda(\theta)\big)_{ij} = \lambda_{ij}^2(\theta) = \frac{1}{4} \sum_{e \in p(i,j)} \theta_e^2, \qquad i, j \in V,\ i \neq j, e \in E.$$

we can employ the method of moments or the composite likelihood method to estimate $\theta = (\theta_e, e \in E)$ from $\bar{\Sigma}(\theta)$.

The method of moments estimator The method of moments estimator is given by

$$\hat{\theta}_{k,n}^{\text{MMave}} = \arg \min_{\theta \in (0,\infty)^E} \|\hat{\Sigma}_U - \bar{\Sigma}_U(\theta)\|_F^2$$

- $\hat{\Sigma}_U$ is the non-parametric covariance matrix

- $\bar{\Sigma}_U(\theta)$ is the parametric covariance matrix
- The parametric matrix $\bar{\Sigma}_U(\theta)$ is given by

$$\bar{\Sigma}_U(\theta) = -M\big(\Lambda(\theta)\big)_{i,j\in U}M$$

with

$$\big(\Lambda(\theta)\big)_{ij} = \lambda_{ij}^2(\theta) = \frac{1}{4}\sum_{e\in p(i,j)}\theta_e^2, i \neq j, e \in E.$$

and certain matrix M.

- If the sample of the original variables is $\xi_{v,i}, v \in U, i = 1,\dots,n$ consider the transformation using the empirical cumulative distribution function $\hat{F}_{v,n}(x) = \big[\sum_{i=1}^n \mathbb{1}(\xi_{v,i} \leq x)\big]/(n+1)$.

$$\hat{X}_{v,i} = \frac{1}{1-\hat{F}_{v,n}(\xi_{v,i})}, \qquad v \in U, \quad i = 1,\dots,n.$$

Then consider their logarithm

$$\hat{Y}_{v,i} = \ln \hat{X}_{v,i}$$

- For given $k \in \{1,\dots n\}$ consider the set of indices

$$I = \left\{i = 1,\dots,n : \overline{\hat{Y}}_i = \left(\frac{1}{|U|}\sum_{v\in U}\hat{Y}_{v,i}\right) > n/k\right\}$$

- For every $v \in U$ and $i \in I$ compose the differences

$$\Delta_{v,i} = \hat{Y}_{v,i} - \overline{\hat{Y}}_i.$$

- The vector of means of these differences is given by

$$\hat{\mu} = \frac{1}{|I|}\sum_{i\in I}(\Delta_{v,i}, v \in U).$$

- The non-parametric covariance matrix $\hat{\Sigma}_U$ is given by

$$\hat{\Sigma} = \frac{1}{|I|}\sum_{i\in I}(\Delta_{v,i} - \hat{\mu}, v \in U)(\Delta_{v,i} - \hat{\mu}, v \in U)^\top.$$

The composite likelihood estimator The composite likelihood estimator is given by

$$\hat{\theta}_{k,n}^{\text{MLEave}} = \arg\max_{\theta\in(0,\infty)^{|E|}} L\Big(\bar{\mu}_U(\theta), \bar{\Sigma}_U(\theta); \{\Delta_{v,i}, i \in I, v \in U\}\Big).$$

The likelihood function L above is the one of $|U|$-variate Gaussian probability density function with mean $\bar{\mu}_U$ and covariance matrix $\bar{\Sigma}_U$.

We illustrate the two estimators on the Seine dataset. Create an object combining the graph and the data. We can extract from it the nodes with latent variables.

```
tobj<- Tree(seg, Seine)
#> From validate.Network: Edges have been assigned names
#> From validate.Network: There are nodes with latent variables
#> From validate.Network: Edges have been assigned names
#> From validate.Network: There are nodes with latent variables
Uc<- getNoDataNodes(tobj)
```

Method of moments estimator Estimate according to the method of moments: first create an object of the appropriate class and then call the method `estimate`.

```
mme_ave<- MMEave(seg)
#> From HRMnetwork: Edges have been assigned names
estimate(mme_ave, Seine, k_ratio=0.2)$depParams
#> From validate.Network: There are nodes with latent variables
#> From validate.Network: There are nodes with latent variables
#> From setRootDepSet.RootDepSet: The order of the subset must
    correspond to the
#>         order of its corresponding root
#> From setParams.HRMtree: Names have been attributed to the
    vector 'value' in the order corresponding to the order of the
    edges: The fist element has the name of the first edge, the
    second element the name of the second edge, etc.
#> From setParams.HRMtree: The parameters have been attached to
    the edges according to their names
#>      e1      e2      e3      e4      e5      e6
#> 0.3991855 1.7674395 0.7114457 0.8146106 1.7026741 0.9087931
```

Composite likelihood estimator Estimate according to the method of maximum likelihood: first create an object of the appropriate class and then call the method `estimate`.

```
mle_ave<- MLEave(seg)
#> From HRMnetwork: Edges have been assigned names
estimate(mle_ave, Seine, k_ratio=0.2)$depParams
#> From validate.Network: There are nodes with latent variables
#> From validate.Network: There are nodes with latent variables
#> From setRootDepSet.RootDepSet: The order of the subset must
    correspond to the
#>         order of its corresponding root
#>      e1      e2      e3      e4      e5      e6
#> 0.3416488 1.6908232 0.9294585 1.0357446 1.6046801 0.9647892
```

The estimates are slightly different from each other. Note also that the estimates are a bit higher compared to MME and CLE in sections 6.3.1 and 6.3.2.

Although local estimation is possible using a collection of subsets around every node with observable variable, we do not pursue this method here.

6.4 Estimation in models on block graphs

The estimator is based on the limit result in section 6.2.2 and it is appropriate when there are latent variables. We have chosen to implement an estimation method which is simple, fastest with respect to all the others proposed for tree models and it gives very satisfactory results, at least for tree models. The idea is to find the edge weights $\delta = (\delta_e^2, e \in E)$ which minimize the distance between the empirical and the theoretical covariance matrices:

$$\hat{\delta}_{n,k}^{\mathrm{MM}} = \arg \min_{\delta \in (0,\infty)^E} \sum_{u \in U} \|\hat{\Sigma}_{W_u,u} - \Sigma_{W_u,u}(\delta)\|_F^2 .$$

Some of the notations are the same as in section 6.3.1, some of them we repeat for completeness.

- W_u is a subset on the node set depending on u. Typically a neighborhood of u or the nodes that are flow connected to u or the intersection of both. Note that

the induced graph on W_u must be connected. A good practice is to compose the sets such that within each subset all parameters are uniquely identifiable. This means that every node in W with latent variable should be connected to at least three other cliques in the same set W.

- $\hat{\Sigma}_{W_u,u}$ is the non-parametric covariance matrix
- $\Sigma_{W_u,u}(\delta)$ is the parametric covariance matrix
- For fixed u and W_u the parametric matrix $\Sigma_{W_u,u}$ is given by

$$\big(\Sigma_{W,u}(\Lambda)\big)_{ij} = 2(\lambda_{iu}^2 + \lambda_{ju}^2 - \lambda_{ij}^2), \qquad i,j \in W \setminus u.$$

with

$$\big(\Lambda(\delta)\big)_{ij} = \lambda_{ij}^2(\delta) = \sum_{e \in p(i,j)} \delta_e^2, \qquad i,j \in V,\ i \neq j, e \in E.$$

- If the sample of the original variables is $\xi_{v,i}, v \in U, i = 1,\dots,n$ consider the transformation using the empirical cumulative distribution function $\hat{F}_{v,n}(x) = \big[\sum_{i=1}^n \mathbb{1}(\xi_{v,i} \leq x)\big]/(n+1)$.

$$\hat{X}_{v,i} = \frac{1}{1 - \hat{F}_{v,n}(\xi_{v,i})}, \qquad v \in U, \quad i = 1,\dots,n.$$

- Fix u and W_u. For given $k \in \{1,\dots n\}$ consider the set of indices

$$I_u = \{i = 1,\dots,n : \hat{X}_{u,i} > n/k\}$$

- For every $v \in W_u \setminus u$ and $i \in I_u$ compose the differences

$$\Delta_{uv,i} = \ln \hat{X}_{v,i} - \ln \hat{X}_{u,i}.$$

- The vector of means of these differences is given by

$$\hat{\mu}_{W_u,u} = \frac{1}{|I_u|} \sum_{i \in I_u} (\Delta_{uv,i}, v \in W_u \setminus u).$$

- The non-parametric covariance matrix $\hat{\Sigma}_{W_u,u}$ is given by

$$\hat{\Sigma}_{W_u,u} = \frac{1}{|I_u|} \sum_{i \in I_u} (\Delta_{uv,i} - \hat{\mu}_{W_u,u}, v \in W_u \setminus u)(\Delta_{uv,i} - \hat{\mu}_{W_u,u}, v \in W_u \setminus u)^\top.$$

We illustrate the estimation methods on a simulated data. We create the `igraph` object which is with the diagram as in Fig. 6.2

```
g<- graph(c(1,3,1,2,2,3,
      3,4,4,5,5,3,
      3,7,3,6,6,7), directed = FALSE)
g<- set.vertex.attribute(g, "name", V(g), c("a", "b", "c", "d",
    "e", "f", "g"))
```

Create the edge weights to be assigned to the edges of the graph.

```
# all deltas are squares already
C1<- c(0.2, 0.8, 0.6) # d_13^2, d_12^2, d_23^2
C2<- c(0.3, 0.5, 0.1) # d_34^2, d_45^2, d_35^2
C3<- c(0.4, 0.05, 0.25) # d_37^2, d_36^2, d_67^2
```

We create an object of class HRMBG and attach the edge weights to the edges using the method `setParams`.

```
hrmbgobj<- HRMBG(g)
#> From HRMnetwork: Edges have been assigned names
hrmbgobj<- setParams(hrmbgobj, c(C1, C2, C3))
#> From setParams.HRMtree: Names have been attributed to the
    vector 'value' in the order corresponding to the order of the
    edges: The fist element has the name of the first edge, the
    second element the name of the second edge, etc.
#> From setParams.HRMtree: The parameters have been attached to
    the edges according to their names
hrmbgobj
#> $graph
#> IGRAPH 28f4820 UN-- 7 9 --
#> + attr: name (v/c), name (e/c)
#> + edges from 28f4820 (vertex names):
#> [1] a--c a--b b--c c--d d--e c--e c--g c--f f--g
#>
#> $depParams
#>  e1  e2  e3  e4  e5  e6  e7  e8  e9
#> 0.20 0.80 0.60 0.30 0.50 0.10 0.40 0.05 0.25
#>
#> attr(,"class")
#> [1] "HRMnetwork" "HRMBG" "MME"
```

We create the matrix Λ, whose entry λ_{ij} is the sum of the edge weights on the unique shortest path between node i and node j.

```
hrmlam<- HRLambda(hrmbgobj)
hrmlam
#>    a   b   c   d   e   f   g
#> a 0.00 0.80 0.20 0.50 0.30 0.25 0.60
#> b 0.80 0.00 0.60 0.90 0.70 0.65 1.00
#> c 0.20 0.60 0.00 0.30 0.10 0.05 0.40
#> d 0.50 0.90 0.30 0.00 0.50 0.35 0.70
#> e 0.30 0.70 0.10 0.50 0.00 0.15 0.50
#> f 0.25 0.65 0.05 0.35 0.15 0.00 0.25
#> g 0.60 1.00 0.40 0.70 0.50 0.25 0.00
```

Generate 1000 observations from Hüsler-Reiss distribution with parameter matrix Λ and some independent random noise.

```
X<- rHRM(hrmbgobj, hrmlam, 1000, noise = TRUE)
```

We will treat the variable X_c as unobserved. If c is unobserved, given the graph structure we don't have much choice in creating the subsets. If we follow the recommendation to choose subset such that in the subgraph induced by the subset the edges are uniquely identifiable we simply take all observable nodes six times.

```
rdsobj<- RootDepSet()
rdsobj<- setRootDepSet(rdsobj, subset = list(c("a", "b", "d", "e",
    "f", "g"),
c("a", "b", "d", "e", "f", "g"),
c("a", "b", "d", "e", "f", "g"),
c("a", "b", "d", "e", "f", "g"),
c("a", "b", "d", "e", "f", "g"),
c("a", "b", "d", "e", "f", "g")),
c("a", "b", "d", "e", "f", "g"))
```

Estimate the model treating the third variable as latent: create first an object of class

HRMBG and then use on it the method `estimate`.

```
hrmbg<- HRMBG(g)
#> From HRMnetwork: Edges have been assigned names
hrmbg<- suppressMessages(
estimate(hrmbg, X[,-3], rdsobj, k_ratio = 0.2)
)
hrmbg$depParams
#>       e1       e2       e3       e4       e5       e6       e7
#> 0.11626913 0.33142701 0.22438322 0.12975244 0.28151050
    0.08391488 0.20506818
#>       e8       e9
#> 0.05308032 0.16180281
```

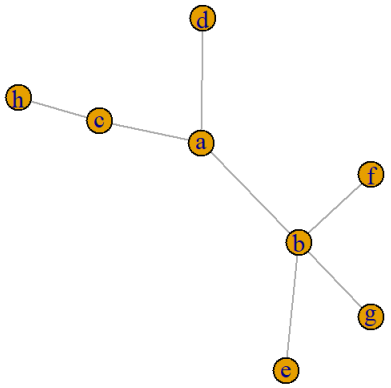
6.5 Additional functionalities

The package contains several additional functionalities that may serve for generating data from the model and post-estimation analysis.

6.5.1 Random sample from a Markov tree parameterized cliquewise by Hüsler-Reiss distributions

Consider the random vector $X = (X_v, v \in V)$ with three properties: it satisfies the global Markov property with respect to a tree $T = (V, E)$, every two adjacent nodes have Hüsler-Reiss distribution with some parameter $\theta_e, e \in E$, and the univariate margins are unit Fréchet, $F(x) = \exp(-1/x)$. For simulation purposes, it is useful to be able to generate from the model for X. The method that does this is `rHRM.HRMtree`. There are two options - with or without independent normal noise.

```
seg<- make_tree(8,3, mode = "undirected") # create the tree
seg<- set.vertex.attribute(seg, "name", V(seg), letters[1:8]) #
    name the nodes
plot(seg)
```




```
hrm<- HRMtree(seg) # initialize the object of of class HRMtree
#> From HRMnetwork: Edges have been assigned names
hrm<- setParams(hrm, seq(0.1, 0.7, 0.1)) # set its parameters
#> From setParams.HRMtree: Names have been attributed to the
    vector 'value' in the order corresponding to the order of the
    edges: The fist element has the name of the first edge, the
    second element the name of the second edge, etc.
#> From setParams.HRMtree: The parameters have been attached to
    the edges according to their names
X<- rHRM(hrm, 1000) # generate a random sample
round(head(X), 4)
#>         a      b      c      d      e      f      g      h
#> [1,] 3.4472 3.9739 4.0591 3.9641 3.6598 1.4403 5.9760 1.4458
#> [2,] 0.5888 0.5696 0.6494 0.3557 0.4057 0.7999 0.5375 0.2906
#> [3,] 1.3907 1.4325 1.9571 1.5407 1.7510 1.1323 3.5939 0.8053
#> [4,] 1.9043 2.0291 2.5510 1.4914 2.3184 2.6095 2.2060 2.3886
#> [5,] 157.8761 144.0455 182.9044 328.4986 153.4239 127.6369
    161.5266 429.3421
#> [6,] 20.7303 24.5769 17.7094 13.4588 32.2752 35.7236 18.4102
    16.3959
XX<- rHRM(hrm, 1000, noise = TRUE) # generate a random samle with
    independent normal noise
round(head(XX), 4)
#>         a      b      c      d      e      f      g      h
#> [1,] 3.0549 3.9415 3.9036 3.1851 2.3651 4.5122 3.4977 2.3783
#> [2,] 1.0991 1.2879 1.3439 0.5735 1.8647 0.7414 0.9795 0.5649
#> [3,] 12.6987 11.2157 9.6800 12.9609 6.5391 12.3512 5.4284 9.6943
#> [4,] 2.4669 2.9649 3.2005 3.3334 6.0280 4.4191 9.9984 4.3267
#> [5,] 3.0698 3.2577 4.4243 2.2971 3.1269 2.9672 3.0714 2.5784
#> [6,] 2.3324 2.4746 1.4741 1.2014 1.9920 1.2968 4.5669 1.5034
```

For any $u \in V$ the joint density function of X is

$$f(x) = f_u(x_u) \prod_{(v,j)\in E_u} f_{j|v}(x_j \mid x_v),$$

with $E_u \subseteq E$ the set of edges directed away from u, i.e., $(v, j) \in E_u$ if and only if $v = u$ or v separates u and j. The joint density f is determined by $d - 1$ bivariate densities f_{vj}. The univariate margins f_u are unit Fréchet densities, $f_j(x_j) = \exp(-1/x_j)/x_j^2$ for $x_j \in (0, \infty)$, and the bivariate margins for each pair of variables on adjacent vertices j, v are Hüsler–Reiss distributions with parameter θ_{jv}.

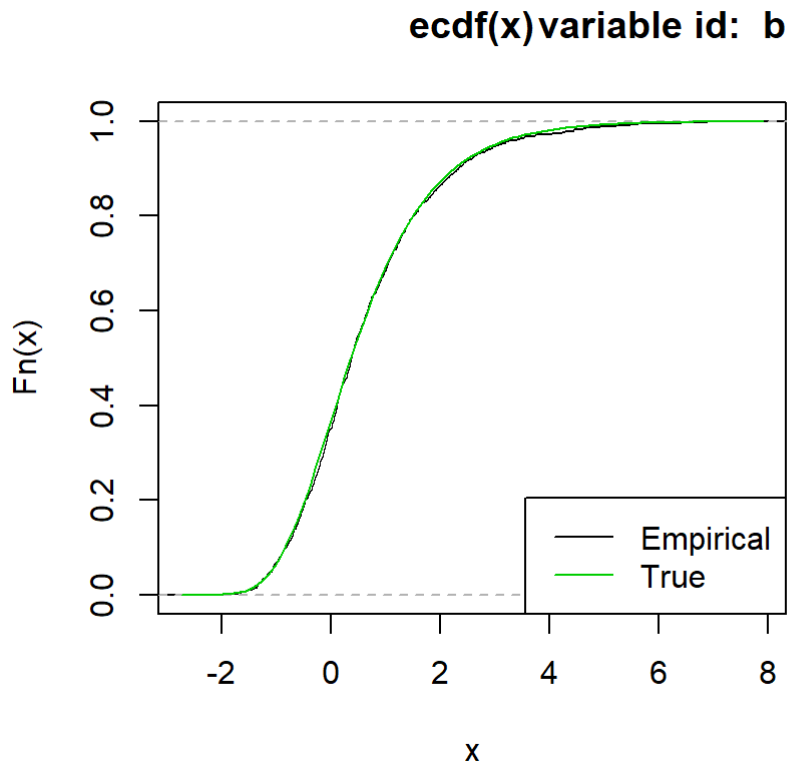
To generate an observation from the left hand-side of the equation above we use the right hand-side of that equation, proceeding iteratively, walking along paths starting from u using the conditional densities. An observation of X_j given $X_v = x_v$ is generated via the inverse function of the conditional cdf $x_j \mapsto F_{j|v}(x_j \mid x_v)$, the conditional cdf of x_j given x_v. To do so, the equation $F_{j|v}(x_j \mid x_v) - p = 0$ is solved numerically as a function in x_j for fixed $p \in (0, 1)$. The choice of the Hüsler–Reiss bivariate distribution gives the following expression for $F_{j|v}(x_j \mid x_v)$:

$$\Phi\left(\frac{\theta_{jv}}{2} + \frac{1}{\theta_{jv}} \ln \frac{x_j}{x_v}\right)$$
$$\cdot \exp\left[-\frac{1}{x_v}\left\{\Phi\left(\frac{\theta_{jv}}{2} + \frac{1}{\theta_{jv}} \ln \frac{x_j}{x_v}\right) - 1\right\} - \frac{1}{x_j}\Phi\left(\frac{\theta_{jv}}{2} + \frac{1}{\theta_{jv}} \ln \frac{x_v}{x_j}\right)\right].$$

6.5.2 Diagnostic for the random sample

As a diagnostic tool of the joint distribution of the sample generated above, we provide comparison between the bivariate true copula and the bivariate empirical copula. We fix a node, say a, and if b and c are adjacent to a, we compute the true bivariate copulas and the empirical copulas of pairs (X_a, X_b) and (X_a, X_c) and compare them. We also provide plot of the true and the empirical cumulative distribution function of the selected variable, e.g., of a in the preceding example.

```
diagnost(hrm, X, "b", y = c(0.3,0.5))
```

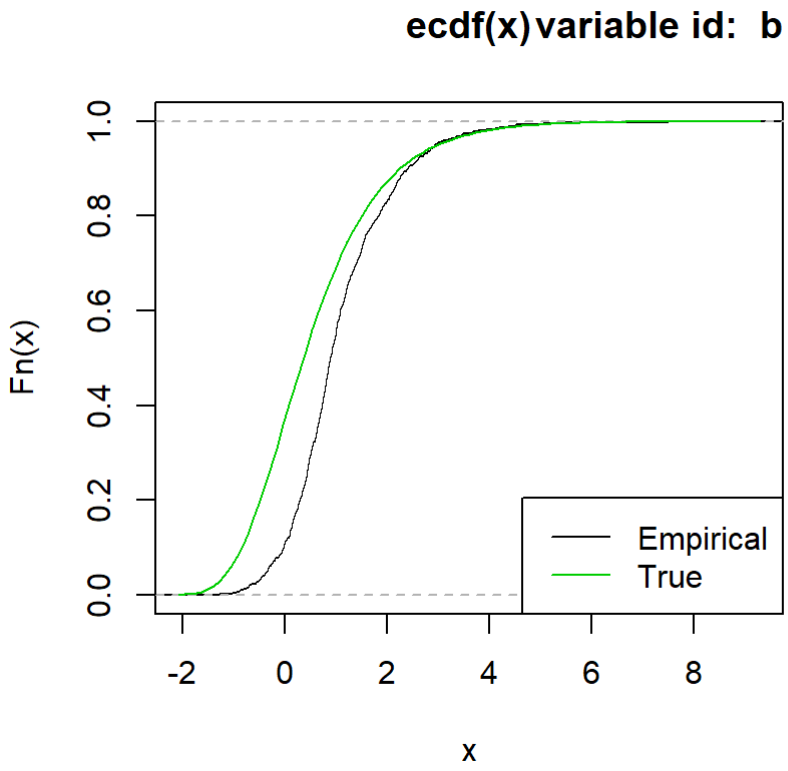



```
#> [1] "variable b adj. variable a ; true copula 0.299999999921372
    ; empirical copula 0.3"
#> [1] "variable b adj. variable e ; true copula 0.29589026650931
    ; empirical copula 0.292"
#> [1] "variable b adj. variable f ; true copula 0.291002999050035
    ; empirical copula 0.287"
#> [1] "variable b adj. variable g ; true copula 0.28491676575979
    ; empirical copula 0.289"
```

When we add some noise to the model the univariate empirical curve is a bit below the true one, but in the tail the difference diminishes.

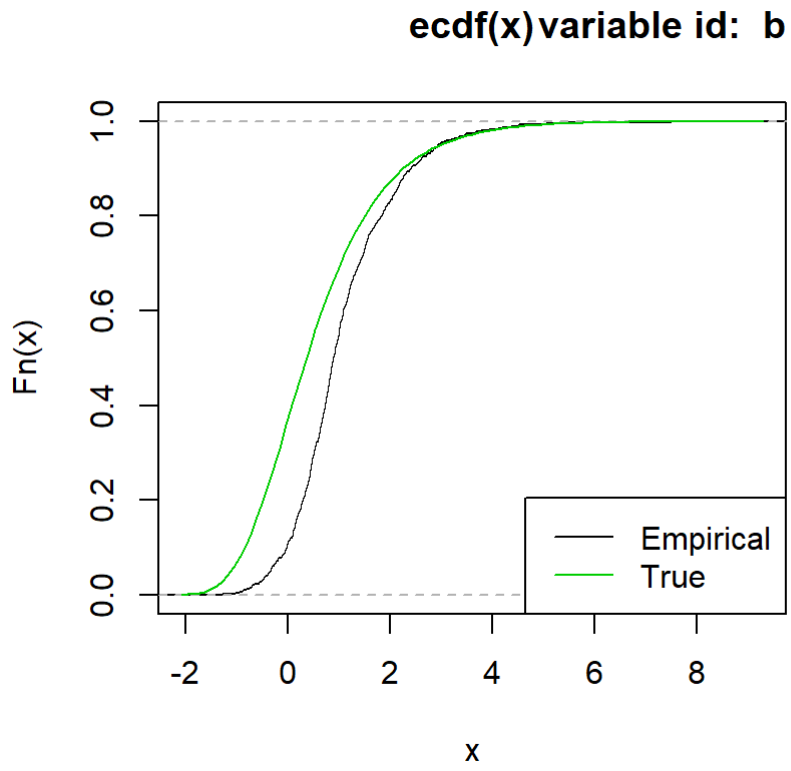
Similarly for the tail of the bivariate copulas: for higher coordinates, (0.8, 0.9), the bivariate true and empirical copulas become closer to each other than for smaller values of the coordinates (0.3, 0.5):

```
diagnost(hrm, XX, "b", y = c(0.3, 0.5))
```




```
#> [1] "variable b adj. variable a ; true copula 0.299999999921372
    ; empirical copula 0.266"
#> [1] "variable b adj. variable e ; true copula 0.29589026650931
    ; empirical copula 0.275"
#> [1] "variable b adj. variable f ; true copula 0.291002999050035
    ; empirical copula 0.253"
#> [1] "variable b adj. variable g ; true copula 0.28491676575979
    ; empirical copula 0.243"
```

```
diagnost(hrm, XX, "b", y = c(0.8, 0.9))
```

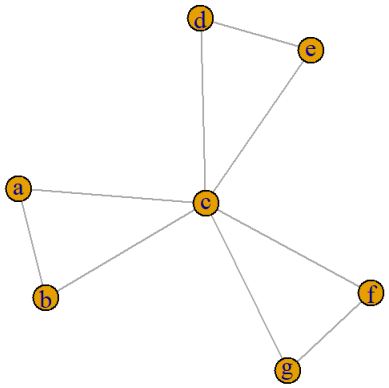



```
#> [1] "variable b adj. variable a ; true copula 0.8 ; empirical
    copula 0.8"
#> [1] "variable b adj. variable e ; true copula 0.799431996325495
    ; empirical copula 0.799"
#> [1] "variable b adj. variable f ; true copula 0.798248020703279
    ; empirical copula 0.798"
#> [1] "variable b adj. variable g ; true copula 0.796402838196817
    ; empirical copula 0.795"
```

6.5.3 Random sample from max-stable Hüsler-Reiss distribution with structured matrix

When X is Markov with respect to a block graph and parameterized cliquewise by a family of Hüsler-Reiss distributions, we do not dispose of a method to generate from the exact distribution of such a model as we do when the graph is a tree. We can generate from the max-stable attractor of this model. To do this we use the package `mev` of Belzile et al. (2020).

```
bg<- graph(c(1,2,2,3,1,3,
3,4,3,5,4,5,
3,7,3,6,6,7), directed = FALSE) # create the graph
bg<- set.vertex.attribute(bg, "name", V(bg), letters[1:7]) # name
    the nodes
plot(bg)
```




```
hrbg<- HRMBG(bg) # initialize an object with zero dependence
    parameters
#> From HRMnetwork: Edges have been assigned names
hrbg<- setParams(hrbg, seq(0.1, 0.9, 0.1)) # set the parameters
#> From setParams.HRMtree: Names have been attributed to the
    vector 'value' in the order corresponding to the order of the
    edges: The fist element has the name of the first edge, the
    second element the name of the second edge, etc.
#> From setParams.HRMtree: The parameters have been attached to
    the edges according to their names
lam<- HRLambda(hrbg) # compute the structured matrix Lambda,
XB<- rHRM(hrbg, lam, 1000, noise = TRUE)
round(head(XB), 4)
#>       a    b    c    d    e    f     g
#> [1,] 0.8995 2.4714 2.1611 2.2669 1.5366 0.9505 1.0385
#> [2,] 7.9079 8.1340 3.3715 1.8902 1.3748 1.7690 3.6102
#> [3,] 2.2044 1.5360 2.1932 3.8561 5.6708 4.3940 4.7238
#> [4,] 3.3555 2.9799 2.5561 3.0176 2.7400 4.7078 11.2548
#> [5,] 1.3955 0.8804 1.5267 1.9929 1.4649 0.9123 3.5049
#> [6,] 2.8644 3.7879 2.5622 1.8191 1.5134 1.1738 1.7523
```

This code samples from a max-stable Hüsler-Reiss copula with unit Fréchet univariate margins and with parameter matrix

$$\lambda_{ij}^2 = \sum_{e \in p(i,j)} \delta_e^2,$$

where $p(i, j)$ is the unique shortest path between two nodes i, j.

If the matrix Lambda corresponds to a tree, then we sample from the distribution which determines the max-stable domain of attraction of the distribution in section 6.5.1. The theoretical result has been proved in Asenova, Mazo, and Segers (2021).

6.5.4 The stable tail dependence function (stdf)

The stdf is a key quantity in the topics of multivariate extremal dependence. It is defined as the following limit

$$l(x_v, v \in V) = \lim_{t\to\infty} tP\Big(\bigcup_{v\in V}\{X_v > t/x_v\}\Big), x \in (0,\infty)^{|V|}$$

where X has unit Pareto univariate margins. The relation between the stdf and a max-stable (extreme value) distribution is given by

$$G(x_v, v \in V) = \exp\big\{-l(1/x_v, v \in V)\big\},$$

if G is an extreme value distribution with unit Fréchet margins. More about the stdf can be read in Drees and Huang (1998), Beirlant et al. (2004), and de Haan and Ferreira (2007). The distribution G can have a parametric model such as the Hüsler-Reiss model used throughout the package `gremes`.

The package provides tools for computing parametric and non-parametric stdf for models on trees and block graphs.

The parametric estimate of l when l is parameterized using the Hüsler-Reiss distribution is $l(x_v, v \in V; \hat{\theta}_{n,k})$ in case of tree models and $l(x_v, v \in V; \hat{\delta}_{n,k})$ in case of models on block graphs. The vector $\hat{\theta}_{n,k} = (\hat{\theta}_{e;n,k}, e \in E)$ is any estimate of the edge weights on a tree, and $\hat{\delta}_{n,k} = (\hat{\delta}_{e;n,k}, e \in E)$ is any estimate of the edge weights on a block graph.

The non-parametric estimate of the stdf is given in Drees and Huang (1998).

$$\hat{l}_{n,k}(x_v, v \in V) = \frac{1}{k}\sum_{i=1}^{n} 1\left(\bigcup_{v\in V}\left\{n\hat{F}_{v,n}(X_{v,i}) > n + 1/2 - kx_v\right\}\right),$$

where $\hat{F}_{v,n}(x) = (1/n)\sum_{i=1}^{n} 1(X_{v,i} \le x)$, i.e., the non-parametric estimate of the cumulative distribution function.

The methods which compute stdf are

- `stdf.Network` for non-parametric estimates of the stdf disregarding of the graph, tree of block graph.
- `stdf.HRMtree` for parametric estimates for models on trees.
- `stdf.HRMBG` for parametric estimates for models on block graphs.

Here we have some examples for these.

Non-parametric estimates on an object containing a tree

```
x<- runif(8)
names(x)<- letters[1:8]
tobj<- Tree(seg, XX) # an object of containing block graph and the
    data associated to it
#> From validate.Network: Edges have been assigned names
#> From validate.Network: No latent variables
#> From validate.Network: Edges have been assigned names
#> From validate.Network: No latent variables
stdf(tobj, x, 0.2) # 20% of the upper order statistics
#> [1] 1.035
x<- x[3:8] # with latent variables on nodes "a" and "b"
names(x)<- letters[3:8]
XU<- X[,3:8]
tobjU<- Tree(seg, XU) # an object containing a tree an the data
    associated to it
#> From validate.Network: Edges have been assigned names
#> From validate.Network: There are nodes with latent variables
```

```
#> From validate.Network: Edges have been assigned names
#> From validate.Network: There are nodes with latent variables
stdf(tobjU, x, 0.25)
#> [1] 0.916
x<- matrix(runif(40), 5,8)
colnames(x)<- letters[1:8]
stdf(tobj, x, 0.25) # 25% of the upper order statistics
#> [1] 1.160 1.228 1.000 0.888 1.120
```

Non-parametric estimates on an object containing a block graph

```
x<- runif(7)
names(x)<- letters[1:7]
bgobj<- BlockGraph(bg, XB) # an object containing a tree an the
    data associated to it
#> From validate.Network: Edges have been assigned names
#> From validate.Network: No latent variables
#> From validate.Network: Edges have been assigned names
#> From validate.Network: No latent variables
stdf(bgobj, x, 0.15) # 15% of the upper order statistics
#> [1] 1.786667
```

Parametric estimates on model with respect to a tree

```
x<- c(0, 0.1, 0, 2.5, 0, 1.3, 2.3, 1.5)
names(x)<- letters[1:8]
stdf(hrm, x )
#> From setRootDepSet.RootDepSet: The order of the subset must
    correspond to the
#>         order of its corresponding root
#>     b
#> 3.254016
#> attr(,"error")
#> [1] 3.077146e-43
#> attr(,"msg")
#> [1] "Normal Completion"
```

Parametric estimates on a model with respect to a block graph

```
x<- runif(7)
names(x)<- letters[1:7]
stdf(hrbg, x)
#> From setRootDepSet.RootDepSet: The order of the subset must
    correspond to the
#>         order of its corresponding root
#> From setRootDepSet.RootDepSet: The order of the subset must
    correspond to the
#>         order of its corresponding root
#> From setRootDepSet.RootDepSet: The order of the subset must
    correspond to the
#>         order of its corresponding root
#> From setRootDepSet.RootDepSet: The order of the subset must
    correspond to the
#>         order of its corresponding root
#> From setRootDepSet.RootDepSet: The order of the subset must
    correspond to the
#>         order of its corresponding root
#> From setRootDepSet.RootDepSet: The order of the subset must
    correspond to the
```

```
#>         order of its corresponding root
#> From setRootDepSet.RootDepSet: The order of the subset must
    correspond to the
#>         order of its corresponding root
#>     a
#> 2.162505
#> attr(,"error")
#> [1] 6.044999e-05
#> attr(,"msg")
#> [1] "Normal Completion"
```

6.5.5 Extremal coefficients

We look at extremal coefficients as some tools for post-estimation analysis. A typical analysis includes a comparison between estimates of extremal coefficients using the estimates of the edge weights and non-parametric extremal coefficients.

The extremal coefficient of variables in the set $J \subseteq V$ is given by

$$l_J = l(1_J),$$

where l is the stable tail dependence function and $1_J = (1_{i \in J}, i \in V)$, i.e., a vector of length $|V|$ whose elements are zero or one. An element i will be one if it belongs to J and zero otherwise. The range of a J-variate extremal coefficient is between 1 and $|J|$ with dependence decreasing for large value of the coefficient.

There are three methods for computing extremal coefficients:

- `extrCoeff.Network` can be used to compute non-parametric extremal coefficients for data related to tree or to block graphs. Here the graph it doesn't matter, as the only input in the computation is the data.
- `extrCoeff.HRMtree` for parametric extremal coefficients of models on trees
- `extrCoeff.HRMBG` for parametric extremal coefficients of models on block graphs

Non-parametric extremal coefficients - trees and block graphs Given two objects of classes `BlockGraph` or `Tree` respectively we can compute the bivariate extremal coefficients as follows (k-ratio = 0.2).

```
extrCoeff(bgobj, 0.2)
#>   a    b    c    d    e    f    g
#> a 0 1.25 1.375 1.505 1.555 1.600 1.605
#> b 0 0.00 1.295 1.480 1.565 1.590 1.585
#> c 0 0.00 0.000 1.410 1.455 1.525 1.510
#> d 0 0.00 0.000 0.000 1.485 1.595 1.595
#> e 0 0.00 0.000 0.000 0.000 1.625 1.655
#> f 0 0.00 0.000 0.000 0.000 0.000 1.555
#> g 0 0.00 0.000 0.000 0.000 0.000 0.000
extrCoeff(tobj, 0.2)
#>   a    b    c    d    e    f    g    h
#> a 0 1.07 1.09 1.095 1.130 1.205 1.235 1.250
#> b 0 0.00 1.08 1.100 1.115 1.190 1.260 1.240
#> c 0 0.00 0.00 1.110 1.155 1.180 1.270 1.220
#> d 0 0.00 0.00 0.000 1.145 1.205 1.255 1.255
#> e 0 0.00 0.00 0.000 0.000 1.245 1.265 1.275
#> f 0 0.00 0.00 0.000 0.000 0.000 1.325 1.315
#> g 0 0.00 0.00 0.000 0.000 0.000 0.000 1.350
#> h 0 0.00 0.00 0.000 0.000 0.000 0.000 0.000
```

For an arbitrary dimension of the extremal coefficient, we need to create the vector of coordinates and name it according to the names of the nodes. If we want a four variate extremal coefficient of the variables (X_a, X_c, X_d, X_e) we need to do

```
y<- c(1, 0, 1, 1, 1, 0, 0)
names(y)<- letters[1:7]
extrCoeff(bgobj, 0.25, y)
#> [1] 1.9
extrCoeff(tobj, 0.25, y)
#> [1] 1.26
```

If there are latent variables the bivariate extremal coefficients are computed between only observable variables.

```
# on the tree
extrCoeff(tobjU, 0.2)
#>   c     d     e     f     g     h
#> c 0 1.125 1.135 1.180 1.215 1.235
#> d 0 0.000 1.165 1.195 1.215 1.240
#> e 0 0.000 0.000 1.205 1.255 1.275
#> f 0 0.000 0.000 0.000 1.275 1.265
#> g 0 0.000 0.000 0.000 0.000 1.310
#> h 0 0.000 0.000 0.000 0.000 0.000

# on the block graph
XBU<- XB[, -3]
bgobjU<- BlockGraph(bg, XBU)
#> From validate.Network: Edges have been assigned names
#> From validate.Network: There are nodes with latent variables
#> From validate.Network: Edges have been assigned names
#> From validate.Network: There are nodes with latent variables
extrCoeff(bgobjU, 0.3)
#>   a        b        d        e        f        g
#> a 0 1.206667 1.446667 1.483333 1.563333 1.540000
#> b 0 0.000000 1.446667 1.473333 1.553333 1.500000
#> d 0 0.000000 0.000000 1.443333 1.550000 1.533333
#> e 0 0.000000 0.000000 0.000000 1.593333 1.546667
#> f 0 0.000000 0.000000 0.000000 0.000000 1.493333
#> g 0 0.000000 0.000000 0.000000 0.000000 0.000000
```

Note that for the parameters to be all identifiable on the tree, only variables on nodes a, b can be latent - from section 6.5.1 we see that only these nodes have at least three neighbors. On the block graph only variable on node c can be latent - from section 6.5.3 we see that only node c takes part in three cliques (Asenova and Segers, 2021).

If we want an extremal coefficient of other dimension we need to pass a vector with non-zero coordinates only for observed variables.

```
v<- c(0,0,1,1,0,1,0,1)
names(v)<- letters[1:8]
extrCoeff(tobjU, 0.2, v)
#> [1] 1.42

v<- c(1, 1, 0, 1, 0, 0, 1)
names(v)<- letters[1:7]
extrCoeff(bgobjU, 0.15, v)
#> [1] 2.16
```

Parametric extremal coefficients The method `extrCoeff.HRMtree` is used for models on trees.

```
extrCoeff(hrm) # bivariate
#>  a     b     c     d     e     f     g     h
#> a 0 1.039878 1.079656 1.119235 1.163330 1.201239 1.238977
    1.284146
#> b 0 0.000000 1.089021 1.125633 1.158519 1.197413 1.235823
    1.286697
#> c 0 0.000000 0.000000 1.143065 1.181231 1.215809 1.251150
    1.273661
#> d 0 0.000000 0.000000 0.000000 1.201239 1.232620 1.265478
    1.306198
#> e 0 0.000000 0.000000 0.000000 0.000000 1.251150 1.281568
    1.324294
#> f 0 0.000000 0.000000 0.000000 0.000000 0.000000 1.303842
    1.343254
#> g 0 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
    1.364744
#> h 0 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
    0.000000
v<- c(0,0,1,1,0,1,0,1)
names(v)<- letters[1:8]
extrCoeff(hrm, v) # for a particular set of variables
#> From setRootDepSet.RootDepSet: The order of the subset must
    correspond to the
#>         order of its corresponding root
#>     c
#> 1.518277
#> attr(,"error")
#> [1] 5.488687e-05
#> attr(,"msg")
#> [1] "Normal Completion"
```

The method `extrCoeff.HRMBG` is used for models on block graphs.

```
extrCoeff(hrbg) # bivariate
#>  a     b     c     d     e     f     g
#> a 0 1.24817 1.416118 1.597216 1.628907 1.705734 1.682689
#> b 0 0.00000 1.345279 1.561422 1.597216 1.682689 1.657218
#> c 0 0.00000 0.000000 1.472911 1.520500 1.628907 1.597216
#> d 0 0.00000 0.000000 0.000000 1.561422 1.726678 1.705734
#> e 0 0.00000 0.000000 0.000000 0.000000 1.745787 1.726678
#> f 0 0.00000 0.000000 0.000000 0.000000 0.000000 1.657218
#> g 0 0.00000 0.000000 0.000000 0.000000 0.000000 0.000000
v<- c(0,0,1,1,0,1,0)
names(v)<- letters[1:7]
extrCoeff(hrbg, v) # for a particular set of variables
#> From setRootDepSet.RootDepSet: The order of the subset must
    correspond to the
#>         order of its corresponding root
#> From setRootDepSet.RootDepSet: The order of the subset must
    correspond to the
#>         order of its corresponding root
#> From setRootDepSet.RootDepSet: The order of the subset must
    correspond to the
#>         order of its corresponding root
#>     c
```

```
#> 2.062956
#> attr(,"error")
#> [1] 1e-15
#> attr(,"msg")
#> [1] "Normal Completion"
```

6.5.6 Tail dependence coefficients for models on trees

The tail dependence coefficient (tdc) of a subset $W \subseteq V$ and a node $u \notin W$ is defined as

$$\lim_{t\to\infty} \chi_{W|u}(t) := \lim_{t\to\infty} \Pr(X_W > t \mid X_u > t) \qquad u \notin W, t > 1.$$

The event $\{X_W > t\}$ is to be read as the intersection $\bigcap_{w\in W}\{X_w > t\}$. When X has unit Pareto univariate margins $\Pr(X_u > t) = 1/t$ for $t > 1$, so that

$$\Pr(X_W > t \mid X_u > t) = t \Pr(X_{W\cup u} > t)$$

If we set $\overline{W} := W \cup u$ we have

$$\chi_{\overline{W}} = \lim_{t\to\infty} \chi_{\overline{W}}(t)$$

as a bivariate or multivariate tail dependence coefficient (tdc). In terms of the stable tail dependence function l of X_U, the inclusion–exclusion formula yields

$$\chi_{\overline{W}} = \sum_{i=1}^{|\overline{W}|} (-1)^{i-1} \sum_{J\subseteq\overline{W},|J|=i} l(1_J),$$

where $1_J = (1_{j\in J}, j \in U)$ is a vector of zeroes and ones. Note that $l(1_J)$ is the extremal coefficient as considered above.

The parametric estimate of $\chi_{\overline{W}}$ involves the parametric expressions of the stdf evaluated at a particular estimate of the parameter vector $\theta = (\theta_e, e \in E)$.

Non-parametrically we estimate $\chi_{\overline{W}}(t)$ at $t = n/k$ via

$$\hat{\chi}_{\overline{W}} = \frac{n}{k}\frac{1}{n}\sum_{i=1}^{n} 1\left\{n\hat{F}_{v,n}(X_{v,i}) > n-k,\ \forall v \in \overline{W}\right\}.$$

For the set $\overline{W} = (a, c, d, e)$ we should do

```
v<- c(1,0,1,1,1,0,0,0)
names(v)<- letters[1:8]
suppressMessages(taildepCoeff(hrm, v)) # parametric tdc
#>       a
#> 0.7419965
#> attr(,"error")
#> [1] 0
#> attr(,"msg")
#> [1] "univariate: using pnorm"
taildepCoeff(tobj, 0.2, v) # non-parametric tdc
#> [1] 0.795
```

6.5.7 Confidence intervals for pairwise ECE - models on trees

Confidence intervals can be computed for the edge weights for models on trees if the *pairwise* extremal coefficients estimator is used. This is thanks to the distribution

available of this estimator in Einmahl, Kiriliouk, and Segers (2018). The method that computes the confidence intervals is `confInt.EKS`.

Let $\hat{\theta}_{n,k} = \hat{\theta}^{\mathrm{ECE}}_{n,k}$ denote the pairwise (bivariate) extremal coefficient estimator and let θ_0 denote the true vector of parameters. By Einmahl et al. (2018, Theorem 2) with Ω equal to the identity matrix, the ECE is asymptotically normal,

$$\sqrt{k}(\hat{\theta}_{n,k} - \theta_0) \sim \mathcal{N}_{|E|}\big(0, M(\theta_0)\big), \qquad n \to \infty,$$

The asymptotic covariance matrix takes the form

$$M(\theta_0) = (\dot{L}^\top \dot{L})^{-1} \dot{L}^\top \Sigma_L \dot{L} (\dot{L}^\top \dot{L})^{-1}.$$

The matrices $\dot{L}$ and Σ_L depend on θ_0 and are based on partial derivatives of the stable tail dependence function.

For every k and every $e \in E$, an asymptotic 95% confidence interval for the edge parameter $\theta_{0,e}$ is given by

$$\theta_{0,e} \in \left[\hat{\theta}_{k,n;e} \pm 1.96\sqrt{\{M(\hat{\theta}_{k,n})\}_{ee}/k}\right].$$

For more details on this interval we refer to section A.5 in Asenova, Mazo, and Segers (2021).

In the example below we suppose that the estimates obtained from the pairwise extremal coefficient estimator are given by the sequence $(0.1, 0.2, \ldots, 0.8)$ and the matrix of evaluation points based on pairs is the one based on all pairs. We suppose also that the estimates are based on $k = 150$.

```
# create the matrix of evaluation points
tup<- Tuples()
x<- rep(1, 8)
names(x)<- letters[1:8]
pair<- evalPoints(tup, tobj, x)
# create an object of class EKS with the supposed estimates of the
    parameters
eks<- EKS(seg)
#> From HRMnetwork: Edges have been assigned names
eks<- setParams(eks, seq(0.1, 0.8, 0.1))
#> From setParams.HRMtree: Names have been attributed to the
    vector 'value' in the order corresponding to the order of the
    edges: The fist element has the name of the first edge, the
    second element the name of the second edge, etc.
#> From setParams.HRMtree: The parameters have been attached to
    the edges according to their names
suppressMessages(confInt(eks, pair, 150))
#>       [,1]   [,2]
#> e1 0.04324791 0.1567521
#> e2 0.12486398 0.2751360
#> e3 0.20901358 0.3909864
#> e4 0.29547348 0.5045265
#> e5 0.38320722 0.6167928
#> e6 0.47152312 0.7284769
#> e7 0.55790699 0.8420930
```

When there are latent variables we should use a matrix of evaluation points where all pairs are between observed variables only. Hence we should use this matrix that should have been used in estimation for computing the confidence intervals too.

```
x<- rep(1, 6)
names(x)<- letters[3:8]
pairU<- evalPoints(tup, tobjU, x )
suppressMessages(confInt(eks, pairU, 150))
#>       [,1]   [,2]
#> e1 -0.18515204 0.3851520
#> e2 0.08115924 0.3188408
#> e3 0.19466120 0.4053388
#> e4 0.28500252 0.5149975
#> e5 0.37881600 0.6211840
#> e6 0.46951129 0.7304887
#> e7 0.55860393 0.8413961
```

6.6 Conclusion

The main functionalities in `gremes` are methods that estimate edge weights on Hüsler–Reiss models on trees or block graphs.

For models on trees

- `estimate.MME` provides method of moments estimates, based on the limiting distribution in (L1).
- `estimate.MLE`, `estimate.MLE1`, `estimate.MLE2` provide composite likelihood estimates, based on the limiting distribution in (L1).
- `estimate.EKS`, `estimate.EKS_part` provide estimates based on extremal coefficients. The estimator uses the limiting distribution in (L2).
- `estimate.EngHitz` provides a type of composite likelihood estimate and it is a variation of the cliquewise estimator in Engelke and Hitz (2020). It is based on the limit discussed in (L3).
- `estimate.MMEave` and `estimate.MLEave` apply the method of moments estimator and a composite likelihood estimator and are based on asymptotics presented in an unpublished note Segers (2019). The estimator is based on the limit presented in (L4).

For models on block graphs the available functionality for estimation the method `estimate.HRMBG` provides method of moments estimates based on the limit discussed in section 6.2.2.

Conclusion 7

The current research aims to study extreme value (EV) limits of graphical models with respect to undirected and directed graphs that represent generalizations of trees. The generalizations considered are the block graphs and the trees of transitive tournaments (ttt). The latter ones are directed block graphs which are acyclic. Two major models are discussed in the thesis:

- a random vector which satisfies the undirected global Markov property with respect to a tree or a block graph and which is parameterized block-wise by a family of Hüsler–Reiss distributions, and
- a linear structural equations model (SEM) with unit Fréchet independent factors with respect to a ttt, namely the max- or the sum-linear model.

Apart from EV limits on such models we study the identifiability of the distributional parameters from a certain marginal distribution. The topic on EV limits on such models is inspired from studies on Markov chains and dates back to Smith (1992) and more recently on Markov trees (Segers, 2020b). This thesis shows that results holding for Markov chains and trees can be extended to more complex structures such as Markov block graphs and Bayesian networks with respect to a ttt. The identifiability problem in its particular context is new in the EV literature but it is based on the usual concept of identifiability according to which, for two distinct parameter points in the allowed parameter space, say $\theta_1 \neq \theta_2$, we also have two distinct probability measures, i.e., $\mathbb{P}_{\theta_1} \neq \mathbb{P}_{\theta_2}$.

For a graphical model $(X_v, v \in V)$ on a graph $\mathcal{G} = (V, E)$ the EV limits that this thesis focuses on are represented by its max-domain of attraction, its multivariate Pareto type limit, or the convergence of

$$(X_v/t, v \in V) \mid X_u = t, \qquad t \to \infty \tag{7.1}$$

and/or

$$(X_v/X_u, v \in V) \mid X_u > t, \qquad t \to \infty, \tag{7.2}$$

or also

$$(\ln X_v - \ln X_u, v \in V) \mid X_u > t, \qquad t \to \infty. \tag{7.3}$$

In Chapter 2 it is shown that if the graphical model, say X, is defined by a class of bivariate Hüsler–Reiss copulas, one for every adjacent pair of variables, its extreme value distribution is a Hüsler–Reiss distribution with structured parameter matrix where the ij-th element depends on the path between the two nodes. Also, the limit in (7.3) is a $|V| - 1$-variate Gaussian distribution whose parameters depend on the edge weights along the paths from node u to the rest of the nodes. These limiting distributions are expected in view of the results in Segers (2020b) and Engelke et al. (2014). In Chapter 3 it is shown that the multivariate pareto distribution (MPD) linked

to the EV distribution of X is an extremal graphical model according to the definition in Engelke and Hitz (2020).

Chapter 2 introduces the identifiability problem mentioned earlier. In particular, we come up with a necessary and sufficient criterion such that for a sub-vector X_U of $X = (X_v, v \in V)$ the marginal distribution $\mathbb{P}_{U,\theta}$ determines uniquely the extremal dependence in the complete vector X. We demonstrate the relevance of this with an application on data from the Seine, France. The advantage of establishing such a criterion is that we can compute different metrics of extremal dependence involving nodes where we do not observe the corresponding variable.

Chapter 3 extends results from Segers (2020b) and from Chapter 2 to Markov block graphs by showing that parameters of the EV limits depend on the unique shortest paths between any pair of nodes, the existence of which is a key characteristic of the block graph. Thanks to the model constructed in Chapter 3 we were able to find the expression of the unique parameter matrix of the Hüsler–Reiss extremal graphical model from Engelke and Hitz (2020, Theorem 3 and 4).

Two linear SEMs popular in the EV literature are studied in Chapters 4 and 5. These are the sum- and max-linear graphical models with respect to a certain DAG which represents a directed version of a block graph and which we called a ttt. SEMs or also Bayesian networks are used to model causality with the interpretation that one variable is directly caused by its parents together with an independent factor term. In the EV literature it makes sense to consider heavy tailed factors. The key message of Chapter 4, which focuses on the max-linear models, is that such a model behaves in the tails as a Markov block graph if and only if the ttt has a unique source. The identifiability criterion in such models is different from the one from Chapters 2 and 3 but again easy to check. The problem is worked out on the basis of the angular measure, which is shown to be discrete and of the same form for both models (Einmahl et al., 2012).

Chapter 5 of the thesis is dedicated to the sum-linear model. It presents results which allow the comparison of the two linear SEMs and it demonstrates that despite the obvious differences between the two settings they share many similarities too. For instance, the identifiability criterion is the same although the angular measure of the sum-linear model has atoms and masses which depend on the edge weights in a completely different way than those of the max-linear model. Sum-linear models are more prone to manipulations using classical matrix algebra and conditional independence theory as opposed to tropical algebra and conditional independence theory developed in Améndola et al. (2022).

The topics studied in the framework of max-linear models are new in the EV literature, and it is early or difficult to assess to what extent they are of relevance. This gap could be filled with an application to a real problem, where we model a certain dataset as max-linear with respect to a fixed ttt. This is a problem for further research.

SEMs are interpreted as causal, and the dominating research interest so far is in causal discovery or also graph structure learning from data. When it comes to extremes this means discovering variables which represent direct causes for extreme levels at a certain node. Most applications in the literature are on quantities related to rivers where the outcome of the proposed causal discovery method is compared to the true river network to illustrate the goodness-of-fit of the approach. To enter into the field of active research it might be possible to exploit the properties of a ttt with unique source to search for methods for graph discovery under the constraint that the graph should be a ttt. An interesting question would be also if the Markov property and the factorization of the limiting distribution along unique shortest paths can be used in order to infer some minimal graph structure.

The last chapter of the thesis presents some estimation approaches implemented

in the R package `gremes`. The models to estimate are based on Hüsler–Reiss parameterizations. Some of the estimation methods are presented in Chapter 2 and others are variations of these which are supposed to be computationally faster on a larger graph. Three estimation methods are applied — the MME, CLE and ECE. The results are estimates of the edge weights representing the tail dependence within a Hüsler–Reiss distribution. The MME and ECE are variations of estimators in Engelke et al. (2014) and Einmahl et al. (2018) respectively. The estimators are suitable when there are unobservable variables as long as the identifiability criterion is satisfied. While all these methods estimate the parameters at once, the so called clique-wise estimator proposed in Engelke and Hitz (2020) estimates parameters independently one from another. Their estimator is adapted to models with latent variables in the package. Some comparison of estimation methods is presented in Chapter 2 where it can be seen that the MME and the CLE give very similar results. The highest dimension of the data that the estimation methods have been tested on is 30 variables. The fastest one is the MME as it represents a quadratic optimization problem.

A major gap of the thesis is the lack of practical motivation of the presented models, namely those in Chapters 3, 4 and 5 which treat graphical models with respect to generalizations of trees. In the introduction to the thesis it has been argued that we could consider the latter as approximations to some real networks where clusters are formed and are linked together in a tree-like manner. Recent literature in this direction is Hu et al. (2022). The advantage of the approximations of real networks by trees or generalizations of trees is that the approximation carries some structure, some order, and perhaps more useful information about the underlying network then the network itself.

In the same spirit, a possible line of further research is if the probabilistic results regarding the EV limits can be extended to even larger classes of graphs. Key factors for the special representation of the tail limiting vectors are the unique shortest paths and the fact that all minimal separators are singletons. These properties are lost for a graph which is not a block graph, so we suspect that this is the largest class of graphs for which the results in Chapter 3 hold. For graphs where minimal separators are not singletons we might need to condition on a different event than the separator variable exceeding a high threshold.

Another intriguing research idea would be to study a model according to which the graph is random and growing as the threshold exceeded at some node goes to infinity. Think for instance of an extreme event happening at some location which starts activating different locations where extreme events can be observed too, for instance an earthquake.

Bibliography

K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, 44(2):182–198, 2009.

C. Améndola, B. Hollering, S. Sullivant, and N. Tran. Markov equivalence of max-linear Bayesian networks. In C. de Campos and M. H. Maathuis, editors, *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, volume 161 of *Proceedings of Machine Learning Research*, pages 1746–1755. PMLR, 27–30 Jul 2021.

C. Améndola, C. Klüppelberg, S. Lauritzen, and N. M. Tran. Conditional independence in max-linear Bayesian networks. *The Annals of Applied Probability*, 32(1):1–45, 2022.

P. Asadi, A. C. Davison, and S. Engelke. Extremes on river networks. *The Annals of Applied Statistics*, 9(4):2023–2050, 12 2015.

S. Asenova and J. Segers. Extremes of markov random fields on block graphs: max-stable limits and structured hüsler-reiss distributions, 2021. https://arxiv.org/abs/2112.04847.

S. Asenova, G. Mazo, and J. Segers. Inference on extremal dependence in the domain of attraction of a structured Hüsler–Reiss distribution motivated by a Markov tree with latent variables. *Extremes*, pages 1–40, 2021.

K. Basnayake, D. Mazaud, A. Bemelmans, N. Rouach, E. Korkotian, and D. Holcman. Fast calcium transients in dendritic spines driven by extreme statistics. *PLoS biology*, 17(6):e2006202, 2019.

T. Bedford and R. M. Cooke. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial intelligence*, 32(1):245–268, 2001.

T. Bedford and R. M. Cooke. Vines–a new graphical model for dependent random variables. *The Annals of Statistics*, 30(4):1031–1068, 2002.

A. Behtoei, M. Jannesari, and B. Taeri. A characterization of block graphs. *Discrete Applied Mathematics*, 158(3):219–221, 2010.

J. Beirlant, Y. Goegebeur, J. Segers, J. Teugels, D. De Waal, and C. Ferro. *Statistics of Extremes: Theory and Applications*. Wiley Series in Probability and Statistics. Wiley, New Jersey, 2004.

J. Beirlant, A. Kijko, T. Reynkens, and J. H. Einmahl. Estimating the maximum possible earthquake magnitude using extreme value methodology: the groningen case. *Natural Hazards*, 98(3):1091–1113, 2019.

L. Belzile, J. L. Wadsworth, P. J. Northrop, S. D. Grimshaw, and R. Huser. *mev: Multivariate Extreme Value Distributions*, 2020. URL `https://CRAN.R-project.org/package=mev`. R package version 1.13.1.

G. Boero. Comparing ex-ante forecasts from a sem and var model: An application to the italian economy. *Journal of Forecasting*, 9(1):13–24, 1990.

J. A. Bondy, U. S. R. Murty, et al. *Graph Theory with Applications*, volume 290. Macmillan, London, 1976.

J. Buck and C. Klüppelberg. Recursive max-linear models with propagating noise. *Electronic Journal of Statistics*, 15(2):4770–4822, 2021.

K. Budhathoki, L. Minorics, P. Bloebaum, and D. Janzing. Causal structure-based root cause analysis of outliers. In K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 2357–2369. PMLR, 17–23 Jul 2022.

G. Chen, D. Glen, Z. Saad, J. P. Hamilton, M. Thomason, I. Gotlib, and R. Cox. Vector autoregression, structural equation modeling, and their synthesis in neuroimaging data analysis. *Computers in biology and medicine*, 41:1142–55, 12 2011.

S. G. Coles and J. A. Tawn. Modelling extreme multivariate events. *Journal of the Royal Statistical Society. Series B*, 53(2):377–392, 1991.

A. C. Davison and D. V. Hinkley. *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1997.

L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2007.

L. de Haan and S. I. Resnick. Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 40(4):317–337, 1977.

H. Drees and X. Huang. Best attainable rates of convergence for estimators of the stable tail dependence function. *Journal of Multivariate Analysis*, 64(1):25–46, 1998.

D. Edwards. Hierarchical interaction models. *Journal of the Royal Statistical Society*, 52(1):3–20, 1990.

J. Einmahl and S. Smeets. Ultimate 100m world records through extreme-value theory. *Statistica Neerlandica*, 65(1):32–42, 2011.

J. Einmahl, A. Kiriliouk, and J. Segers. A continuous updating weighted least squares estimator of tail dependence in high dimensions. *Extremes*, 21(2):205–233, 2018.

J. H. J. Einmahl, A. Krajina, and J. Segers. An M-estimator for tail dependence in arbitrary dimensions. *Annals of Statistics*, 40(3):1764–1793, 2012.

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events: for Insurance and Finance*, volume 33. Springer Science & Business Media, Berlin, 2013.

S. Engelke and A. S. Hitz. Graphical models for extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(3):1–38, 2020.

S. Engelke and S. Volgushev. Structure learning for extremal tree models, 2020. `https://arxiv.org/abs/2012.06179v2`.

S. Engelke, Z. Kabluchko, and M. Schlather. An equivalent representation of the brown–resnick process. *Statistics and Probability Letters*, 81(8):1150–1154, 2011.

S. Engelke, A. Malinowski, Z. Kabluchko, and M. Schlather. Estimation of Hüsler-Reiss distributions and Brown-Resnick processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):239–265, 2014.

J. French, P. Kokoszka, S. Stoev, and L. Hall. Quantifying the risk of heat waves using extreme value theory and spatio-temporal functional data. *Computational Statistics and Data Analysis*, 131:176–193, 2019. High-dimensional and functional data analysis.

M. G. Genton, Y. Ma, and H. Sang. On the likelihood function of Gaussian max-stable processes. *Biometrika*, 98(2):481–488, 2011.

N. Gissibl and C. Klüppelberg. Max-linear models on directed acyclic graphs. *Bernoulli*, 24(4A):2693–2720, 2018.

N. Gissibl, C. Klüppelberg, and M. Otto. Tail dependence of recursive max-linear models with regularly varying noise variables. *Econometrics and Statistics*, 6:149–167, 2018.

N. Gissibl, C. Klüppelberg, and S. Lauritzen. Identifiability and estimation of recursive max-linear models. *Scandinavian Journal of Statistics*, 48(1):188–211, 2021.

N. Gnecco, N. Meinshausen, J. Peters, and S. Engelke. Causal discovery in heavy-tailed models. *The Annals of Statistics*, 49(3):1755–1778, 2021.

D. Hand, T. Rudas, and A. Dawid. Discussion of the papers by Edwards, and Wermuth and Lauritzen. *Journal of the Royal Statistical Society*, 52(1):51–72, 1990.

F. Harary and L. Moser. The theory of round Robin tournaments. *The American Mathematical Monthly*, 73(3):231–246, 1966.

J. Heffernan and S. Resnick. Hidden regular variation and the rank transform. *Advances in Applied Probability*, 37(2):393–414, 2005.

J. E. Heffernan and S. I. Resnick. Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, 17(2):537–571, 2007.

J. E. Heffernan and J. A. Tawn. A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):497–546, 2004.

M. Hentschel, S. Engelke, and J. Segers. Statistical inference for hüsler-reiss graphical models through matrix completions. 2022. arXiv:2210.14292v1.

F. Hernandez-Campos, K. Jeffay, C. Park, J. S. Marron, and S. I. Resnick. Extremal dependence: Internet traffic applications. *Stochastic Models*, 21(1):1–35, 2005.

S. Hu, Z. Peng, and J. Segers. Modelling multivariate extreme value distributions via markov trees. *arXiv preprint arXiv:2208.02627*, 2022. https://doi.org/10.48550/arXiv.2208.02627.

X. Huang. Statistics of bivariate extremes. *Tinbergen Institute Research Series*, 22, 1992.

R. Huser and A. C. Davison. Composite likelihood estimation for the Brown–Resnick process. *Biometrika*, 100(2):511–518, 2013.

S. Højsgaard. *gRim: Graphical Interaction Models*, 2017. URL `https://CRAN.R-project.org/package=gRim`. R package version 0.2-0.

J. Hüsler and R.-D. Reiss. Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters*, 7(4):283–286, 1989.

A. Janssen and J. Segers. Markov tail chains. *Journal of Applied Probability*, 51(4): 1133–1153, 2014.

H. T. Kiiveri, T. P. Speed, and J. B. Carlin. Recursive causal models. *Journal of The Australian Mathematical Society*, 36:30–52, 1984.

A. Kiriliouk, J. Segers, and L. Tafakori. An estimator of the stable tail dependence function based on the empirical beta copula. *Extremes*, 22(4):581–600, 2018.

C. Klüppelberg and S. Lauritzen. Bayesian networks for max-linear models. In *Network Science*, pages 79–97. Springer, Cham., 2019.

C. Klüppelberg and M. Krali. Estimating an extreme Bayesian network via scalings. *Journal of Multivariate Analysis*, 181:1046–1072, 2021.

C. Klüppelberg and E. Sönmez. Max-linear models in random environment. *Journal of Multivariate Analysis*, 190:1049–1099, 2022.

D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, Cambridge, US, 2009.

H. J. P. Krupskii. *CopulaModel: CopulaModel: Dependence Modeling with Copulas*, 2014. R package version 0.6.

W. Kulpa. Poincaré and domain invariance theorem. *Acta Univ. Carolin. Math. Phys.*, 39(1-2):127–136, 1998.

S. Lauritzen and N. Wermuth. Graphical models for associations between variables, some of which are qualitative and some quantitative. *The Annals of Statistics*, 17(1): 31–57, 1989.

S. L. Lauritzen. *Graphical Models*. Oxford University Press, Oxford, 1996.

V. B. Le and N. N. Tuy. The square of a block graph. *Discrete Mathematics*, 310(4): 734–741, 2010.

A. W. Ledford and J. A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187, 1996.

A. W. Ledford and J. A. Tawn. Modelling dependence within joint tail regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(2):475–499, 1997.

D. Lee and H. Joe. Multivariate extreme value copulas with factor and tree dependence structures. *Extremes*, 21:1–30, 06 2017.

M. Manera. Modelling factor demands with sem and var: an empirical comparison. *Journal of Productivity Analysis*, 26:121–146, 02 2006.

A. K. Nikoloulopoulos, H. Joe, and H. Li. Extreme value properties of multivariate t copulas. *Extremes*, 12(2):129–148, 2009.

OECD. Preventing the flooding of the seine in the paris – Île-de-france region, 2018. https://www.oecd.org/gov/risk/preventing-the-flooding-of-the-seine-2018.pdf.

F. Orsini, G. Gecchele, M. Gastaldi, and R. Rossi. Collision prediction in roundabouts: a comparative study of extreme value theory approaches. *Transportmetrica A: Transport Science*, 15(2):556–572, 2019.

I. Papastathopoulos and K. Strokorb. Conditional independence among max-stable laws. *Statistics & Probability Letters*, 108:9–15, 2016.

I. Papastathopoulos and J. A. Tawn. Hidden tail chains and recurrence equations for dependence parameters associated with extremes of higher-order markov chains, 2019.

I. Papastathopoulos, K. Strokorb, J. A. Tawn, and A. Butler. xtreme events of Markov chains. *Advances in Applied Probability*, 49(1):134–161, 2017.

O. C. Pasche, V. Chavez-Demoulin, and A. C. Davison. Causal modelling of heavy-tailed variables and confounders with application to river flow, 2021. `https://arxiv.org/abs/2110.06686v1`.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, UK, 2000.

R. Perfekt. Extremal Behaviour of Stationary Markov Chains with Applications. *The Annals of Applied Probability*, 4(2):529–548, 1994.

V. Pisarenko, A. Sornette, D. Sornette, and M. Rodkin. Characterization of the tail of the distribution of earthquake magnitudes by combining the gev and gpd descriptions of extreme value theory. *Pure and Applied Geophysics*, 171(8):1599–1624, 2014.

S. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer-Verlag, Berlin, 1987.

S. Resnick. Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5:303–336, 2002.

S. I. Resnick. Discussion of the danish data on large fire insurance losses. *ASTIN Bulletin*, 27(1):139–151, 1997.

S. I. Resnick and D. Zeber. Asymptotics of Markov kernels and the tail chain. *Advances in Applied Probability*, 45(1):186–213, 2013.

R. T. Rockafellar and R. J.-B. Wets. *Variational Analysis*, volume 317. Springer-Verlag, Berlin, 1998.

H. Rootzén and N. Tajvidi. Multivariate generalized Pareto distributions. *Bernoulli*, 12(5):917–930, 2006.

H. Rootzén, J. Segers, and J. L. Wadsworth. Multivariate generalized Pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis*, 165:117–131, 2018.

J. Segers. Multivariate regular variation of heavy-tailed Markov chains, 2007. `https://arxiv.org/abs/math/0701411v1`.

J. Segers. On the property of the domain of attraction of the simple huesler-reiss distribution: Lognormal limit when conditioning on the geometric mean being large, 2019. johan.segers@uclouvain.be.

J. Segers. Conditional independence within multivariate Hüsler–Reiss Pareto distributions. Discussion of "Graphical Models for Extremes" by S. Engelke and A.S. Hitz, 2020a.

J. Segers. One-versus multi-component regular variation and extremes of Markov trees. *Advances in Applied Probability*, 52(3):855–878, 2020b.

Y. Shen, G. B. Giannakis, and B. Baingana. Nonlinear structural vector autoregressive models with application to directed brain networks. *IEEE Transactions on Signal Processing*, 67(20):5325–5339, 2019.

N. Shephard and A. Rambachan. Econometric analysis of potential outcomes time series: instruments, shocks, linearity and the causal response function. 2020. arXiv:1903.01637v3.

R. L. Smith. The extremal index for a Markov chain. *Journal of Applied Probability*, 29 (1):37–45, 1992.

H. Southworth. Predicting potential liver toxicity from phase 2 data: a case study with ximelagatran. *Statistics in Medicine*, 33(17):2914–2923, 2014.

P. Spirtes. Conditional independence in directed cyclical graphical models representing feedback or mixtures. 1994.

P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*, volume 81. 1993.

K. Strokorb. Extremal independence old and new, 2020.

N. M. Tran, J. Buck, and C. Klüppelberg. Estimating a latent tree for extremes. 2021a. `http://128.84.4.18/abs/2102.06197v2`.

N. M. Tran, J. Buck, and C. Klüppelberg. Causal discovery of a river network from its extremes, 2021b. `https://arxiv.org/abs/2102.06197v2`.

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998.

M. J. Wainwright, M. I. Jordan, et al. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.

S. Wright. The method of path coefficients. *The Annals of Mathematical Statistics*, 5 (3):161–215, 1934.

H. Yu, W. I. T. Uy, and J. Dauwels. Modeling spatial extremes via ensemble-of-trees of pairwise copulas. *IEEE Transactions on Signal Processing*, 65(3):571–586, 2016.

S. Yun. The extremal index of a higher-order stationary Markov chain. *Annals of Applied Probability*, pages 408–437, 1998.

R. Zhao, X. He, and J. Wang. Learning linear non-gaussian directed acyclic graph with diverging number of nodes, 2021.

Éditions OCDE. Étude de l'OCDE sur la gestion des risques d'inondation la Seine en Île-de-France, 2014. https://doi.org/10.1787/9789264207929-fr.