EXTREMES OF MARKOV RANDOM FIELDS ON BLOCK GRAPHS

Stefka Asenova, Johan Segers

LIDAM Discussion Paper ISBA
2022 / 13

ISBA
Voie du Roman Pays 20 - L1.04.01
B-1348 Louvain-la-Neuve
Email : lidam-library@uclouvain.be
https://uclouvain.be/en/research-institutes/lidam/isba/publication.html

Extremes of Markov random fields on block graphs

Stefka Asenova* Johan Segers†

December 10, 2021

Abstract

We study the joint occurrence of large values of a Markov random field or undirected graphical model associated to a block graph. On such graphs, containing trees as special cases, we aim to generalize recent results for extremes of Markov trees. Every pair of nodes in a block graph is connected by a unique shortest path. These paths are shown to determine the limiting distribution of the properly rescaled random field given that a fixed variable exceeds a high threshold.

When the sub-vectors induced by the blocks follow Hüsler–Reiss extreme value copulas, the global Markov property of the original field induces a particular structure on the parameter matrix of the limiting max-stable Hüsler–Reiss distribution. The multivariate Pareto version of the latter turns out to be an extremal graphical model according to the original block graph. Moreover, thanks to these algebraic relations, the parameters are still identifiable even if some variables are latent.

Keywords — Markov random field; graphical model; block graph; multivariate extremes; tail dependence; latent variable; Hüsler–Reiss distribution; conditional independence

1 Introduction

Graphical models are statistical models for random vectors whose components are associated to the nodes of a graph, where 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 behaviour 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) show that conditional independence implies unconditional independence. At least for absolutely continuous max-stable distributions, this means that there is no point in studying Markov random fields. Note that the max-stable graphical models in Gissibl and Klüppelberg (2018) and Améndola et al. (2020) concern directed acyclic graphs as well as max-linear, hence singular, distributions. In our paper, conditional independence relations are induced by separation properties in undirected graphs, not by parent-child relations in directed ones.

*Corresponding author. UCLouvain, LIDAM/ISBA, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium. E-mail: stefka.asenova@uclouvain.be

†UCLouvain, LIDAM/ISBA, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium. E-mail: johan.segers@uclouvain.be

Instead, Markov random fields with continuous and positive densities have made their way in extreme value analysis extremes 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 this paper 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 certain variable exceeds a high threshold. As a prime example, we consider the case where the random vectors induced by the graph's cliques have themselves a max-stable distribution. Note that such a Markov random field is itself not max-stable, due to the earlier cited result by Papastathopoulos and Strokorb (2016).

Our main result, Theorem 3.5, is inspired by the one about Markov random fields on connected trees, or Markov trees in short, in Segers (2020). 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. In the present paper we show a similar result for Markov random fields with respect to connected block graphs. 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) and Perfekt (1994). The assumptions that our theorem shares with these and subsequent papers include existence of tail limits of the transition kernels together with a regularity condition designed to exclude processes which can become extreme again after reaching non-extreme levels; see Resnick and Zeber (2013) for an in-depth discussion.

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 4, we study the implications of our main convergence result for a Markov block graph in which the joint law on each clique is the max-stable Hüsler–Reiss distribution (Hüsler and Reiss, 1989). This model generalizes the Hüsler–Reiss Markov tree in Asenova et al. (2021) used to study extremes of a river network. Note again that by Papastathopoulos and Strokorb (2016), the joint law of such a Markov block graph is itself not a Hüsler–Reiss max-stable distribution.

In Proposition 4.2, we show that for a Markov block graph with clique-wise Hüsler–Reiss max-stable distributions, the tail limits in Theorem 3.5 are all multivariate log-normal. In Engelke et al. (2015), such log-normal limits were found to characterize the domain of attraction of the Hüsler–Reiss max-stable distribution. For our Hüsler–Reiss Markov block graphs, Proposition 4.4 states that the parameter matrix of the max-stable limit has an explicit and elegant form determined by the path sums in the block graph, generalizing a structure found for Markov trees in Segers (2020) and Asenova et al. (2021). By Proposition 4.5, the multivariate Pareto distribution associated to the Hüsler–Reiss max-stable distribution is an extremal graphical model with respect to the same graph. Incidentally, the result completes Proposition 4 in Engelke and Hitz (2020), which states existence and uniqueness of a Hüsler–Reiss parameter matrix inducing an extremal graphical model on a given block graph and with given clique distributions, but without explicit formula. Finally, thanks to the structure of the parameter matrix, it is proved in Proposition 4.6 that all parameters of the limiting Hüsler–Reiss max-stable distribution remain identifiable even when some of the variables are latent, as long as these variables lie on nodes with neighbours in at least three different cliques. This generalizes a similar identifiability claim for trees in Asenova et al. (2021).

Our perspective is a typical one in extreme value analysis. The point of departure is a distribution that has a certain structure not related to extreme values, in this case a graphical model in the usual sense. Then we study the tail of such a random vector: we verify that the limiting distribution of the scaled vector given that some component exceeds a high threshold is an extremal graphical model with respect to the same graph. This property provides the justification of calibrating the limit model to the extremes in a data-set and to use the fitted model as a basis for extrapolation beyond the data range. Our contribution thus adds an essential complement to the work in Engelke and Hitz (2020), who study the limit distribution but not its domain of attraction. 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 et al. (2021, Section A.2) for a discussion.

The outline of the paper is as follows. In the preliminary Section 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. Section 4 concerns a Markov block graph with max-stable Hüsler–Reiss clique distributions and presents the limiting field, the limiting max-stable distribution, the relation to extremal graphical models and the identifiability criterion in case of latent variables. The conclusion in Section 5 summarizes the main points of the paper. Proofs are deferred to Appendices A and B.

2 Preliminaries

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 $(a, b) \in E$ 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.

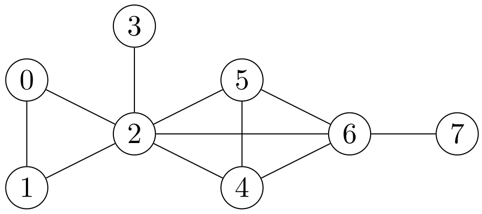


Figure 1: An example of a block graph. There are four blocks or (maximal) cliques, $\mathcal{C} = \big\{\{0,1,2\},\{2,3\},\{2,4,5,6\},\{6,7\}\big\}$, as well as two minimal separators, $\mathcal{S} = \big\{\{2\},\{6\}\big\}$. The unique shortest path from 7 to 0 has edge set $(7 \rightsquigarrow 0) =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

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 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 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.

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

$$\forall x \in (0,\infty)^V, \qquad \lim_{n\to\infty} F^n(nz) = G(z), \tag{1}$$

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 (1) then means that

$$M^{(n)}/n \xrightarrow{d} G, \qquad n \to \infty, \tag{2}$$

the arrow $\xrightarrow{d}$ signifying convergence in distribution. The choice of the scaling sequence n in (1) and (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

$$G(z) = \exp\left(-\mu\left[\left\{x \in [0,\infty)^V : \exists v \in V, x_v > z_v\right\}\right]\right), \qquad z \in (0,\infty]^V, \tag{3}$$

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

$$\ell(y) := -\ln G(1/y_v, v \in V) = \mu\left[\left\{x \in [0,\infty)^V : \exists v \in V, x_v > 1/y_v\right\}\right] \tag{4}$$

is known as the stable tail dependence function (stdf) (Drees and Huang, 1998). In terms of the copula K of the original distribution F, the stdf is given by

$$\ell(y) = \lim_{t\to\infty} t\left\{1 - K(1 - y_v/t, v \in V)\right\} = \lim_{t\to\infty} t\,\mathbb{P}\left(\exists v \in V, X_v > t/y_v\right). \tag{5}$$

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

We can replace the integer n in (1) by the real scalar $t > 0$: the condition $F \in D(G)$ is equivalent to

$$\forall z \in (0,\infty)^V, \qquad \lim_{t\to\infty} F^t(tz) = G(z). \tag{6}$$

By a direct calculation starting from (6) it follows that

$$\lim_{t\to\infty} \mathbb{P}\left(\forall v \in V, X_v/t \leq z_v \;\middle|\; \max_{v\in V} X_v > t\right) = \frac{\ln G(\min(z_v, 1), v \in V) - \ln G(z)}{\ln G(1, \ldots, 1)}, \tag{7}$$

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

$$\left(t^{-1}X \;\middle|\; \max_{v\in V} X_v > t\right) \xrightarrow{d} Y, \qquad t \to \infty, \tag{8}$$

where $Y = (Y_v, v \in V)$ is a random vector whose cumulative distribution function is equal to the right-hand side in (7). 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 thresholds; see also Rootzén et al. (2018).

3 Tails of Markov 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 3.5, 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.

We will be making two assumptions on the conditional distribution of X at high levels. Assumption 3.1 is the main one, as it will determine the limit distribution in Theorem 3.5 through the construction in Definition 3.2 below. For a set A and an element $b \in A$, we write $A \setminus b$ rather than $A \setminus \{b\}$.

Assumption 3.1. *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*

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

If the distribution of the clique vector X_C is max-stable, the limit $\nu_{C,s}$ can be calculated by means of Proposition 3.9.

Definition 3.2 (Increments). *Under Assumption 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 3.1, 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 3.2 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 ranges over $\mathcal{C}$. 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. Consider a Markov random field on the block graph in Figure 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	$\{6\}$

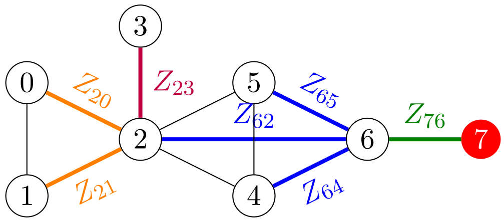
Note that the union over the sets $C \setminus s$ is equal to $\{0,1,\ldots,6\} = V \setminus u$. Assumption 3.1 requires certain joint conditional distributions to converge weakly: as $t \to \infty$, we have

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

The random vector Z in Definition 3.2, step (Z3), is a 7-dimensional random vector whose joint distribution is equal to the product of the above four distributions:

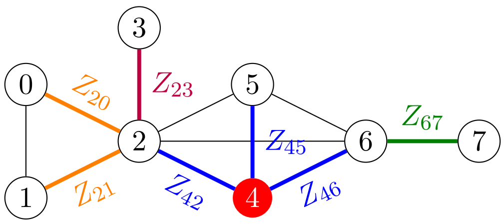
$$Z := (Z_{20}, Z_{21}; Z_{23}; Z_{62}, Z_{64}, Z_{65}; Z_{76}) \sim \nu_{\{0,1,2\},2} \otimes \nu_{\{2,3\},2} \otimes \nu_{\{2,4,5,6\},6} \otimes \nu_{\{6,7\},7}. \tag{9}$$

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 (9). 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 3.1. $\diamondsuit$

In Assumption 3.1, let $\nu_{C,s}^{v}$ 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 3.4. *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*

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

Assumption 3.4 excludes processes which can become extreme again after reaching non-extreme levels; for Markov chains, see Resnick and Zeber (2013). It will imply that in the limit, zero is an absorbing state: if, starting from node u, zero is reached at a separator node s, it persists throughout the graph after that node.

Theorem 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 3.1 and 3.4 be satisfied. For a given $u \in V$, let Z be the random vector in Definition 3.2. Then as $t \to \infty$, we have*

$$\left(\frac{X_v}{t}, v \in V \setminus u \,\middle|\, X_u = t\right) \xrightarrow{d} (A_{uv}, v \in V \setminus u) =: A_{u,V\setminus u} \quad \text{where} \quad A_{uv} := \prod_{e \in (u \rightsquigarrow v)} Z_e. \tag{10}$$

Remark 3.6. If the block graph is a tree, Theorem 3.5 reduces to Theorem 1 in Segers (2020). 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 (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 3.7. We continue with Example 3.3. Let the variable exceeding a high threshold be the one on node $u = 7$. The conclusion of Theorem 3.5 is that as $t \to \infty$, we have

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

where, in the notation of Example 1, 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 3.6). ◇

Remark 3.8. A useful result that we will need further on is that the convergence in (10) implies a self-scaling property of the Markov random field at high levels: as $t \to \infty$, we have

$$\left(\frac{X_v}{X_u}, v \in V \setminus u \,\middle|\, X_u > t\right) \xrightarrow{d} (A_{u,v}, v \in V \setminus u). \tag{11}$$

The proof of (11) is the same as the one of Corollary 1 in Segers (2020).

The limit distribution in Theorem 3.5 is determined by the graph structure and the clique-wise limit distributions $\nu_{C,s}$ in Assumption 3.1 via Definition 3.2. The next result provides those limits $\nu_{C,s}$ for certain max-stable distributions.

Proposition 3.9. *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

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

4 Markov random field with Hüsler–Reiss cliques

We will apply Theorem 3.5 to a Markov random field X with respect to a (connected) block graph $\mathcal{G} = (V, E)$ such that for every (maximal) clique $C \in \mathcal{C}$, the sub-vector $X_C = (X_v, v \in C)$ has a max-stable Hüsler–Reiss distribution with unit-Fréchet margins. Note that the joint distribution of X is then itself not Hüsler–Reiss, unless $\mathcal{G}$ consists of a single clique: indeed, absolutely continuous max-stable distributions cannot satisfy conditional independence relations except for unconditional independence (Papastathopoulos and Strokorb, 2016).

In Proposition 4.2, we find that the limit random vector A_u in Theorem 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 Hüsler–Reiss max-stable distribution whose parameter matrix can be derived in a simple way from those of the Hüsler–Reiss distributions of the sub-vectors X_C (Proposition 4.4). Further, the multivariate Pareto distribution associated to this joint Hüsler–Reiss max-stable 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 4.5). Finally, the parameters of the limiting Hüsler–Reiss max-stable 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 4.6). The proofs of the results in this section are given in Appendix B.

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 (4) the stdf ℓ of a general max-stable distribution distribution G.

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

$$\ell_\delta(x, y) = x \, \Phi\left(\delta + \frac{\ln(x/y)}{2\delta}\right) + y \, \Phi\left(\delta + \frac{\ln(y/x)}{2\delta}\right), \qquad (x, y) \in (0, \infty)^2, \tag{12}$$

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 3.9 can be calculated explicitly and is equal to the one of the lognormal random variable $\exp\{2\delta(Z-\delta)\}$, with Z a standard normal random variable (Segers, 2020, Example 4).

To introduce the multivariate Hüsler–Reiss distribution, we follow the exposition in Engelke et al. (2015). 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:

$$\lim_{n\to\infty} (1-\rho_{ij}(n)) \ln(n) = \delta_{ij}^2, \qquad i,j \in W. \tag{13}$$

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 (13).] 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

$$\left(\Psi_{J,s}(\Delta)\right)_{i,j} = 2(\delta_{si}^2 + \delta_{sj}^2 - \delta_{ij}^2), \qquad i,j \in J \setminus s. \tag{14}$$

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

$$H_\Delta(x) = \exp\left\{ \sum_{j=1}^{|W|} (-1)^j \sum_{J\subseteq W : |J| = j} h_{\Delta,J}(x_J) \right\}, \qquad x \in (0,\infty)^W, \tag{15}$$

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 (14).

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:

$$H_\Delta(x) = \exp\left\{ - \sum_{s\in W} \frac{1}{x_s} \Phi_{|W|-1}\left(2\delta_{vs}^2 + \ln(x_v/x_s), v \in W\setminus s; \Psi_{W,s}(\Delta)\right) \right\}, \qquad x \in (0,\infty)^W,$$

with $\Phi_d(\,\cdot\,;\Sigma)$ the d-variate normal cdf with covariance matrix Σ. The stdf is thus

$$\ell_\Delta(y) = \sum_{s\in W} y_s\, \Phi_{|W|-1}\left(2\delta_{vs}^2 + \ln(y_s/y_v), v \in W\setminus s; \Psi_{W,s}(\Delta)\right), \qquad y \in (0,\infty)^W.$$

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 (12).

4.2 Hüsler–Reiss Markov field and its tail

Assumption 4.1 (Hüsler–Reiss Markov field)**.** *Let $\mathcal{G} = (V, E)$ be a (connected) block graph with (maximal) cliques $\mathcal{C}$. 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. Let X be a Markov random field with respect to $\mathcal{G}$ such that the distribution of every sub-vector X_C is equal to the max-stable Hüsler–Reiss distribution with parameter matrix Δ_C and unit-Fréchet margins.*

To show that such a Markov field indeed exists, argue as follows. As a block graph is decomposable, there is an enumeration $C_1, \ldots, C_m$ of the cliques in $\mathcal{C}$ that respects the running intersection property (Lauritzen, 1996, Chapter 2): for any $i \in \{2, \ldots, m\}$ there exists $k(i) \in \{1, \ldots, i-1\}$ such that $D_i := C_i \cap (C_1 \cup \cdots \cup C_{i-1})$ is contained in $C_{k(i)}$. Since $\mathcal{G}$ is a block graph, every such intersection D_i is a singleton, $\{v_i\}$, say. In Assumption 4.1, if h_C is the probability density function of the max-stable Hüsler–Reiss distribution with parameter matrix Δ_C, then the probability density function of the Hüsler–Reiss Markov field X is

$$h(x) = h_{C_1}(x_{C_1}) \prod_{i=2}^{m} h_{C_i \setminus v_i | v_i}(x_{C_i \setminus v_i} | x_{v_i}).$$

The property that the above function integrates to one and thus defines a valid probability density function follows by recursion on m. By the Hammersley–Clifford theorem, h is then indeed the probability density function of a Markov random field with respect to $\mathcal{G}$. Moreover, the clique-wise marginal densities of h must be h_C for $C \in \mathcal{C}$. As the univariate distributions are all unit-Fréchet, the distributions of the variables X_v on the separator nodes v are consistently specified, that is, independently of the block C including v.

As mentioned before, by Papastathopoulos and Strokorb (2016, Theorem 1), the random field X in Assumption 4.1 is itself not max-stable, unless the graph $\mathcal{G}$ is complete and thus has only a single clique. We apply Theorem 3.5 to study the limit of 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

$$p_{ij}(\Delta) := \sum_{e \in (i \rightsquigarrow j)} \delta_e^2, \tag{16}$$

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

Proposition 4.2 (Logarithm of the limiting field)**.** *Under Assumption 4.1, we have, for each $u \in V$ and as $t \to \infty$,*

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

with mean vector and covariance matrix written in terms of $p_{ij} = p_{ij}(\Delta)$ in (16) *by*

$$\big(\mu_u(\Delta)\big)_i = -2p_{ui}, \qquad i \in V \setminus u, \tag{17}$$

$$\big(\Sigma_u(\Delta)\big)_{i,j} = 2(p_{ui} + p_{uj} - p_{ij}), \qquad i, j \in V \setminus u, \tag{18}$$

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 4.3. Consider the Hüsler–Reiss Markov field with respect to the block graph in Figure 2. The graph has three cliques, to which correspond the parameter matrices

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

If a high threshold is exceeded at node $u = 1$, the limiting 5-variate normal distribution in Proposition 4.2 has means $(\mu_1(\Delta))_i$ and variances $(\Sigma_1(\Delta))_{ii}$ proportional to the path sums

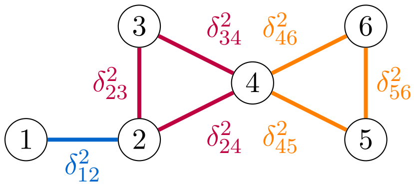


Figure 2: Parameters of a 6-variate Hüsler–Reiss Markov field (Assumption 4.1) with respect to a block graph with three cliques. The non-zero entries of the symmetric matrices $\Delta_1, \Delta_2, \Delta_3$ are associated to the edges of the graph.

$p_{1i} = \sum_{e\in(1\rightsquigarrow i)} \delta_e^2$ for $i \in \{2,\ldots,5\}$, while the off-diagonal entries of the covariance matrix are given by

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

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

The exact distribution of the Hüsler–Reiss Markov field in Assumption 4.1 is, in general, not a Hüsler–Reiss max-stable distribution, but, still, it is in the max-domain of attraction of such a distribution.

Proposition 4.4 (Max-domain of attraction of the Hüsler–Reiss Markov field)**.** *The Hüsler–Reiss Markov random field X in Assumption 4.1 is in the max-domain of attraction of the Hüsler–Reiss max-stable distribution* (15) *with unit-Fréchet margins and parameter matrix $P(\Delta)$ in* (16)*, that is,*

$$\lim_{n\to\infty} \left(\mathbb{P}\left(\forall v\in V: X_v \le n x_v\right)\right)^n = H_{P(\Delta)}(x), \qquad x\in(0,\infty)^V. \tag{19}$$

Recall from Section 2.2 the link between max-domains of attractions on the one hand and weak convergence of high-threshold excesses to multivariate Pareto distributions on the other hand. Because X belongs to the max-domain of attraction of a max-stable Hüsler–Reiss law, the asymptotic distribution of the vector of high-threshold excesses is the Hüsler–Reiss Pareto distribution, which is the distribution of the random vector Y in the next proposition and which is called a Hüsler–Reiss Pareto distribution in Engelke and Hitz (2020). The latter turns out to be an extremal graphical model in the sense of Engelke and Hitz (2020, Definitions 1 and 2). The extremal graphical model with absolutely continuous distribution is defined on the multivariate Pareto random vector in (8) with distribution given in (7). For $u \in V$, let $Y^{(u)}$ be a random vector equal in distribution to $Y \mid Y_u > 1$. According to Engelke and Hitz (2020, Definition 2), Y is an extremal graphical model with respect to a graph $\mathcal{G} = (V,E)$ if we have $Y_i^{(u)} \perp\!\!\!\perp Y_j^{(u)} \mid Y^{(u)}_{V\setminus\{u,i,j\}}$ for all $i,j \in V\setminus u$ such that $(i,j)\notin E$.

Proposition 4.5 (Extremal graphical model)**.** *The Hüsler–Reiss Markov random field X in Assumption 4.1 satisfies the weak convergence relation* (8) *with Y being distributed as in* (7) *for G equal to $H_{P(\Delta)}$, the limit in Proposition 4.4. This Y is an extremal graphical model with respect to $\mathcal{G}$ in the sense of Engelke and Hitz (2020, Definition 2).*

This leads to the elegant result that the graphical model X obtained by endowing every clique C of a block graph $\mathcal{G}$ by a 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 Hüsler–Reiss distributions is an extremal graphical model constructed clique-wise by Hüsler–Reiss Pareto distributions.

Proposition 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 (16).

4.3 Latent variables and parameter identifiability

In Asenova et al. (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 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 4.4 and let $H_{P(\Delta)}(x)$ be the $|V|$-variate Hüsler–Reiss distribution with parameter matrix $P(\Delta)$ in (16). 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 4.6 (Identifiability)**.** *Given the block graph $\mathcal{G} = (V, E)$ and a node set $U \subset V$, the Hüsler–Reiss parameter matrix $P(\Delta)$ in* (16) *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 $v \in V \setminus U$.*

Example 4.7. Consider the block graph $\mathcal{G} = (V, E)$ in Figure 3 with $V = \{1, \ldots, 7\}$ and $U = V \setminus \{3\}$. Node $v = 3$ belongs to three different cliques and thus has clique degree $\mathrm{cd}(v) = 3$. By Proposition 4.6, all edge parameters δ_e^2 for $e \in E$ can be identified from the path sums p_{ij} for $i, j \in U = V \setminus 3$. Indeed, for the edges $e = (a, b) \in \{(1, 2), (4, 5), (6, 7)\}$ this follows from the identity $\delta_e^2 = p_{ab}$, while for the edges $\delta_{i,3}^2$ for $i \neq 3$ this follows from a calculation such as

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

If, however, node $v = 1$ would not belong to U, then the edge parameters δ_{12}^2 and δ_{13}^2 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 $(1, 3)$.

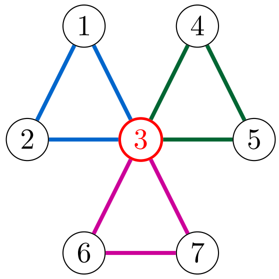


Figure 3: A block graph with three cliques. In the first clique with node set $C_1 = \{1, 2, 3\}$ the parameters are $\delta^2_{12}, \delta^2_{13}, \delta^2_{23}$, in the second clique $C_1 = \{3, 4, 5\}$ the parameters are $\delta^2_{34}, \delta^2_{35}, \delta^2_{45}$, and in the third clique $C_1 = \{3, 6, 7\}$ the parameters are $\delta^2_{36}, \delta^2_{37}, \delta^2_{67}$. These nine parameters determine the Hüsler–Reiss parameter matrix $P(\Delta)$ in (16). The nine parameters and thus the entire matrix $P(\Delta)$ is identifiable from the submatrix $P(\Delta)_U$ with $U = V \setminus \{3\}$ because node 3 belongs to three different cliques (Proposition 4.6 and Example 4.7).

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 (2020). 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.

We then focused on a particular random field with respect to a block graph, namely one for which the distribution on each clique is max-stable Hüsler–Reiss. 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 shows up in the parameter matrix of the max-stable Hüsler–Reiss distribution to which the Markov field is attracted. We show the relation between the original Markov field as an ordinary graphical model to the Hüsler–Reiss extremal graphical models in Engelke and Hitz (2020). Finally, we show that, 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, and what this means for specific parametric families.

Acknowledgements

Stefka Asenova wishes to thank Antonio Abruzzo for his much appreciated help and suggestions on graphical models. Johan Segers wishes to thank Sebastien Engelke and Manuel Hentschel for inspiring discussions on Hüsler–Reiss distributions.

Data availability statement

There is no data used in this manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Améndola, C., Klüppelberg, C., Lauritzen, S., and Tran, N. (2020). Conditional independence in max-linear Bayesian networks. `https://arxiv.org/abs/2002.09233`.

Asenova, S., Mazo, G., and Segers, J. (2021). 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.

Behtoei, A., Jannesari, M., and Taeri, B. (2010). A characterization of block graphs. *Discrete Applied Mathematics*, 158(3):219–221.

Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., De Waal, D., and Ferro, C. (2004). *Statistics of Extremes: Theory and Applications*. Wiley, Chichester.

de Haan, L. and Ferreira, A. (2007). *Extreme Value Theory: An Introduction*. Springer Series in Operations Research and Financial Engineering. Springer, New York.

de Haan, L. and Resnick, S. I. (1977). Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 40(4):317–337.

Drees, H. and Huang, X. (1998). Best attainable rates of convergence for estimators of the stable tail dependence function. *Journal of Multivariate Analysis*, 64(1):25–46.

Engelke, S. and Hitz, A. S. (2020). Graphical models for extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(3):1–38.

Engelke, S., Malinowski, A., Kabluchko, Z., and Schlather, M. (2015). Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):239–265.

Genton, M. G., Ma, Y., and Sang, H. (2011). On the likelihood function of Gaussian max-stable processes. *Biometrika*, 98(2):481–488.

Gissibl, N. and Klüppelberg, C. (2018). Max-linear models on directed acyclic graphs. *Bernoulli*, 24(4A):2693–2720.

Hüsler, J. and Reiss, R.-D. (1989). Maxima of normal random vectors: Between independence and complete dependence. *Statistics & Probability Letters*, 7(4):283–286.

Huser, R. and Davison, A. C. (2013). Composite likelihood estimation for the Brown–Resnick process. *Biometrika*, 100(2):511–518.

Lauritzen, S. L. (1996). *Graphical Models*. Oxford University Press, Oxford.

Le, V. B. and Tuy, N. N. (2010). The square of a block graph. *Discrete Mathematics*, 310(4):734–741.

Lee, D. and Joe, H. (2017). Multivariate extreme value copulas with factor and tree dependence structures. *Extremes*, 21:1–30.

Nikoloulopoulos, A. K., Joe, H., and Li, H. (2009). Extreme value properties of multivariate t copulas. *Extremes*, 12(2):129–148.

Papastathopoulos, I. and Strokorb, K. (2016). Conditional independence among max-stable laws. *Statistics & Probability Letters*, 108:9–15.

Perfekt, R. (1994). Extremal behaviour of stationary Markov chains with applications. *The Annals of Applied Probability*, 4(2):529–548.

Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes.* Springer-Verlag, New York.

Resnick, S. I. and Zeber, D. (2013). Asymptotics of Markov kernels and the tail chain. *Advances in Applied Probability*, 45(1):186–213.

Rockafellar, R. T. and Wets, R. J.-B. (1998). *Variational Analysis*, volume 317. Springer-Verlag, Berlin.

Rootzén, H., Segers, J., and Wadsworth, J. L. (2018). Multivariate generalized Pareto distributions: Parametrizations, representations, and properties. *Journal of Multivariate Analysis*, 165:117–131.

Rootzén, H. and Tajvidi, N. (2006). Multivariate generalized Pareto distributions. *Bernoulli*, 12(5):917–930.

Segers, J. (2020). One- versus multi-component regular variation and extremes of Markov trees. *Advances in Applied Probability*, 52:855–878.

Smith, R. L. (1992). The extremal index for a Markov chain. *Journal of Applied Probability*, 29(1):37–45.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

A Proofs for Section 3

A.1 Proof of Theorem 3.5

The proof follows the lines of the one of Theorem 1 in Segers (2020). To show (10) it is sufficient to show that for a real bounded Lipschitz function f, for any fixed $u \in V$ it holds that

$$\lim_{t\to\infty} \mathbb{E}[f(X_{V\setminus u}/t) \mid X_u = t] = \mathbb{E}[f(A_{u,V\setminus u})], \tag{20}$$

(van der Vaart, 1998, Lemma 2.2). Without loss of generality, we assume that $0 \le f(x) \le 1$ and $|f(x) - f(y)| \le L \sum_j |x_j - y_j|$ 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 3.1 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, $m \ge 2$. 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 $s \ne u$. 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 $s \ne u$ 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

$$(X_{C_{1:m-1}}/t) \mid X_u = t \xrightarrow{d} A_{u,C_{1:m-1}}, \tag{21}$$

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

$$\lim_{t\to\infty} \mathbb{E}\left[h(X_{C_{1:m-1}}/t) \mid X_u = t\right] = \mathbb{E}[h(A_{u,C_{1:m-1}})]. \tag{22}$$

To prove the convergence in (20) we start with the following inequality: for $\delta > 0$,

$$\begin{aligned}
&\left|\mathbb{E}[f(X_{V\setminus u}/t) \mid X_u = t] - \mathbb{E}[f(A_{u,V\setminus u})]\right| \\
&\le \left|\mathbb{E}\left[f(X_{V\setminus u}/t)\mathbb{1}(X_s/t \ge \delta) \mid X_u = t\right] - \mathbb{E}\left[f(A_{u,V\setminus u})\mathbb{1}(A_{us} \ge \delta)\right]\right| \qquad (23)\\
&+ \left|\mathbb{E}\left[f(X_{V\setminus u}/t)\mathbb{1}(X_s/t < \delta) \mid X_u = t\right] - \mathbb{E}\left[f(A_{u,V\setminus u})\mathbb{1}(A_{us} < \delta)\right]\right|. \qquad (24)
\end{aligned}$$

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 (23). We first deal with (23). The first expectation is equal to

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

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

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

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

$$\int_{[0,\infty)^{C_{1:m-1}}} \mathbb{E}\left[f\left(y, X_{C_m\setminus s}/t\right) \,\Big|\, X_s = ty_s\right] \mathbb{1}(y_s \ge \delta)\, \mathbb{P}\left(X_{C_{1:m-1}}/t \in \mathrm{d}y \mid X_u = t\right). \qquad (25)$$

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

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

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 $y_s \ne \delta$. We need to show that

$$\lim_{t\to\infty} g_t(y(t)) = g(y). \qquad (26)$$

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 3.1, we have

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

Since f is continuous, also

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

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 (26).

By the induction hypothesis (21) and Theorem 18.11 in van der Vaart (1998), the continuous convergence in (26) 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

$$\lim_{t\to\infty} \mathbb{E}\left[g_t\left(\frac{X_{C_{1:m-1}}}{t}\right) \;\middle|\; X_u = t\right] = \mathbb{E}[g(A_{C_{1:m-1}})]. \tag{27}$$

The expectation on the left-hand side of (27) is the integral in (25) while the right-hand side of (27) 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 (23) converges to 0 as $t \to \infty$, for any continuity point δ of A_{us}.

Analysis of (24). 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, (24) is bounded from above by

$$\mathbb{P}[X_s/t < \delta \mid X_u = t] + \mathbb{P}[A_{us} < \delta]. \tag{28}$$

By the induction hypothesis

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

and (28) 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 3.4. 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 (24):

$$\left|\mathbb{E}\left[f\left(\frac{X_{C_{1:m-1}}}{t}, \frac{X_{C_m\setminus s}}{t}\right)\mathbb{1}\left(\frac{X_s}{t} < \delta\right) \middle| X_u = t\right] - \mathbb{E}\left[f\left(\frac{X_{C_{1:m-1}}}{t}, 0\right)\mathbb{1}\left(\frac{X_s}{t} < \delta\right) \middle| X_u = t\right]\right| \tag{29}$$

$$+ \left|\mathbb{E}\left[f(A_{u,C_{1:m-1}}, A_{u,C_m\setminus s})\mathbb{1}(A_{u,s} < \delta)\right] - \mathbb{E}\left[f(A_{u,C_{1:m-1}}, 0)\mathbb{1}(A_{u,s} < \delta)\right]\right| \tag{30}$$

$$+ \left|\mathbb{E}\left[f\left(\frac{X_{C_{1:m-1}}}{t}, 0\right)\mathbb{1}\left(\frac{X_s}{t} < \delta\right) \middle| X_u = t\right] - \mathbb{E}\left[f(A_{u,C_{1:m-1}}, 0)\mathbb{1}(A_{u,s} < \delta)\right]\right|. \tag{31}$$

We treat each of the three terms in turn.

Equation (31) 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 (29). 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 (29) is bounded by

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

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

$$\int_{[0,\delta]} \mathbb{E}\left[\min\left(1, L\|X_{C_m\setminus s}/t\|_1\right) \;\mid\; X_s/t = x_s\right] \mathbb{P}\left(X_s/t \in \mathrm{d}x_s \mid X_u = t\right). \tag{32}$$

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

$$\mathbb{E}\left[\min\left(1, L\|X_{C_m\setminus s}/t\|_1\right) \mathbb{1}\left\{\forall v \in C_m \setminus s : X_v/t \leq \eta\right\} \mid X_s/t = x_s\right] \tag{33}$$
$$+\, \mathbb{E}\left[\min\left(1, L\|X_{C_m\setminus s}/t\|_1\right) \mathbb{1}\left\{\exists v \in C_m \setminus s : X_v/t > \eta\right\} \mid X_s/t = x_s\right]. \tag{34}$$

The integrand in (33) 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 (34) 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 (32) we can thus bound it by

$$\int_{[0,\delta]} \left[L\,|C_m \setminus s|\,\eta + \mathbb{P}\left(\exists v \in C_m \setminus s : X_v/t > \eta \mid X_s/t = x_s\right)\right] \mathbb{P}\left(X_s/t \in \mathrm{d}x_s \mid X_u = t\right). \tag{35}$$

Consider the supremum of the probability in the integrand over the values $x_s \in [0,\delta]$ to bound the integral further. Hence (35) 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 3.4 and the fact that η can be chosen arbitrarily small we conclude that (29) converges to 0 as $t \to \infty$.

Finally we look at the term in (30). As f has range contained in $[0,1]$ and is Lipschitz continuous, the expression in (30) is smaller than or equal to

$$\mathbb{E}\left[\mathbb{1}(A_{us} < \delta) \min\left(1,\, L \textstyle\sum_{v \in C_m\setminus s} A_{uv}\right)\right]. \tag{36}$$

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 (36) as

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

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 (36) converges to 0 as $\delta \downarrow 0$. □

A.2 Proof of Proposition 3.9

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 (4), the copula, K, of G is

$$K(u) = G(-1/\ln u_1, \dots, -1/\ln u_d) = \exp(-\ell(-\ln u_1, \dots, -\ln u_d)), \qquad u \in (0,1)^d.$$

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

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

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

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

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

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

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 $x = (x_2, \dots, x_d) \in (0,\infty)^{d-1}$, we find

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

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

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

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 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 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 tightness follows. $\square$

B Proofs for Section 4

B.1 Proof of Proposition 4.2

By Proposition 3.9, the Markov random field X of Assumption 4.1 satisfies Assumption 3.1. Moreover, Assumption 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 (Segers, 2020, Example 4) and therefore does not have an atom at zero. We can thus apply Theorem 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 3.5. First we identify the weak limit $\nu_{C,s}$ of $(X_{C\setminus s}/t \mid X_s = t)$ in Assumption 3.1. By a reasoning similar to the one in the proof of Corollary 1 in Segers (2020), the limit must be the same as the one of $(X_{C\setminus s}/X_s \mid X_s > t)$ as $t \to \infty$. But by Engelke et al. (2015, Theorem 2) we have, as $t \to \infty$,

$$(\ln X_v - \ln X_s,\, v \in C \setminus s \mid X_s > t) \xrightarrow{d} \mathcal{N}_{|C\setminus s|}(\mu_{C,s}(\Delta_C), \Psi_{C,s}(\Delta_C)), \tag{37}$$

where the mean vector is

$$\big(\mu_{C,s}(\Delta)\big)_v = -2\delta_{sv}^2, \qquad v \in C \setminus s, \tag{38}$$

and the covariance matrix $\Psi_{C,s}(\Delta)$ is as in (14). 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 (37). 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 3.5. 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 3.2 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 (38), 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 (14), we have

$$\operatorname{cov}(\ln Z_e, \ln Z_f) = 2(\delta_{si}^2 + \delta_{sj}^2 - \delta_{ij}^2) \tag{39}$$

By Theorem 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

$$\ln A_{uv} = \ln\left(\prod_{e \in (u \rightsquigarrow v)} Z_e\right) = \sum_{e \in (u \rightsquigarrow v)} \ln Z_e, \qquad v \in V \setminus u. \tag{40}$$

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 (40):

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

which coincides with the element v of the vector $\mu_u(\Delta)$ in (17). It remains to show that the covariance matrix of $(\ln A_{uv}, v \in V \setminus u)$ is $\Sigma_u(\Delta)$ in (18).

Calculating $\Sigma_u(\Delta)$. Let $i, j \in V \setminus u$. By (40) and the bilinearity of the covariance operator, we have

$$\operatorname{cov}(\ln A_{ui}, \ln A_{uj}) = \sum_{e \in (u \rightsquigarrow i)} \sum_{f \in (u \rightsquigarrow j)} \operatorname{cov}(\ln Z_e, \ln Z_f).$$

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 (39), the covariance becomes

$$\begin{aligned}\operatorname{cov}(\ln A_{ui}, \ln A_{uj}) &= \sum_{e \in (u \rightsquigarrow i)} \operatorname{var}(\ln Z_e) \\ &= \sum_{e \in (u \rightsquigarrow i)} 4\delta_e^2 = 4p_{ui} = 2(p_{ui} + p_{uj} - p_{ij}),\end{aligned}$$

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

$$\begin{aligned}\operatorname{cov}(\ln A_{ui}, \ln A_{uj}) &= \sum_{e \in (u \rightsquigarrow a)} \operatorname{var}(\ln Z_e) + \operatorname{cov}(\ln Z_{e_a}, \ln Z_{f_a}) \\ &= 4p_{ua} + \operatorname{cov}(\ln Z_{e_a}, \ln Z_{f_a}).\end{aligned}$$

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

$$\begin{aligned}\operatorname{cov}(\ln A_{ui}, \ln A_{uj}) = 4p_{ua} &= 2((p_{ui} - p_{ai}) + (p_{uj} - p_{aj})) \\ &= 2(p_{ui} + p_{uj} - (p_{ai} + p_{aj})) \\ &= 2(p_{ui} + p_{uj} - p_{ij}),\end{aligned}$$

since the shortest path between i and j passes through a.

(3.b) Suppose e_a and f_a belong to the same clique; see Figure 4. Writing $e_a = (a, k)$ and $f_a = (a, l)$, we find, in view of (39),

$$\begin{aligned}\operatorname{cov}(\ln A_{ui}, \ln A_{uj}) &= 4p_{ua} + 2\left(\delta_{ak}^2 + \delta_{al}^2 - \delta_{kl}^2\right) \\ &= 2\left(\left(p_{ua} + \delta_{ak}^2\right) + \left(p_{ua} + \delta_{al}^2\right) - \delta_{kl}^2\right) \\ &= 2\left(p_{uk} + p_{ul} - \delta_{kl}^2\right) \\ &= 2\left((p_{ui} - p_{ki}) + (p_{uj} - p_{lj}) - \delta_{kl}^2\right) \\ &= 2\left(p_{ui} + p_{uj} - \left(p_{ki} + p_{lj} + \delta_{kl}^2\right)\right) \\ &= 2(p_{ui} + p_{uj} - p_{ij}),\end{aligned}$$

since the shortest path between i and j passes by k and l.

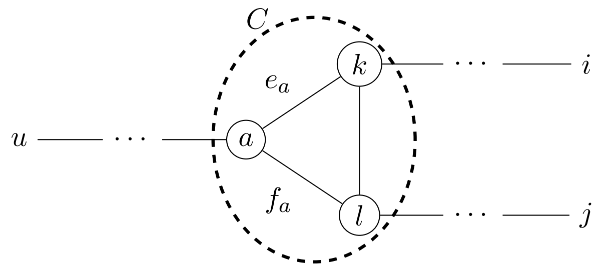


Figure 4: Calculation of cov $(\ln A_{ui}, \ln A_{uj})$ in the proof of Proposition 4.2. The paths from u to i and from u to j have the path from u to a in common. On these two paths, the edges right after node a are $e_a = (a, k)$ and $f_a = (a, l)$, respectively. The picture considers the case (3.b) where the three nodes a, k, l belong to the same clique, say C.

We conclude that the covariance matrix of $(\ln A_{uv}, v \in V \setminus u)$ is indeed $\Sigma_u(\Delta)$ in (18).

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 (40) 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 (39) 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 $\Theta_u^Z(\Delta) = \left(\Sigma_u^Z(\Delta)\right)^{-1}$, we find that $\Sigma_u(\Delta)$ is invertible as well with inverse

$$\Theta_u(\Delta) = (M_u^{-1})^\top \, \Theta_u^Z(\Delta) \, M_u^{-1}. \tag{41}$$

$P(\Delta)$ 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. □

B.2 Proof of Proposition 4.4

Let $H_{P(\Delta)}$ be the $|V|$-variate max-stable Hüsler–Reiss distribution in (15) with parameter matrix $P(\Delta)$ in (16). To show (19), it is sufficient to verify that

$$\lim_{t\to\infty} t\,\mathbb{P}\left(\exists v\in V : X_v > tx_v\right) = -\ln H_{P(\Delta)}(x), \qquad x\in(0,\infty)^V. \tag{42}$$

By the inclusion–exclusion formula,

$$t\,\mathbb{P}\left(\exists v\in V : X_v > tx_v\right) = \sum_{i=1}^{|V|}(-1)^{i-1}\sum_{W\subseteq V:|W|=i} t\,\mathbb{P}\left(\forall v\in W : X_v > tx_v\right).$$

For non-empty $W\subseteq V$, we have, for any $u\in W$,

$$t\,\mathbb{P}\left(\forall v\in W : X_v > tx_v\right) = t\,\mathbb{P}(X_u > tx_u)\,\mathbb{P}\left(\forall v\in W\setminus u : X_v > tx_v \mid X_u > tx_u\right).$$

Proposition 4.2 implies

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

where the random vector $(\ln A_{uv}, v\in V\setminus u)$ is distributed according to the multivariate normal law stated in the proposition. By Segers (2020, Corollary 1), the above convergence implies

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

Since the marginal distributions of X are all unit-Fréchet, the equivalence between statements (a) and (c) in Theorem 2 in Segers (2020) implies that

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

where ζ is a unit-Pareto random variable, i.e., $\mathbb{P}(\zeta > y) = 1/y$ for $y\geq 1$, independent of $(A_{uv}, v\in V\setminus u)$. It follows that, for non-empty $W\subseteq V$ and for $u\in W$,

$$\lim_{t\to\infty} t\,\mathbb{P}\left(\forall w\in W : X_v > tx_v\right) = x_u^{-1}\,\mathbb{P}\left(\forall v\in W\setminus u : \zeta A_{uv} > x_v/x_u\right).$$

As $1/\zeta$ is uniformly distributed on $[0,1]$, we can rewrite the limit probability as an expectation:

$$\begin{aligned} x_u^{-1}\,\mathbb{P}\left(1/\zeta < \min_{v\in W\setminus u}(x_u/x_v)A_{uv}\right) &= x_u^{-1}\,\mathbb{E}\left[\min\left\{1, \min_{v\in W\setminus u}(x_u/x_v)A_{uv}\right\}\right] \\ &= \mathbb{E}\left[\min_{v\in W} x_v^{-1}A_{uv}\right], \end{aligned}$$

since $A_{uu}=1$. 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:

$$\begin{aligned} \mathbb{E}\left[\min_{v\in W} x_v^{-1}A_{uv}\right] &= \int_0^\infty \mathbb{P}\left(\forall v\in W : x_v^{-1}A_{uv} > y\right)\mathrm{d}y \\ &= \int_0^{x_v^{-1}} \mathbb{P}\left(\forall v\in W\setminus u : A_{uv} > x_v y\right)\mathrm{d}y \\ &= \int_{\ln x_v}^\infty \mathbb{P}(\forall v\in W\setminus u : \ln A_{uv} > \ln(x_v) - z)\, e^{-z}\,\mathrm{d}z. \end{aligned}$$

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

$$\begin{aligned}
&\lim_{t\to\infty} t\,\mathbb{P}\left(\exists v \in V : X_v > tx_v\right) \\
&\qquad = \sum_{i=1}^{|V|} (-1)^{i-1} \sum_{W\subseteq V:|W|=i} \int_{\ln x_v}^{\infty} \mathbb{P}(\forall v \in W \setminus u : \ln A_{uv} > \ln(x_v) - z)\, e^{-z}\, \mathrm{d}z.
\end{aligned}$$

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 (17) and (18), respectively. In view of the expression of $H_{P(\Delta)}$ in (15), we find that (42) indeed holds. □

B.3 Proof of Proposition 4.5

By Proposition 4.4, the Hüsler–Reiss Markov field X in Assumption 4.1 satisfies (6) with F its joint cumulative distribution function and $G = H_{P(\Delta)}$. It follows that (8) holds too, yielding the weak convergence of high-threshold excesses to a Pareto random vector Y with distribution given in (7). It remains to show that Y is an extremal graphical model with respect to the given graph $\mathcal{G}$ in the sense of Definition 1 in Engelke and Hitz (2020).

Proposition 4.2 implies

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

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 (8) 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 (18); see (41). 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

$$\left(\Theta_u(\Delta)\right)_{ij} = 0.$$

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

$$\begin{aligned}
\left(\Theta_u(\Delta)\right)_{ij} &= \sum_{a\in V\setminus u} \sum_{b\in V\setminus u} \left((M_u^{-1})^\top\right)_{ia} \left(\Theta_u^Z(\Delta)\right)_{ab} \left(M_u^{-1}\right)_{bj} \\
&= \sum_{a\in V\setminus u} \sum_{b\in V\setminus u} \left(M_u^{-1}\right)_{ai} \left(\Theta_u^Z(\Delta)\right)_{ab} \left(M_u^{-1}\right)_{bj}.
\end{aligned}$$

Now, $\left(M_u^{-1}\right)_{ai}$ and $\left(M_u^{-1}\right)_{bj}$ 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 $\left(\Theta_u^Z(\Delta)\right)_{ab} = 0$, since $\Sigma_u^Z(\Delta)$ and thus $\Theta_u^Z(\Delta)$ are block-diagonal. □

B.4 Proof of Proposition 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 $P(\Delta)$.

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, as required.

- 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 the path sum $p_{b\bar{\imath}}$ in terms of observable path sums, we can compute $\delta^2_{ab} = p_{a\bar{\imath}} - p_{b\bar{\imath}}$. □