Markov Equivalence of Max-Linear Bayesian Networks

Carlos Améndola[1] **Benjamin Hollering**[2] **Seth Sullivant**[2] **Ngoc Tran**[3]

[1]Department of Mathematics , Technical University of Munich, Germany
[2]Department of Mathematics , North Carolina State University, USA
[3]Department of Mathematics , University of Austin, Texas , USA

Abstract

Max-linear Bayesian networks have emerged as highly applicable models for causal inference via extreme value data. However, conditional independence (CI) for max-linear Bayesian networks behaves differently than for classical Gaussian Bayesian networks. We establish the parallel between the two theories via tropicalization, and establish the surprising result that the Markov equivalence classes for max-linear Bayesian networks coincide with the ones obtained by regular CI. Our paper opens up many problems at the intersection of extreme value statistics, causal inference and tropical geometry.

1 INTRODUCTION

A *max-linear Bayesian network* is a special class of a graphical model on a directed acyclic graph (DAG) used to model causal relations between extreme values [Gissibl and Klüppelberg, 2018, Klüppelberg and Lauritzen, 2019, Engelke and Hitz, 2020]. Denoting the maximum operator max by $\vee$, its defining equation is

$$X_i = \bigvee_{j=1,\ldots,n} c_{ij} X_j \vee Z_i, \qquad c_{ij}, Z_i \geq 0 \qquad (1)$$

for each $i \in [n] = \{1, \ldots, n\}$, where the Z_i are independent and unobserved random variables and C is a matrix of coefficients supported on a DAG with n nodes. Like their classical counterparts, they are versatile and easy to interpret. They are the simplest class of models that exhibit *cascading failure*, where extreme measurements X_j (rainfall, contaminant level, risk, financial return) occurring at a large number of nodes can be traced to a few common sources Z_i (storm, chemical spill, catastrophic failure, financial shock). Such cascading failures are commonly experienced in hydrology, engineering, and finance, and therefore max-linear Bayesian

networks are finding many applications in these domains [Gissibl, 2018, Gissibl and Klüppelberg, 2018, Buck and Klüppelberg, 2020, Janßen and Wan, 2020, Klüppelberg and Krali, 2021]. Most recently, Tran et al. [2021] fitted max-linear Bayesian trees to data and achieved state-of-the-art results on the Hidden River problem, the current benchmark for causal discovery from extreme data [Asadi et al., 2015]. This indicates that max-linear Bayesian networks are highly suited to model causal relations between large observed values of random variables.

Conditional independence (CI) theory is fundamental to causal inference on Bayesian networks [Spirtes et al., 2000, Pearl and Verma, 1995]. While max-linear Bayesian networks are special instances of Bayesian networks, they have, however, a *different* CI theory [Améndola et al., 2021]. While CI statements in Bayesian networks can be found by the classical *d-separation* criterion [Geiger et al., 1990a,b, Meek, 1995], CI statements on max-linear Bayesian networks are given by ∗-*separation*, a stronger form of separation (see Example 1.1). Furthermore, in sharp contrast with the classical case, CI statements on a max-linear Bayesian network can depend on both the coefficients C and the context, that is, on some (partial) realization of the conditioning set [Améndola et al., 2021, Example 1.3].

The natural next step is the question of Markov equivalence. Two graphs G and H are called *Markov equivalent* if they yield the same set of CI statements under a separation criterion. Based on CI statements alone, one can only hope to recover a DAG up to Markov equivalence. Nevertheless, while Markov equivalent graphs yield the same conditional independence structures, they have very different causal structures. Understanding the Markov equivalence classes allows for the development of algorithms that infer Markov equivalence classes from data such as the PC algorithm for Bayesian networks [Spirtes et al., 2000].

A necessary and sufficient condition for determining Markov equivalence for classical Bayesian networks is well-known [Verma and Pearl, 1990, 1992]. Specifically, two directed
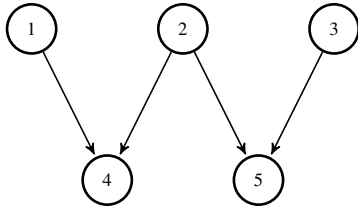
Figure 1: The Cassiopeia graph.

acyclic graphs are Markov equivalent if and only if they have the same skeleton and same unshielded colliders (also known as v-structures or immoralities) [Andersson et al., 1997]. In this paper we answer the natural question of determining what are the Markov equivalence classes for max-linear Bayesian networks.

Here is a motivating example. The relevant definitions will be given in Section 2, and more details in Example 2.9.

Example 1.1 (Cassiopeia). *Consider the graph in Figure 1. It holds that 1 and 3 are d-connected by 4 and 5. Therefore, the global Markov property is not enough to conclude that $1 \perp\!\!\!\perp 3 \mid 4, 5$. Nevertheless, 1 and 3 are $*$-separated by 4 and 5 and hence we can conclude that for a max-linear Bayesian network supported on the Cassiopeia graph, the conditional independence statement $1 \perp\!\!\!\perp 3 \mid 4, 5$ holds.*

A priori, one would expect that the Markov equivalence class for max-linear Bayesian networks refine those seen for Gaussian Bayesian networks since it was shown in [Améndola et al., 2021] that max-linear Bayesian networks have additional valid conditional independence statements. We show in our main theorem that in fact it is not true that the Markov equivalence classes are refined, and that instead the equivalence classes of max-linear Bayesian networks coincide with the equivalence classes of Gaussian Bayesian networks.

Theorem (Theorem 3.4). *Gaussian Bayesian networks and max-linear Bayesian networks have the same Markov equivalence classes.*

Aside from the tree case [Tran et al., 2021], finding a consistent estimator for the parameters c_{ij}^* of the max-linear Bayesian networks to data remains difficult. One primary reason is that C^* is a matrix of max-weighted paths on an unknown DAG G. The set of such matrices is a non-convex piecewise-linear manifold made up of many low-dimensional cones [Tran, 2014]. Thus, estimators of C^* are often sensitive to noise [Gissibl et al., 2021]. For Gaussian graphical models, conditional independence statements are governed by rank constraints on the covariance matrix (see e.g. [Sullivant, 2018, Proposition 4.1.9]). Our second main result establishes an analogue of this result for the max-linear model via tropical geometry. In particular, we show that some tropical rank constraints on the tropical-

ization of the covariance matrix correspond to conditional independence statements for max-linear Bayesian networks.

Theorem (Theorem 5.2). *The conditional independence statements implied by d-separation in a max-linear Bayesian network impose tropical rank constraints that hold for every tropical covariance matrix supported on the network.*

This result opens up new directions for conditional independence testing in max-linear Bayesian networks. Namely, instead of finding a consistent estimator for C^*, which so far has proven difficult, one could opt to find a consistent estimator for the tropical covariance matrix and/or the tropical ranks of its sub-blocks. Furthermore, this theorem offers an *algebraic* way to handle CI, complementing the path-based $*$-separation criterion of Améndola et al. [2021]. In particular, this brings us closer in the task of identifying sufficient conditions for CI statements to hold in max-linear Bayesian networks.

This paper is organized as follows. In Section 2 we discuss the preliminary concepts and notation, including some basics on tropical algebra. The first main result, Theorem 3.4, and its proof are the content of Section 3. We explore deeper the connection of max-linear models to tropical geometry by proving a tropical analogue of the classical trek rule in Section 4. We use the tropical trek rule to obtain tropical rank constraints, culminating in our second main result, Theorem 5.2, in Section 5. Finally, we consider in Section 6 some interesting open problems and future research directions.

2 PRELIMINARIES

2.1 TROPICAL BASICS

Here we recall some concepts of tropical geometry for our discussion. For an introductory exposition to this field we recommend Maclagan and Sturmfels [2015].

We work in the *max-times semiring* $(\mathbb{R}_{\geq}, \vee, \odot)$, defined by

$$a \vee b := \max(a, b), \quad a \odot b := ab \quad \text{for } a, b \in \mathbb{R}_{\geq} := [0, \infty).$$

These tropical operations extend to $\mathbb{R}_{\geq}^n$ coordinate-wise, to scalar-vector multiplication as

$$\lambda \odot x = (\lambda x_1, \ldots, \lambda x_d) \quad \text{for } \lambda \in \mathbb{R}_{\geq} \text{ and } x \in \mathbb{R}_{\geq}^n,$$

and to corresponding matrix product as

$$(A \odot B)_{ij} = \bigvee_{\ell=1}^{n} a_{i\ell} b_{\ell j}$$

for $A \in \mathbb{R}_{\geq}^{m \times n}$ and $B \in \mathbb{R}_{\geq}^{n \times p}$. In particular, this defines tropical matrix powers $A^{\odot k}$ for $k \in \mathbb{N}$ where $A^{\odot 0} = I_n$ is the identity matrix.

Analogously, we can define the *tropical determinant* of a matrix $A \in \mathbb{R}^{n \times n}$ as

$$\text{tdet}(A) = \bigvee_{\sigma \in S_n} a_{1\sigma(1)} a_{2\sigma(2)} \ldots a_{n\sigma(n)}$$

where S_n denotes the symmetric group on $[n]$.

A matrix is *tropically singular* if the maximum in the evaluation of the tropical determinant is attained at least twice.

Definition 2.1. *The* tropical rank $\text{trank}(M)$ *of a matrix* $M \in \mathbb{R}^{m \times n}$ *is the largest integer r such that M has a tropically non-singular $r \times r$ minor.*

Example 2.2. *The matrix M given by*

$$\begin{pmatrix} 6 & 3 & 0 \\ 0 & 8 & 4 \\ 6 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} \odot \begin{pmatrix} 0 & 4 & 2 \end{pmatrix} \vee \begin{pmatrix} 3 \\ 0 \\ 3 \end{pmatrix} \odot \begin{pmatrix} 2 & 1 & 0 \end{pmatrix}$$

has tropical rank 2. Indeed, the 2×2 minor

$$A = \begin{pmatrix} 6 & 3 \\ 0 & 8 \end{pmatrix}$$

is tropically non-singular since

$$\text{tdet}(A) = 6 \odot 8 \vee 0 \odot 3 = 48 \vee 0 = 48$$

achieves its maximum uniquely. On the other hand, M is tropically singular since

$$\text{tdet}(M) = 96 \vee 84 \vee 0 \vee 0 \vee 0 \vee 96 = 96$$

attains its maximum twice, namely at the terms $6 \odot 8 \odot 2$ and $6 \odot 4 \odot 4$.

Remark 2.3. *While tropical geometry is often introduced as min-plus or max-plus operations over $\mathbb{R} \cup \{\infty\}$ or $\mathbb{R} \cup \{-\infty\}$ respectively, our max-times semiring $(\mathbb{R}_{\geq}, \vee, \odot)$ is isomorphic to the latter by taking logarithms.*

2.2 MAX-LINEAR BAYESIAN NETWORKS

A *max-linear Bayesian network* is given by a random vector $X = (X_1, \ldots X_n)$ with vertices on a directed acyclic graph $G = ([n], E)$, edge weights $c_{ij} \geq 0$, and independent positive random variables $Z_1, \ldots, Z_n$ called *innovations*. The Z_i have support $\mathbb{R}_{>} = (0, \infty)$ and have atom-free distributions. Then X is given by the *recursive structural equations*

$$X_i = \bigvee_{j \in \text{pa}(i)} c_{ij} X_j \vee Z_i,$$

or $X = C \odot X \vee Z$ in tropical notation. This system has solution $X = C^* \odot Z$, that is,

$$X_i = \bigvee_{j \in \text{an}(i) \cup i} c_{ij}^* Z_j \tag{2}$$

where $C^* = \bigvee_{k=0}^{n-1} C^{\odot k}$ is the *Kleene star* of the matrix C. In these equations $\text{pa}(i)$ denotes the parents of node i and $\text{an}(i)$ denotes the ancestors of i.

Conditional independence in max-linear models can be quite different from conditional independence in classical Bayesian networks. For the latter, the *d-separation* criterion gives a complete set of valid conditional independence statements for the model [Meek, 1995].

We use the following standard notation (see e.g. Klüppelberg and Lauritzen [2019]). A *path* in a DAG G is a sequence of vertices $i_0, i_1, \ldots i_k$ such that $i_\ell \to i_{\ell+1}$ is an edge in G or $i_{\ell+1} \to i_\ell$ is an edge in G for each $\ell = 0, \ldots k$. A *directed path* has edges $i_\ell \to i_{\ell+1}$ for all ℓ. If there is a directed path from i to j, we say that i is an *ancestor* of j and denote by $\text{an}(K)$ the set of all ancestors of nodes in K. A *collider* on a path is a vertex i_ℓ in the path such that $i_{\ell-1} \to i_\ell \leftarrow i_{\ell+1}$.

Definition 2.4. *Two vertices i and j in G are d-connected given a set $K \subseteq [n] \setminus \{i, j\}$ if there is a path π from i to j such that all colliders on π are in $K \cup \text{an}(K)$ and no non-collider on π is in K. If $I, J, K \subseteq [n]$ are pairwise disjoint sets with I and J nonempty, then K d-separates I and J if no pair of nodes $i \in I$ and $j \in J$ are d-connected given K. We denote this by $I \perp_d J | K$.*

It was noted in Klüppelberg and Lauritzen [2019] that d-separation does not give all valid conditional independence statements for a max-linear Bayesian network. Recently, a new criterion named *∗-separation* which gives a complete set of conditional independence statements for max-linear models was given in Améndola et al. [2021].

Definition 2.5. *A path π between i and j in a DAG is ∗-connecting given a set $K \subseteq [n] \setminus \{i, j\}$ if and only if π is d-connecting given K and π contains at most one collider. Two nodes i and j are ∗-connected given K if there is a ∗-connecting path. If $I, J, K \subseteq [n]$ are pairwise disjoint sets with I and J nonempty, then K ∗-separates I and J if no pair of nodes $i \in I$ and $j \in J$ are ∗-connected given K. We denote this by $I \perp_* J | K$.*

The 5 basic shapes of ∗-connecting paths are illustrated in Figure 2.

An alternate formulation of ∗-separation involves the notion of the conditional reachability DAG. This alternate formulation is useful for giving the proofs of our main result.

Definition 2.6. *Let G be a DAG and $K \subseteq [n]$. A directed path π from i to j factors through K if there exists a vertex $k \in \pi$ such that $k \neq i, j$ and $k \in K$. The conditional reachability DAG, denoted G_K^* is a graph on $[n]$ with edges given by $i \to j \in G_K^*$ if and only if there exists a directed path from $i \to j$ that does not factor through K.*

Example 2.7. *Let G be the DAG pictured on the left in Figure 3 and let $K = \{3\}$. Then the conditional reachability*
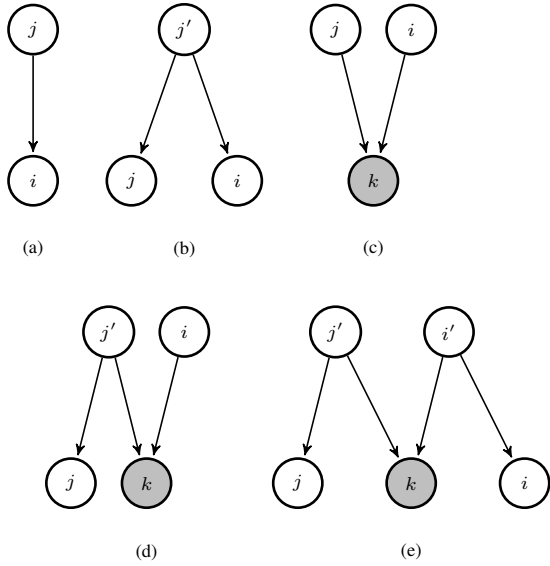
Figure 2: Types of $*$-connecting paths between two nodes i and j with shaded nodes in K.
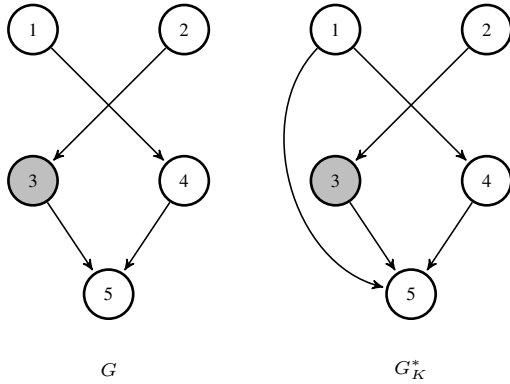

Figure 3: A DAG G and the corresponding conditional reachability graph G_K^* for $K = \{3\}$.

graph G_K^* is the graph pictured on the right in Figure 3. Note that the additional edge $1 \to 5$ is due to the fact that the path $1 \to 4 \to 5$ is a directed path that does not factor through K. On the other hand, there is no edge between 2 and 5 since the only directed path between 2 and 5 in G is $2 \to 3 \to 5$ which factors through K.

Observe that since $1 \to 5$ is a $*$-connecting path in G_K^* we have that $1 \not\perp_* 5|3$. The only path between 2 and 5 in G_K^* is $2 \to 3 \to 5$ which is not $*$-connecting since $3 \in K$ so $2 \perp_* 5|3$.

Remark 2.8. *With this definition, we can say that for $I, J, K \subseteq [n]$ pairwise disjoint sets, I and J are $*$-separated by K in G if and only if there are no $*$-connecting paths from I to J in G_K^*.*

Checking $*$-separation in G_K^* is the analogue of checking

d-separation as undirected separation in a corresponding moralized graph, see [Lauritzen, 1996, Proposition 3.25].

Example 2.9. *To provide some intuition on the* 'at most one collider' *condition in Definition 2.5, we consider again the Cassiopeia graph from Example 1.1 and the path between 1 and 3 with two colliders in $K = \{4, 5\}$. For simplicity, let all nonzero coefficients c_{ij} equal to one. Then the max-linear model states that*

$$X_1 \vee X_2 \leq x_4 \quad and \quad X_2 \vee X_3 \leq x_5.$$

If $x_4 < x_5$, then $X_3 \leq x_5$ so that x_5 cannot be caused by X_2 but only by X_3 or Z_5. Analogously, if $x_4 \geq x_5$ then x_4 can only be caused by X_1 or Z_4. Finally, if $x_4 = x_5$ we must have $X_2 = x_4 = x_5$ (since the Z_i are atom-free) so that both are caused by X_2. In any of the three cases, we have that $1 \perp\!\!\!\perp 3 \,|\, \{4, 5\}$.

3   SOLVING MARKOV EQUIVALENCE

In this section we compare Markov equivalence under d-separation and $*$-separation and show that they give the same Markov equivalence classes.

Definition 3.1. *Two graphs G and H with vertex set V are called* Markov equivalent *if they yield the same set of conditional independence statements under a global Markov property, i.e., for all pairwise disjoint $I, J, K \subseteq V$,*

$$I \perp J \,|\, K \implies I \perp\!\!\!\perp J \,|\, K$$

where $\perp$ is a separation criterion.

The following theorem by Verma and Pearl [1990, 1992] characterizes which graphs are Markov equivalent when considering the d-separation criterion. An alternative proof can also be found in [Andersson et al., 1997].

Theorem 3.2 (Verma and Pearl [1990], Theorem 1). *Two directed acyclic graphs G and H are Markov equivalent under the d-separation criterion if and only if the following two conditions hold:*

1. *G and H have the same skeleton, which is the undirected graph obtained by removing edge directions.*

2. *G and H have the same unshielded colliders, which are triples $i, j, k \in [n]$ which induce a subgraph of the form $i \to k \leftarrow j$.*

We use $\sim_d$ and $\sim_*$ to denote Markov equivalence under d-separation and $*$-separation respectively. Since every $*$-connecting path is also d-connecting, it holds that if $I \perp_d J|K$ then $I \perp_* J|K$. Consequently, one might think that the additional statements obtained from $*$-separation would refine the Markov equivalence classes of d-separation but this is not the case. We will show that the Markov equivalence classes are actually the same. The following lemma will be useful in the proof.

Lemma 3.3. *Let $K \subseteq [n]$. Let G and H be DAGs on $[n]$ such that $G \sim_d H$ and suppose $i \to j \in G_K^*$. Then one of the following holds:*

1. *$i \to j \in H_K^*$,*

2. *$j \to i \in H_K^*$,*

3. *i and j have a common parent ℓ in H_K^*, i.e., there exists $l \in [n]$ such that $\ell \to i, \ell \to j \in H_K^*$.*

Proof. The existence of the edge $i \to j \in G_K^*$ implies that there is a directed path π_G from i to j in G that does not factor through K. Since it holds that $G \sim_d H$ hence G and H have the same skeleton. This means there exists a path π_H between i and j in H whose edge directions we now consider.

First we suppose that π_H has no colliders on it. Then π_H must be a directed path from i to j, a directed path from j to i, or has exactly one source $\ell \neq i, j$ on it. If π_H is a directed path from i to j then (1) holds since there is no vertex in K on π and similarly if π_H is a directed path from j to i then (2) holds. If there is a source $\ell \neq i, j$ then we get that (3) holds since there must be a directed paths from ℓ to both i and j and no vertex on these paths can be in K.

Now we suppose that there is at least one collider on π_H. We will show a different path exists by looking at the colliders on the path. Suppose v is a collider on π_H and let $u \to v \leftarrow w$ be path π_H locally around v. The local path around v in π_G must have the form $u \to v \to w$ since π_G is the path that induces the edge $i \to j \in G_K^*$. Since we have that $G \sim_d H$ they must have the same unshielded colliders by Theorem 3.2 and since the triple (u, v, w) is not an unshielded collider in G, it cannot be an unshielded collider in H. Combining this with the fact that locally around v, π_H has the form $u \to v \leftarrow w$, there must either exist an edge $u \to w$ or $u \leftarrow w$ in H. So we can create a new path, π_H', in H by replacing $u \to v \leftarrow w$ with the edge between u and w. The path π_H' is one edge shorter than π_H and has one less collider or one of the vertices u, w has become a collider. We can inductively apply the same argument to π_H' though until we obtain a path between i and j with no colliders or we end up with a direct edge between i and j. If we have a direct edge between i and j then either (1) or (2) holds and if we have a path with no colliders than the result holds by the previous paragraph. $\square$

We are now ready to prove our promised first main result.

Theorem 3.4. *d-separation and $*$-separation induce the same Markov equivalence classes on a DAG G.*

Proof. Let $G \sim_d H$. We will show that for any $K \subseteq [n]$, if there exists a $*$-connecting path π_G between i and j in G then there is a $*$-connecting path π_H between i and j in H. This implies that G and H have the same $*$-separations and

hence are Markov equivalent with respect to $*$-separation. So fix K and let π_G be a $*$-connecting path between i and j in G conditioned on K. We now argue that the desired path π_H exists based on which of the five possible forms displayed in Figure 2 π_G may take. For each of the possible $*$-connecting paths π_G, we apply Lemma 3.3 to each edge in π_G and analyze the possible resulting graphs. Throughout the rest of the proof, we denote by (1), (2), and (3) the three outcomes that we can get by applying Lemma 3.3 to an edge in π_G.

(a) Suppose π_G has the form of path (a) in Figure 2. Then Lemma 3.3 gives three possibilities which are all $*$-connecting paths between i and j.

(b) Suppose π_G has the form of path (b) in Figure 2. We analyze these cases up to the symmetry obtained by interchanging i and j.

- If (1) holds for both edges then the path $j \leftarrow j' \to i$ is also in H_K^*.

- If (1) holds for either edge with (2) holding for the other then we have a directed path between i and j in H_K^* which implies the existence of a directed path between i and j in H that does not factor through K hence we $i \to j \in H_K^*$ which is $*$-connecting as well.

- If (1) holds for $j' \to j$ and (3) holds for $j' \to i$ then there exists an ℓ such that $j' \leftarrow \ell \to i \in H_K^*$. The existence of these two paths implies that $j \leftarrow \ell \to i \in H_K^*$.

- If (2) holds for both edges, (3) holds for both edges, or (2) holds for one edge and (3) for the other then j' would be a collider in H_K^*. If such a collider exists though then we are guaranteed a path directly from i to j in H by the proof of Lemma 3.3 which means we have a path of type (a) in H_K^*.

(c) Suppose π_G has the form of path (c) in Figure 2. We again use the symmetry between i and j to reduce the number of cases.

- If (1) holds for both edges then the path is unchanged and hence $*$-connecting.

- Suppose (1) holds for the edge $i \to k$ and (2) holds for the edge $j \to k$. Then locally around k, π_G has the form $u \to k \leftarrow v$ but in H this path must be of the form $u \to k \to v$ which means that the triple (u, k, v) is no longer an unshielded collider in H though it was in G. This contradicts the assumption that $G \sim_d H$ though so this scenario is not possible. The same argument applies to the case where (2) holds for both edges and the case where (2) and (3) hold for the edges.

- Suppose (1) holds for the edge $i \to k$ and (3) holds for the edge $j \to k$. Then there exists a common parent j' of j and k which implies the existence of a path of type (d) between the vertices i, j, j', and k.

- Suppose (3) holds for both edges. Then there exists a common parent i' of i and k as well as a common parent j' of j and k. This means that there is a path of type (e) between these vertices which is $*$-connecting.

(d) Suppose π_G has the form of path (d) in Figure 2. We consider the different cases that can arise based on the cases we had when π_G had the form (c) and analyze how adding the edge between j and j' affects these cases.

- Suppose (1) holds for the edge $i \to k$ and (2) holds for the edge $j' \to k$. Regardless of the status of the edge $j' \to j$, we know from the previous case that the graph H is missing an unshielded collider which contradicts $G \sim_d H$. Just as in the previous case, the same argument applies to the case where (2) holds for both edges and the case where (2) and (3) hold for the edges.

- Now suppose that any of the other cases hold. Then we know there is a $*$ connecting path between j' and i of type (c), (d), or (e). Furthermore, there is a $*$-connecting path between j and j' of type (a) or (b). If the path between j' and i is of type (c) then taking the union of this path with any of the configurations of the path between j and j' gives a $*$-connecting path between i and j of type (c), (d), or (e) still. If the path between j' and i is of type (e) and the path between j and j' is of type (b) then j' will be a collider in the union of these paths. This path has the form $j \leftarrow m \to j' \leftarrow \ell \to k \leftarrow \ell' \to i$. Again though by the proof of Lemma 3.3, we know that the parents of j' must be moral since j' is not a collider on π_G and $G \sim_d H$. This means that we have an edge $m \to \ell$ and thus the path $j \leftarrow m \to k \leftarrow \ell \to i$ which is a path of type (d) between i and j. The remaining configurations follow in the exact same way.

(e) Suppose π_G has the form of path (e) in Figure 2. This case follows in a very similar way to the type (d) case. We can again rule out all of the cases where k is no longer a collider. The remaining cases then follow from the same argument used previously. We know that j and i' will have a $*$-connecting path between them of type (c), (d), or (e) by the previous case. We also know from the previous case that taking the union of this path with any of the possible configurations of the path between i' and i also gives a $*$-connecting path.

This completes the proof since the desired path exists in every case. □

Remark 3.5. *The equality of equivalence classes in Theorem 3.4 holds for generic coefficient matrices C supported on G. For special choices of C, there may be more valid CI statements.*

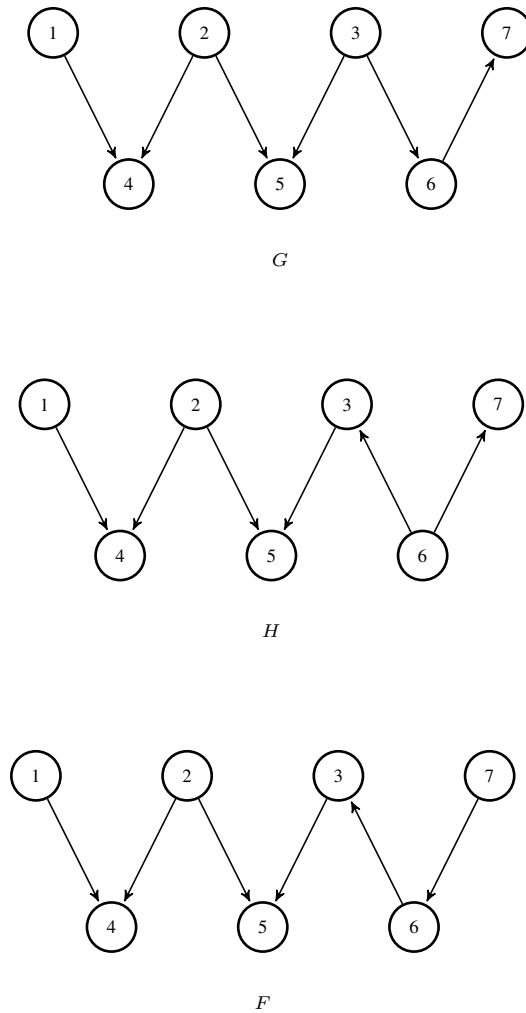The following example illustrates Theorem 3.4.

G

H

F

Figure 4: The graphs G, H, and F used in Example 3.6. These three graphs form a Markov equivalence class with respect to d- and $*$-separation.

Example 3.6. *Let G be the graph pictured in Figure 4. Let H be obtained from G by reversing the edge $3 \to 6$ and let F be obtained by reversing the edge $6 \to 7$ in H. Then the graphs G, H and F form a Markov equivalence class under d-separation since each of the reversals of edges does not change the unshielded colliders but altering any other edges would. It is also straightforward to check that any additional conditional independence statements that come from applying the $*$-separation criterion to G are also valid for H and F. For example, suppose $K = \{4, 5\}$ and observe that in all three conditional reachability graphs there is a unique path between 1 and 6 in which there are always two colliders and thus cannot be $*$-connecting. So the additional statements that $*$-separation gives do not refine this Markov equivalence class for d-separation and thus we see this is also a Markov equivalence class with respect to $*$-separation.*

Note that as a corollary of Theorem 3.4, the problem of counting the number of Markov equivalence classes for max-linear Bayesian networks is equivalent to the classical one, as studied in Gillispie and Perlman [2001], Radhakrishnan et al. [2017, 2018].

4 TROPICAL TREK RULE

In this section we define a new matrix, Σ^{trop}, associated to a max-linear model that is a natural analogue of the covariance matrix for Gaussian distributions. We then show that this matrix can be obtained by tropicalizing the trek rule of Sullivant et al. [2010] which is used to parameterize the covariance matrix of a directed Gaussian graphical model. This immediately implies tropical rank constraints on Σ^{trop} which correspond to conditional independence statements for the model.

We begin with some background on directed Gaussian graphical models and the trek rule. For additional background we refer the reader to [Sullivant, 2018, Chapter 13]. Recall that a random vector X is distributed according to the directed Gaussian graphical model for a graph, G, if it satisfies the recursive structural equations $X = CX + Z$ where C is the weighted adjacency matrix of G and Z is a Gaussian random vector with diagonal covariance matrix Ω. The matrices C and Ω are the parameters of the model. The recursive structural equations have solution $X = (Id - C)^{-1}Z$ where $(Id - C)^{-1}$ plays the role that C^* plays in the max-linear model. The covariance matrix Σ admits the factorization $\Sigma = (Id - C)^{-1}\Omega(Id - C)^{-T}$ and this factorization can be used to interpret the entries of Σ combinatorially.

Definition 4.1. *A trek τ from i to j is an alternating sequence of nodes and edges of the form*

$$i \leftarrow i_l \leftarrow \cdots \leftarrow i_1 \longleftarrow i_0 \longrightarrow j_1 \to \cdots \to j_r \to j.$$

(a trek takes you up and down a 'mountain'). The top *of the trek is* $\mathrm{top}(\tau) = i_0$, *the* left-hand side *of the trek,* $\mathrm{left}(\tau) = \{i_0, i_1, \ldots, i_l, i\}$ *and the* right-hand side *of the trek is* $\mathrm{right}(\tau) = \{i_0, j_1, \ldots, j_r, j\}$. *We also allow* trivial *treks with a single node i that have* $\mathrm{left}(\tau) = \mathrm{right}(\tau) = \{i\}$.

For a trek τ with top i_0 in a DAG with edge weights given by C we can naturally define a trek monomial:

$$\tau(C, \Omega) = \omega_{i_0 i_0} \prod_{k \to l \in \tau} c_{lk}. \tag{3}$$

Proposition 4.2 (Trek Rule). *(see e.g. Sullivant et al. [2010]) Let X be distributed according to the directed Gaussian graphical model on G with parameters C and Ω. Then the covariance matrix, Σ, of X satisfies*

$$\Sigma_{ij} = \sum_{\tau \in \mathcal{T}(i,j)} \tau(C, \Omega), \qquad i, j \in [n]. \tag{4}$$
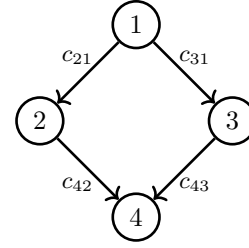

Figure 5: The Diamond graph G with its edges labeled by coefficients.

where $\mathcal{T}(i, j)$ denotes the set of all treks from i to j.

Since the structural equations of a max-linear model are given by tropicalizing the structural equations of a directed Gaussian graphical model, it is natural to consider the tropicalization of the above trek rule. If the random variable X is distributed according to the max-linear model on G with coefficient matrix C then we call $\Sigma^{\mathrm{trop}} = C^* \odot (C^*)^T$ the *tropical covariance matrix* for X where the matrix multiplication is in max-times arithmetic. This definition is motivated by the factorization that the covariance matrix of a directed Gaussian graphical model admits, as shown above.

Theorem 4.3 (Tropical Trek Rule). *Let $G = ([n], E)$ be a DAG and $\Sigma^{\mathrm{trop}} = C^* \odot (C^*)^T$ for a coefficient matrix C supported on G. Then*

$$(\Sigma^{\mathrm{trop}})_{ij} = \bigvee_{\tau \in \mathcal{T}(i,j)} \tau(C, Id). \tag{5}$$

Proof. First note that by construction c_{ij}^* is the maximum weight of any path from j to i which means

$$c_{ij}^* = \bigvee_{\pi \in \mathcal{P}(j,i)} \prod_{m \to \ell \in \pi} c_{\ell m}.$$

This means that $(C^*)_{ij}^T = c_{ji}^*$ is the maximum weight of any path from i to j. Thus we get that

$$(C^* \odot (C^*)^T)_{ij} = \bigvee_k c_{ik}^* c_{jk}^*$$

$$= \bigvee_k \left(\bigvee_{\pi \in \mathcal{P}(k,i)} \prod_{m \to \ell \in \pi} c_{\ell m} \right) \left(\bigvee_{\pi \in \mathcal{P}(k,j)} \prod_{m \to \ell \in \pi} c_{\ell m} \right)$$

Note that the last expression corresponds exactly to the trek monomial of the max-weighted trek between i and j which gives the desired result. $\square$

Example 4.4. *Let G be the Diamond graph which is pictured in Figure 5. The tropical trek rule can be used to compute the entries of the tropical covariance matrix corresponding to the max-linear model on G.*

For example, there are three treks from 2 to 4:

$$2 \to 4, \quad 2 \leftarrow 1 \to 2 \to 4 \ \text{ and } \ 2 \leftarrow 1 \to 3 \to 4.$$

Then, according to expression (5):

$$\Sigma_{24}^{\mathrm{trop}} = c_{42} \vee c_{21}^2 c_{42} \vee c_{21} c_{31} c_{43}. \qquad (6)$$

5 TROPICAL RANK CONSTRAINTS

The tropical trek rule allows us to easily show that conditional independence statements that come from d-separation correspond to tropical rank constraints on Σ^{trop}.

We first recall the following proposition which is the analogous result for Gaussians.

Proposition 5.1. *Let X be a multivariate Gaussian with covariance matrix Σ and $I, J, K \subseteq [n]$ be disjoint sets. Then the conditional independence statement $X_I \perp\!\!\!\perp X_J | X_K$ holds if and only if $\mathrm{rank}(\Sigma_{I \cup K, J \cup K}) = \#K$.*

The trek rule can also be thought of as a map that parameterizes the set of Σ that can be produced by the directed Gaussian graphical model on a DAG, G. Let $G = ([n], E)$ be a DAG and let

$$\phi_G : \mathbb{R}^E \times \mathbb{R}^n \to \mathbb{R}^{\binom{n+1}{2}}$$

be defined by

$$\phi_G(C, \Omega)_{ij} = \sum_{\tau \in \mathcal{T}(i,j)} \tau(C, \Omega).$$

The image of ϕ_G is exactly the parameterized Gaussian graphical model associated to G. From Section 5 of Sturmfels and Tevelev [2008] we have that

$$\mathrm{image}(\mathrm{trop}(\phi_G)) \subseteq \mathrm{trop}(\mathrm{image}(\phi_G)).$$

The following two results are an immediate consequence of this containment of tropical varieties.

Theorem 5.2. *Let G be a DAG and Σ^{trop} be supported on G. If K d-separates I and J in the DAG G then $\mathrm{trank}(\Sigma_{I \cup K, J \cup K}^{\mathrm{trop}}) = \#K$.*

Proof. Proposition 5.1 implies that if $\Sigma \in \mathrm{image}(\phi_G)$ every $(\#K + 1) \times (\#K + 1)$ minor of $\Sigma_{I,J}$ vanishes. Since we have that $\Sigma^{\mathrm{trop}} \in \mathrm{trop}(\mathrm{image}(\phi_G))$, the tropicalization of any polynomial that vanishes on $\mathrm{image}(\phi_G)$ will vanish on Σ^{trop}. So all of the $(\#K + 1) \times (\#K + 1)$ tropical minors of Σ^{trop} vanish and the result follows. $\square$

Example 5.3. *Consider the Diamond graph G pictured in Figure 5 which has Σ^{trop} equal to*

$$\begin{pmatrix} 1 & c_{21}^* & c_{31}^* & c_{41}^* \\ c_{21}^* & (c_{21}^*)^2 \vee 1 & c_{21}^* c_{31}^* & c_{21}^* c_{41}^* \vee c_{42}^* \\ c_{31}^* & c_{21}^* c_{31}^* & (c_{31}^*)^2 \vee 1 & c_{31}^* c_{41}^* \vee c_{43}^* \\ c_{41}^* & c_{21}^* c_{41}^* \vee c_{42}^* & c_{31}^* c_{41}^* \vee c_{43}^* & (c_{41}^*)^2 \vee (c_{42}^*)^2 \vee (c_{43}^*)^2 \vee 1 \end{pmatrix}$$

Note that the entry $\Sigma_{24}^{\mathrm{trop}}$ coincides with that computed in Equation (6) since

$$\begin{aligned} \Sigma_{24}^{\mathrm{trop}} &= c_{42} \vee c_{21}^2 c_{42} \vee c_{21} c_{31} c_{43} \\ &= c_{42}^* \vee c_{21}^* (c_{21} c_{42} \vee c_{31} c_{43}) \\ &= c_{42}^* \vee c_{21}^* c_{41}^*. \end{aligned}$$

Observe that $K = \{1\}$ d-separates $I = \{2\}$ and $J = \{3\}$ in G so the tropical rank of the submatrix $\Sigma_{I \cup K, J \cup K}^{\mathrm{trop}} = \Sigma_{\{1,2\},\{1,3\}}^{\mathrm{trop}}$ should be $\#K = 1$. More explicitly, the submatrix is

$$\Sigma_{\{1,2\},\{1,3\}}^{\mathrm{trop}} = \begin{pmatrix} 1 & c_{31}^* \\ c_{21}^* & c_{21}^* c_{31}^* \end{pmatrix}$$

and is not zero so it has tropical rank at least one. To show that it is rank is 1, we compute the tropical determinant which is

$$\det(\Sigma_{\{1,2\},\{1,3\}}^{\mathrm{trop}}) = c_{21}^* c_{31}^* \vee c_{31}^* c_{21}^*$$

Since this determinant is tropically singular, we have that $\mathrm{trank}(\Sigma_{\{1,2\},\{1,3\}}^{\mathrm{trop}}) = \#K = 1$.

While conditional independence statements that come from d-separation give tropical rank constraints on Σ^{trop}, the same is not necessarily true for those which come from $*$-separation. The following example illustrates this.

Example 5.4. *Let G be the Cassiopeia graph pictured in Figure 1 and recall that $1 \perp_* 3 | \{4, 5\}$ in G. The corresponding block of Σ^{trop} is*

$$\Sigma_{\{1,4,5\},\{3,4,5\}}^{\mathrm{trop}} = \begin{pmatrix} 0 & c_{41} & 0 \\ 0 & c_{41}^2 \vee c_{42}^2 \vee 1 & c_{42} c_{52} \\ c_{53} & c_{42} c_{52} & c_{52}^2 \vee c_{53}^2 1 \end{pmatrix}.$$

Observe that the tropical determinant of this submatrix is

$$\det(\Sigma_{\{1,4,5\},\{3,4,5\}}^{\mathrm{trop}}) = 0 \vee 0 \vee 0 \vee c_{41} c_{42} c_{52} c_{53} \vee 0 \vee 0$$

which is not tropically singular for any choice of C. This means for every C supported on G, the tropical rank of this submatrix is 3 so this conditional independence statement does not correspond to a drop in tropical rank.

It is worth mentioning that, in general, computing the tropical rank is *NP-Hard* [Shitov, 2014]. However, there exist recent approximation algorithms [Karaev and Miettinen, 2019].

A natural question is if Theorem 5.2 can be used for structure learning of a graph. A first obstacle is access to the tropical covariance matrix Σ^{trop}. Unlike the classical case, there is no known estimator for this matrix from data (see Question 6.2).

Nevertheless, we could assume oracle access to Σ^{trop} and apply the PC algorithm to try recover a max-linear graph.

However, the PC algorithm can fail because the distribution is in general not *faithful*. This means that there exist valid CI statements that do not follow from d-separation. We illustrate with an example.

Example 5.5. *Consider the Diamond graph from Figure 5, and assume that the matrix C satisfies $c_{42}c_{21} < c_{31}c_{43}$ with $c_{31} > 1$. Then we have that*

$$\Sigma^{\text{trop}}_{\{1,3\},\{3,4\}} = \begin{pmatrix} c_{31}^* & c_{41}^* \\ (c_{31}^*)^2 \vee 1 & c_{31}^* c_{41}^* \vee c_{43}^* \end{pmatrix}$$

has tropical rank 1 because

$$c_{31}^*(c_{31}^* c_{41}^* \vee c_{43}^*) = c_{41}^*((c_{31}^*)^2 \vee 1) = (c_{31}^*)^2 c_{41}^*.$$

While it is true that $1 \perp\!\!\!\perp 4 \,|\, 3$ in this scenario, this CI statement cannot be concluded from d-separation, since it is not true that $K = \{3\}$ d-separates $I = \{1\}$ and $J = \{4\}$ in the Diamond graph.

6 OPEN PROBLEMS

In this section we describe some open problems surrounding max-linear models with a particular emphasis on conditional independence.

Our original inspiration for considering Σ^{trop} was its similarity to the tail-dependence matrix χ defined in [Sibuya, 1960]. *Conditional tail dependence* is the extreme value analogue of correlation. Gissibl and Klüppelberg [2018] showed that if X is distributed according to a max-linear model on a DAG G with Fréchet(α) innovations Z_i then the tail dependence between X_i and X_j can be computed in the following way. First define the normalized coefficient matrix $\overline{C}$ with entries

$$\overline{c}_{ij} = \frac{(c_{ij}^*)^\alpha}{\sum_{k \in \text{An}(j)} (c_{kj}^*)^\alpha}$$

then the tail dependence between X_i and X_j is

$$\chi(i,j) = \sum_{k \in \text{An}(i) \cap \text{An}(j)} \overline{c}_{ki} \wedge \overline{c}_{kj}.$$

Since tail dependence is a popular measure of dependence in extreme value theory and the matrix χ can be estimated directly from data, it would be interesting to determine if something analogous to Theorem 5.2 holds for χ. If a relationship like this could be found, then more tools from algebraic geometry and tropical geometry could be used to study max-linear models just as algebraic geometry has been used to study Gaussian Bayesian networks. This motivates the following problem.

Problem 6.1. *Determine if conditional independence statements that hold for the max-linear model X correspond to an algebraic or tropical algebraic constraint on the tail dependence matrix χ.*

It would be interesting to determine other information that Σ^{trop} encodes. We have shown that it satisfies tropical rank constraints similar to those for Gaussians but it would be more helpful if Σ^{trop} had a direct interpretation in terms of the underlying max-linear model or could be determined from data like χ (without having to estimate C^*). This leads us to the following question.

Question 6.2. *Is there a consistent estimator for Σ^{trop}?*

Lastly, we note that our rank constraints on Σ^{trop} only correspond to conditional independence statements that come from d-separation. It would be interesting to determine if the conditional independence statements that come $*$-separation can also be interpreted as an algebraic constraint on Σ^{trop} or a related matrix such as χ.

Problem 6.3. *Suppose K $*$-separates I and J in the DAG G so $X_I \perp\!\!\!\perp X_J | X_K$ for X distributed according to a max-linear model on G. Determine if this conditional independence statement corresponds to an algebraic or tropical algebraic constraint on the matrix Σ^{trop}.*

In fact, in Example 5.5 we see that despite the CI statement $1 \perp\!\!\!\perp 4 \,|\, 3$ not being implied by d-separation, and only by $*$-separation, a tropical rank constraint still holds. However, we see in Example 5.4 there is a $*$-separation statement that does not correspond to a drop in tropical rank. This suggests a complex relationship between tropical rank and $*$-separation.

Finally, it may also be interesting to consider an analogue of Problem 6.3 when the coefficient matrix C is fixed. This means additional $*$-separation statements might hold [Améndola et al., 2021].

References

Carlos Améndola, Claudia Klüppelberg, Steffen Lauritzen, and Ngoc Tran. Conditional independence in max-linear Bayesian networks. *Annals of Applied Probability*, to appear, 2021.

Steen A. Andersson, David Madigan, and Michael D. Perlman. A characterization of Markov equivalence classes for acyclic digraphs. *Annals of Statistics*, 25(2):505–541, 1997.

Peiman Asadi, Anthony C Davison, Sebastian Engelke, et al. Extremes on river networks. *Annals of Applied Statistics*, 9(4):2023–2050, 2015.

Johannes Buck and Claudia Klüppelberg. Recursive max-linear models with propagating noise. *arXiv:2003.00362*, 2020.

Sebastian Engelke and Adrien S Hitz. Graphical models for extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(4):871–932, 2020.

Dan Geiger, Thomas Verma, and Judea Pearl. d-separation: From theorems to algorithms. In *Machine Intelligence and Pattern Recognition*, volume 10, pages 139–148. Elsevier, 1990a.

Dan Geiger, Thomas Verma, and Judea Pearl. Identifying independence in Bayesian networks. volume 20, pages 507–534. 1990b. Special issue on influence diagrams.

Steven B Gillispie and Michael D Perlman. Enumerating markov equivalence classes of acyclic digraph models. In *Proceedings of the Seventeenth Annual Conference on Uncertainty in Artificial Intelligence*, pages 171–177, 2001.

Nadine Gissibl. *Graphical modeling of extremes*. PhD thesis, Technische Universität München, 2018.

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

Nadine Gissibl, Claudia Klüppelberg, and Steffen Lauritzen. Identifiability and estimation of recursive max-linear models. *Scandinavian Journal of Statistics*, 48(1):188–211, 2021.

Anja Janßen and Phyllis Wan. k-means clustering of extremes. *Electronic Journal of Statistics*, 14(1):1211–1233, 2020.

Sanjar Karaev and Pauli Miettinen. Algorithms for approximate subtropical matrix factorization. *Data Mining and Knowledge Discovery*, 33(2):526–576, 2019.

Claudia Klüppelberg and Mario Krali. Estimating an extreme bayesian network via scalings. *Journal of Multivariate Analysis*, 181:104672, 2021.

Claudia Klüppelberg and Steffen Lauritzen. Bayesian networks for max-linear models. In *Network Science*, pages 79–97. Springer, Cham, 2019.

Steffen L Lauritzen. *Graphical models*, volume 17. Clarendon Press, 1996.

Diane Maclagan and Bernd Sturmfels. *Introduction to tropical geometry*, volume 161 of *Graduate Studies in Mathematics*. American Mathematical Society, 2015.

Christopher Meek. Strong completeness and faithfulness in Bayesian networks. In *Uncertainty in artificial intelligence (Montreal, PQ, 1995)*, pages 411–418. Morgan Kaufmann, San Francisco, CA, 1995.

Judea Pearl and Thomas S Verma. A theory of inferred causation. In *Studies in Logic and the Foundations of Mathematics*, volume 134, pages 789–811. Elsevier, 1995.

Adityanarayanan Radhakrishnan, Liam Solus, and Caroline Uhler. Counting Markov equivalence classes by number of immoralities. In *Proceedings of the Thirty Third Annual Conference on Uncertainty in Artificial Intelligence*, 2017.

Adityanarayanan Radhakrishnan, Liam Solus, and Caroline Uhler. Counting Markov equivalence classes for DAG models on trees. *Discrete Applied Mathematics*, 244:170–185, 2018.

Yaroslav Shitov. The complexity of tropical matrix factorization. *Advances in Mathematics*, 254:138–156, 2014.

Masaaki Sibuya. Bivariate extreme statistics. I. *Ann. Inst. Statist. Math. Tokyo*, 11:195–210, 1960. ISSN 0020-3157. doi: 10.1007/bf01682329.

Peter Spirtes, Clark N Glymour, Richard Scheines, and David Heckerman. *Causation, prediction, and search*. MIT press, 2000.

Bernd Sturmfels and Jenia Tevelev. Elimination theory for tropical varieties. *Math. Res. Lett.*, 15(3):543–562, 2008. ISSN 1073-2780.

Seth Sullivant. *Algebraic statistics*, volume 194 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2018.

Seth Sullivant, Kelli Talaska, and Jan Draisma. Trek separation for Gaussian graphical models. *Annals of Statistics*, 38(3):1665–1685, 2010.

Ngoc Mai Tran. Polytropes and tropical eigenspaces: cones of linearity. *Discrete & Computational Geometry*, 51(3):539–558, 2014.

Ngoc Mai Tran, Johannes Buck, and Claudia Klüppelberg. Causal discovery of a river network from its extremes. *arXiv:2102.06197*, 2021.

Thomas Verma and Judea Pearl. Equivalence and synthesis of causal models. In *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, pages 255–270, 1990.

Thomas Verma and Judea Pearl. An algorithm for deciding if a set of observed independencies has a causal explanation. In *Proceedings of the Eighth Annual Conference on Uncertainty in Artificial Intelligence*, pages 323–330, 1992.