CONDITIONAL INDEPENDENCE IN MAX-LINEAR BAYESIAN NETWORKS

BY CARLOS AMÉNDOLA[1,*], CLAUDIA KLÜPPELBERG[1,†], STEFFEN LAURITZEN[2] AND NGOC M. TRAN[3]

[1]*Center for Mathematical Sciences, Technical University of Munich,* *carlos.amendola@tum.de; †cklu@tum.de*

[2]*Department of Mathematical Sciences, University of Copenhagen, lauritzen@math.ku.dk*

[3]*Department of Mathematics, University of Texas at Austin, ntran@math.utexas.edu*

Motivated by extreme value theory, max-linear Bayesian networks have been recently introduced and studied as an alternative to linear structural equation models. However, for max-linear systems the classical independence results for Bayesian networks are far from exhausting valid conditional independence statements. We use tropical linear algebra to derive a compact representation of the conditional distribution given a partial observation, and exploit this to obtain a complete description of all conditional independence relations. In the context-specific case, where conditional independence is queried relative to a specific value of the conditioning variables, we introduce the notion of a *source DAG* to disclose the valid conditional independence relations. In the context-free case, we characterize conditional independence through a modified separation concept, ∗-separation, combined with a tropical eigenvalue condition. We also introduce the notion of an *impact graph*, which describes how extreme events spread deterministically through the network and we give a complete characterization of such impact graphs. Our analysis opens up several interesting questions concerning conditional independence and tropical geometry.

1. Introduction. Max-linear graphical models were introduced in [15] to model causal dependence between extreme events. The underlying graphical structure of the model is a directed acyclic graph (DAG) and to emphasize this aspect, we shall here use the term *max-linear Bayesian network*, to allow for generalizations and extensions (see Section 7.2 at the end of this paper).

A *max-linear Bayesian network* is specified by a random vector $X = (X_1, \ldots, X_d)$, a directed acyclic graph $\mathcal{D} = (V, E)$ with nodes $V = \{1, \ldots, d\}$, nonnegative *edge weights* $c_{ij} \geq 0$ for $i, j \in V$ and independent positive random variables $Z_1, \ldots, Z_d$. These, known as *innovations*, have support $\mathbb{R}_> := (0, \infty)$ and have atom-free distributions. Then X is specified by a *recursive system of max-linear structural equations* as

$$(1.1) \qquad X_i = \bigvee_{j \in \mathrm{pa}(i)} c_{ij} X_j \vee Z_i, \quad i = 1, \ldots, d.$$

Without loss of generality, we assume that the basic probability space is $\Omega = \mathbb{R}_>^V$ equipped with the standard Borel σ-algebra, so all randomness in the system originates from the innovations $Z_1, \ldots, Z_d$. The equation system (1.1) has solution

$$(1.2) \qquad X_i = \bigvee_{j \in \mathrm{an}(i) \cup \{i\}} c_{ij}^* Z_j, \quad i = 1, \ldots, d,$$

where $\mathrm{an}(i)$ denotes the set of nodes j where there is a directed path from j to i, and c_{ij}^* is a maximum taken over all the products along such paths (see [15], Theorem 2.2). Any such
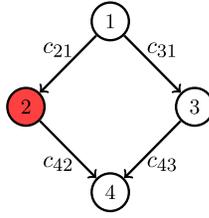
FIG. 1. *Diamond graph with the set $K = \{2\}$ being observed, as indicated by red color. If $c_{42}c_{21} \geq c_{43}c_{31}$, it holds that $X_1 \perp\!\!\!\perp X_4 \mid X_2$.*

path that realizes this maximum is called *critical* (max-weighted under C). The *max-linear coefficient matrix* $C^* = (c_{ij}^*)$ is also known from tropical linear algebra as the *Kleene star* of $C = (c_{ij})$; cf. (2.4) below.

In [21], it was observed that the conditional independence properties for max-linear Bayesian networks are very different from those in a standard Bayesian network. In particular, they are often not faithful to their underlying DAG $\mathcal{D}$. This means that the usual d-separation criterion ([13]) on the DAG typically will not identify all valid conditional independence relations, in contrast to the situation for most Bayesian networks based on discrete random variables or linear structural equations. Example 1.1 below gives a simple example of this phenomenon.

EXAMPLE 1.1 (Diamond).   Consider the DAG in Figure 1. The path $1 \to 2 \to 4$ is critical if and only if $c_{42}c_{21} \geq c_{43}c_{31}$. If this is the case, the joint distribution of (X_1, X_2, X_4) has the representation

$$X_1 = Z_1, \qquad X_2 = c_{21}X_1 \vee Z_2,$$

and

$$
\begin{aligned}
X_4 &= c_{42}X_2 \vee Z_4 \vee c_{43}X_3 \\
&= c_{42}(Z_2 \vee c_{21}Z_1) \vee Z_4 \vee c_{43}(Z_3 \vee c_{31}Z_1) \\
&= c_{42}Z_2 \vee c_{42}c_{21}Z_1 \vee Z_4 \vee c_{43}Z_3 \vee c_{43}c_{31}Z_1 \\
&= c_{42}Z_2 \vee c_{42}c_{21}Z_1 \vee Z_4 \vee c_{43}Z_3 \quad \text{since } c_{42}c_{21} \geq c_{43}c_{31} \\
&= c_{42}X_2 \vee Z_4 \vee c_{43}Z_3
\end{aligned}
$$

and hence we have $X_1 \perp\!\!\!\perp X_4 \mid X_2$ which does *not* follow from the d-separation criterion. Here, the fact that $1 \to 2 \to 4$ is *critical* renders the path $1 \to 3 \to 4$ unimportant for the conditional independence $X_1 \perp\!\!\!\perp X_4 \mid X_2$, even if $1 \to 3 \to 4$ were also critical (i.e., even if $c_{42}c_{21} = c_{43}c_{31}$).

In Example 1.1, the complicating issue was associated with paths being critical or not. However, this is not the only way standard d-separation fails. In Example 1.2 below, the complications are associated with double colliders along a path.

EXAMPLE 1.2 (Cassiopeia).   We shall show later (see Example 4.2) that a max-linear Bayesian network on the graph in Figure 2 will satisfy $X_1 \perp\!\!\!\perp X_3 \mid X_{\{4,5\}}$ for all coefficient matrices C. However, this conditional independence statement does *not* follow from the d-separation criterion since the path from 1 to 3 is d-connecting relative to $\{4, 5\}$.
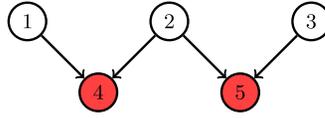
FIG. 2.   *The Cassiopeia graph with observed nodes $K = \{4, 5\}$. Here, it holds that $X_1 \perp\!\!\!\perp X_3 \mid X_{\{4,5\}}$.*

Example 1.2 shows that not only are max-linear Bayesian networks often not faithful to d-separation, but d-separation is also not *complete* in the sense of [13] for conditional independence in these networks. That is, there are conditional independence statements which are valid for any choice of coefficients C, but cannot be derived from d-separation.

Also, in contrast to standard results for Bayesian networks, some conditional independence relations are highly *context-specific*, that is, depend drastically on the particular values of the conditioning variables, as in Example 1.3 below. To control this, we introduce the notion of a *source DAG* $\mathcal{C}(X_K = x_K)$ for a given *context* $\{X_K = x_K\}$; see Definition 3.16 for details.

EXAMPLE 1.3 (Tent).   Consider the DAG $\mathcal{D}$ to the left in Figure 3 with all edge weights $c_{ij} = 1$. Let $K = \{4, 5\}$ be the set of observed nodes; we seek all independence relations conditionally valid in the context $X_4 = X_5 = 2$. Writing out the model (1.1), we find

$$X_1 = Z_1, \qquad X_2 = Z_2, \qquad X_3 = Z_3 \vee X_1 \vee X_2,$$
$$X_4 = Z_4 \vee X_1 \vee X_2 = 2,$$
$$X_5 = Z_5 \vee X_1 \vee X_2 = 2.$$

Since $Z_1, \ldots, Z_5$ are a.s. different when the innovations have atom-free distributions, it holds apart from a null-set that $X_1 \vee X_2 = Z_1 \vee Z_2 = 2$. This introduces bounds on the innovations; we must have $Z_1, Z_2, Z_4, Z_5 \leq 2$ and it also holds that $X_3 \geq 2$. Further, we then have

$$X_1 = Z_1, \qquad X_2 = Z_2, \qquad X_1 \vee X_2 = 2, \qquad X_3 = Z_3 \vee 2,$$
$$X_4 = Z_4 \vee 2 = 2,$$
$$X_5 = Z_5 \vee 2 = 2,$$

whence we conclude that $X_3 \perp\!\!\!\perp (X_1, X_2) \mid X_4 = X_5 = 2$, since now the dependence of X_3 on X_1, X_2 has disappeared. This independence statement is reflected in the lack of edges $1 \to 3$ and $2 \to 3$ in the source DAG $\mathcal{C}(X_4 = X_5 = 2)$, shown to the right in Figure 3.

In this paper, we give a complete description of valid conditional independence statements for a given matrix C, conditional independence statements that hold for all C supported on

$\mathcal{D}$

$\mathcal{C}(X_K = x_K)$

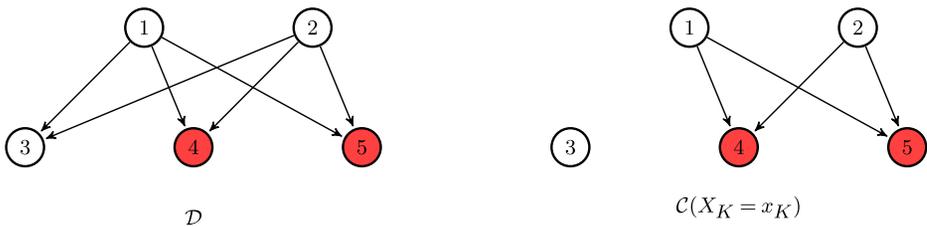FIG. 3.   *The left-hand figure displays what we shall name the tent DAG $\mathcal{D}$. For all coefficients equal to 1, the source DAG $\mathcal{C}(X_K = x_K)$ when the observed nodes are $K = \{4, 5\}$ with observed values $x_4 = x_5 = 2$, is obtained from the left-hand figure by removing the edges $1 \to 3$ and $2 \to 3$, which become redundant in the context $\{X_4 = X_5 = 2\}$; see Section 3.3.*

a given DAG $\mathcal{D}$, as well as those that depend on the specific values of the conditioning variables. We achieve this by introducing three separation criteria. These are less restrictive than d-separation, as they focus on paths that are *critical* (see Example 1.1), do not have *multiple colliders* (see Example 1.2), and, for a given context, refer to the source DAG, obtained by removing edges that are *redundant* in the context (see Example 1.3).

Before we state and prove results for conditional independence, we investigate how extreme events at selected nodes spread through the network. We define an *impact graph* as a random graph on V, containing the edge $j \to i \iff X_i = c_{ij}^* Z_j$, that is, if X_i is realized (determined) by Z_j (see Definition 3.1). Since the distributions of the innovations are atom-free, it holds with probability one that any node i has at most one parent in such a graph. We give a complete description of all impact graphs with positive probability in Theorem 3.3. As we shall see in Remark 2, the impact graphs index partitions of the innovation space into regions of linearity for the max-linear map in (1.2).

Impact graphs can be compatible with a context $\{X_K = x_K\}$ or not, and vice versa, a context $\{X_K = x_K\}$ can be possible under a certain impact graph or not. For instance, for the Cassiopeia graph in Example 1.2, the possible impact graphs are: the empty graph, all subgraphs with a single edge, and the four subgraphs with two edges displayed in Figure 4. On the other hand, the impact graph g_2 implies that $X_4 > X_5$, so only events satisfying this restriction are possible under g_2.

The union of all impact graphs compatible with a context $\{X_K = x_K\}$ describes all possible ways that an extreme innovation could spread across the network while conforming with the context. However, as seen in Example 1.3, the given context can cause max-linear combinations of variables to be constant under specific scenarios, such that they do not influence the distribution of random variables X_v, $v \notin K$ as expected. This effect is taken care of by the removal of edges to yield the *source DAG $\mathcal{C}(X_K = x_K)$* compatible with the context.

Moreover, we classify all nodes into nonconstant nodes (active) and constant nodes with specific properties (see Proposition 3.18). This classification plays an important role when modifying the solution in (1.2) to obtain a compact representation of the conditional distribution.

This compact representation is given in Theorem 4.3 and can be seen as a version of Theorem 6.7 in [15] and of Theorem 1 in [29]. More precisely, [29] studies a general max-linear model where the max-linear coefficient matrix C^* is not necessarily the Kleene star of
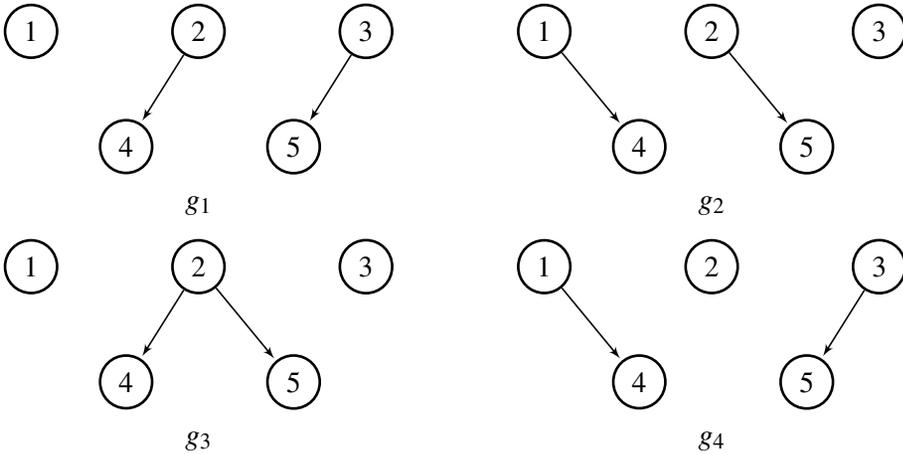

FIG. 4. *The four impact graphs with two edges in the Cassiopeia graph of Example 1.2 as depicted in Figure 2. Suppose all coefficients are equal to one. Then, only the impact graphs g_2 and g_4 are compatible with the context $\{X_4 = 3, X_5 = 2\}$, whereas only the impact graph g_3 is compatible with the context $\{X_4 = X_5 = 2\}$; see Example 3.9 below.*

a max-linear Bayesian network, hence not necessarily idempotent; they further give a more detailed description of the conditional distribution, using a collection of *hitting scenarios*, describing specific elements of Z, which obtain their upper bounds. An important endeavor of the present article is to further identify characteristics of the hitting scenarios, exploiting the graphical structure of the model, and this is done in Theorem 4.3.

We formulate three different theorems to clarify conditional independence for max-linear Bayesian networks. All three have the following structure, using what we shall term $*$-separation ($\perp_*$) in appropriate derived DAGs.

THEOREM. *Let X be a max-linear Bayesian network over a directed acyclic graph $\mathcal{D} = (V, E)$. Then for all $I, J, K \subseteq V$,*

$$I \perp_* J \mid K \quad in \ \widetilde{\mathcal{D}} \quad \implies \quad X_I \perp\!\!\!\perp X_J \mid X_K.$$

The DAG $\widetilde{\mathcal{D}}$—derived from $\mathcal{D}$, C, and the specific *context* $\{X_K = x_K\}$—depends on the situation and we distinguish the following three: Theorem 5.12 refers to a fixed C and also a specific context $\{X_K = x_K\}$, thus yielding conditional independence relations that are valid for the particular values x_K; Theorem 5.14 considers a fixed coefficient matrix C and yields independence relations that may depend on C, which are valid for all possible contexts; whereas in Theorem 5.15, the coefficient matrix C is arbitrary with support included in $\mathcal{D}$ and this yields conditional independence relations that are universally valid under these conditions. In all three scenarios, the derived DAG $\widetilde{\mathcal{D}}$ is different, and the $*$-separation has to be considered in this derived DAG. In addition, we give conditions for these criteria to be *complete* in the sense of [13], that is, they yield all conditional independence statements that are valid under the specified conditions. In Section 6, we investigate in which sense these results are as strong as possible. For a fixed matrix C, Theorem 5.14 may not yield all conditional independence results, as specific relations between the coefficients could imply additional independences, similar to what is known as path cancellations for linear Bayesian networks. We provide the tools to identify exactly what these are in Theorem 6.18.

The paper is organized as follows. In Section 2, we introduce basic concepts and notation. We define the impact graphs, describing how effects of extremes spread to other variables, and the source DAG, describing the possible sources for a given value of observations, in Section 3. Section 4 is devoted to deriving a compact representation of the conditional distribution and the conditional independence results are stated and proved in Section 5. Section 6 is devoted to the discussion of completeness. We conclude by indicating potential future work and research directions in Section 7.

2. Preliminaries.

2.1. *Graph terminology.* We use the same graph notation as in [21]. A *directed graph* is a pair $g = (V, E)$ of a *node set* $V = \{1, \ldots, d\}$ and *edge set* $E = \{j \to i : i, j \in V, i \neq j\}$. An edge $j \to i$ points *from j to i*, with j called a *parent* of i and i is a *child* of j. In a graph g, the set of parents of i is $\mathrm{pa}_g(i)$ and the set of children of i is $\mathrm{ch}_g(i)$. A *path* from j to i of *length* $n \geq 1$ is a sequence of distinct nodes $[j = k_0, k_1, \ldots, k_n = i]$ such that $k_{r-1} \to k_r \in E$ or $k_r \to k_{r-1} \in E$ for all $r = 1, \ldots, n$ and we say that i and j are *connected*. We also say that the path is *between i and j*. A graph is *connected* if there is a path between any two vertices.

A *directed path* from j to i has $k_{r-1} \to k_r \in E$ for all r. If there is a directed path from j to i in g, we say that j is an *ancestor* of i and i a *descendant* of j. A *directed cycle* is a directed path with the modification that $i = j$. A directed acyclic graph (abbreviated DAG) is a directed graph with no directed cycles. A DAG is *well ordered* if all edges point from low

to high, that is, $j \to i \implies j < i$. A connected DAG is a *tree* if every node has at most one parent. The *root* of a tree is the unique node in the tree without parents. The *height* of a tree is the length of the longest directed path in the tree. A *forest* is a collection of trees. A *star* is a tree of height at most one, and we call a forest of stars a *galaxy*. For a forest g on node set V and $i \in V$, we let $R_g(i)$ denote the root of the tree containing i and $R(g)$ denotes the set of roots in g. A matrix $A \in \mathbb{R}_\geq^{d \times d}$ defines a *weighted directed graph* $\mathcal{D}(A)$, where $j \to i \in \mathcal{D}(A)$ if and only if its *edge weight* $a_{ij} > 0$. The *weight of a path* π in $\mathcal{D}(A)$ is then the product of its edge weights.

2.2. *Tropical linear algebra.*   A number of theorems in our paper are proved using techniques from tropical linear algebra. Here, we recall some essential facts of this field. For a comprehensive text, we recommend [5] and [8]; see also [18] and [26].

Tropical linear algebra is linear algebra with arithmetic in the *max-times semiring* $(\mathbb{R}_\geq, \vee, \odot)$, defined by

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

Note that many authors (including those above) use the isomorphic semirings max-plus or min-plus, but we have chosen max-times to conform with the literature on extreme value theory. The operations extend to $\mathbb{R}_\geq^d$ coordinatewise and to corresponding matrix multiplication for $A \in \mathbb{R}_\geq^{m \times n}$ and $B \in \mathbb{R}_\geq^{n \times p}$ as

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

and we also write

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

For example,

$$(2.1) \qquad \begin{pmatrix} 1 & 8 \\ 2 & 3 \end{pmatrix} \odot \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \cdot 2 \vee 8 \cdot 1 \\ 2 \cdot 2 \vee 3 \cdot 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 4 \end{pmatrix} = 4 \odot \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

The recursive structural equation system (1.1) can be written as a tropically linear equation:

$$(2.2) \qquad X = C \odot X \vee Z,$$

where $C = (c_{ij}) \in \mathbb{R}_\geq^{d \times d}$ and $X, Z \in \mathbb{R}_>^d$. We consider the *weak transitive closure* ([8], Section 1.6.2) $\Gamma = \Gamma(C) = (\gamma_{ij})$ of C given as

$$(2.3) \qquad \Gamma = \Gamma(C) = \bigvee_{k=1}^{d-1} C^{\odot k}.$$

Here, $\gamma_{ij} > 0$ if and only if there exists a directed path in $\mathcal{D}(C)$ from j to i, and γ_{ij} equals the maximum weight over all such paths. We name $\mathcal{D}^*(C)$ the *weighted reachability DAG* of $\mathcal{D}(C)$ and $\mathcal{D}^*$ the unweighted counterpart. When $\mathcal{D}(C)$ is a DAG, by [5], Theorem 3.17, (2.2) can be solved uniquely for X as

$$(2.4) \qquad X = C^* \odot Z,$$

where $C^* = I_d \vee \Gamma(C)$ is the *Kleene star* of C and I_d the $d \times d$ identity matrix. Since Kleene stars are idempotent, that is, $C^* \odot C^* = C^*$, we also have

$$(2.5) \qquad X = C^* \odot X.$$

If V is well ordered, the matrix C is lower triangular and so are Γ and C^*. The Kleene star C^* corresponds to the max-linear coefficient matrix B in [15], [21] and in particular, [15], Theorem 2.2, is a special instance of [5], Theorem 3.17. For $K \subseteq V$, we let

$$(2.6) \qquad \mathcal{L}_K^C = \left\{ x_K : \exists z \in \mathbb{R}_>^V \text{ with } x_K = (C^* \odot z)_K \right\}$$

denote the image of the projection to K-coordinates of the max-linear map determined by C^*. A matrix $A = (a_{ij}) \in \mathbb{R}_\geq^{d \times d}$ has *tropical eigenvalue* λ and *tropical eigenvector* $x \in \mathbb{R}_\geq^d$ if

$$(2.7) \qquad A \odot x = \lambda \odot x.$$

For example, equation (2.1) shows that $x = (2, 1)$ is a tropical eigenvector for the given matrix, with tropical eigenvalue $\lambda = 4$.

The maximum geometric mean of weights along a directed cycle in $\mathcal{D}(A)$ is the *maximum cycle mean* of A, denoted $\lambda(A)$. Note that if $\mathcal{D}(A)$ is acyclic then $\lambda(A) = 0$. For any matrix A, the number $\lambda(A) \geq 0$ is always a tropical eigenvalue, called the *principal eigenvalue* of A ([8], Theorem 4.2.4). A cycle achieving the maximum mean is a *critical* cycle. Similarly, a vector $x \in \mathbb{R}_>^d$ is called a *tropical subeigenvector of A* for $\lambda > 0$ if

$$(2.8) \qquad A \odot x \leq \lambda \odot x.$$

The following fact about tropical sub-eigenvectors will be useful.

PROPOSITION 2.1. *Let $A \in \mathbb{R}_\geq^{V \times V}$. We then have*:

(a) *There exists $x \in \mathbb{R}_>^V$ such that $A \odot x \leq x$ if and only if $\lambda(A) \leq 1$.*
(b) *Suppose $\lambda(A) = 1$, $S \subseteq V$ is the union of the support of its critical cycles, and that $x \in \mathbb{R}_>^V$ satisfies $A \odot x \leq x$. Then $A_{SS} \odot x_S = x_S$.*
(c) *There exists $x \in \mathbb{R}_>^V$ such that $A \odot x < x$ if and only if $\lambda(A) < 1$.*

PROOF. Statement (a) is shown in [8], Theorem 1.6.18. Now we prove (b). First, consider the case $S = V$. Let x be such that $A \odot x \leq x$. Fix any $i \in V$. Then i belongs to some critical cycle σ of length r that achieves the tropical eigenvalue. For each edge $v \to u$ in this cycle, $A \odot x \leq x$ implies

$$(2.9) \qquad a_{uv} x_v \leq x_u.$$

Therefore,

$$\prod_{v \to u \in \sigma} a_{uv} \leq \prod_{v \to u \in \sigma} \frac{x_u}{x_v}$$
$$= 1 \quad \text{since } \sigma \text{ is a cycle.}$$

But σ is critical, so

$$\prod_{v \to u \in \sigma} a_{uv} = c(\sigma) = \lambda^r = 1.$$

Thus all the inequalities in (2.9) must be equalities; that is, $a_{uv} x_v = x_u$ for all nodes u, v in the support of σ. In particular, this holds for $u = i$. Thus, for the edge $v \to i \in \sigma$,

$$x_i \geq (A \odot x)_i \geq a_{iv} x_v = x_i.$$

So $(A \odot x)_i = x_i$. Since i was chosen arbitrarily, it follows that $A \odot x = x$. Now suppose $S \subset V$. Let $\bar{S} = V \backslash S$. Then

$$x_S \geq (A \odot x)_S = A_{SS} \odot x_S \vee A_{S\bar{S}} \odot x_{\bar{S}} \geq A_{SS} \odot x_S.$$

Since $\lambda(A_{SS}) = 1$, applying the previous argument to A_{SS} gives $A_{SS} \odot x_S = x_S$.

Now we prove (c). Suppose $\lambda(A) < 1$. Let x be an associated eigenvector to the principal eigenvalue of A. Then $A \odot x = \lambda(A)x < x$. For the converse, if $A \odot x < x$ it also satisfies $A \odot x \leq x$ so by (a) we have $\lambda(A) \leq 1$. If $\lambda(A) = 1$ then by (b), there exists some $S \subseteq V$, $|S| \geq 2$, such that $A_{SS} \odot x_S = x_S$. But then

$$x_S = A_{SS} \odot x_S \leq (A \odot x)_S < x_S,$$

a contradiction. Thus we conclude that $\lambda(A) < 1$ and the proof is complete. $\square$

We recall one more useful fact from tropical linear algebra.

LEMMA 2.2 ([8], Lemma 1.6.19).   *Let $A \in \mathbb{R}_{\geq}^{V \times V}$ with $\lambda(A) = 1$ and eigenvector $x \in \mathbb{R}_{>}^{V}$. Let σ be a critical cycle in A. Then for all edges $v \to u \in \sigma$,*

$$a_{uv}x_v = x_u.$$

2.3. *Conditional independence.*   Conditional independence is concerned with probability distributions on product spaces $\mathcal{X} = \prod_{i \in V} \mathcal{X}_i$, where $\mathcal{X}_i$ are measurable spaces. For $I \subseteq V$, we write $x_I = (x_v, v \in I)$ to denote a generic element in $\mathcal{X}_I = \prod_{v \in I} \mathcal{X}_v$, and similarly $X_I = (X_v)_{v \in I}$. If $\mathbb{P}$ is a probability distribution on $\mathcal{X}$, we use the short notation

$$I \perp\!\!\!\perp J \mid K \quad \Longleftrightarrow \quad X_I \perp\!\!\!\perp X_J \mid X_K,$$

where $\perp\!\!\!\perp$ denotes probabilistic conditional independence w.r.t. $\mathbb{P}$.

Graphical models identify conditional independence relations through a *separation criterion* $\perp_\sigma$ applied to a graph. Such a separation criterion is, for example, given by *d-separation* $\perp_{\mathcal{D}}$ ([13]) for a given DAG $\mathcal{D}$; see, for example, [22], [24] or [25] for further details. A separation criterion $\perp_\sigma$ is *complete* relative to a family $\mathcal{P}$ of probability distributions if

$$I \perp\!\!\!\perp J \mid K \quad \text{for all } \mathbb{P} \in \mathcal{P} \quad \Longleftrightarrow \quad I \perp_\sigma J \mid K.$$

A probability distribution $\mathbb{P}$ of X is *faithful* to $\perp_\sigma$ if for all disjoint subsets I, J, K of V and that specific $\mathbb{P}$ it holds that

$$I \perp\!\!\!\perp J \mid K \quad \Longleftrightarrow \quad I \perp_\sigma J \mid K.$$

Thus the distribution of X is in particular *Markov* w.r.t. $\perp_\sigma$. If there exists a faithful distribution $\mathbb{P} \in \mathcal{P}$ for $\perp_\sigma$, the separation criterion is obviously complete.

3. Auxiliary graphs.   In this section, we introduce the concept of an impact graph, an impact graph compatible with a context, and a source DAG. These are devices that translate probabilistic statements to graph-theoretic and algebraic statements, and at the same time keep track of all deterministic relationships in a max-linear Bayesian network.

3.1. *The context-free impact graph.*

DEFINITION 3.1.   Consider the max-linear Bayesian network (2.2) with fixed coefficient matrix C. The (context-free) *impact graph* is a random graph $G = G(Z)$ on V consisting of the following edges:

$$j \to i \quad \Longleftrightarrow \quad X_i = c_{ij}^* Z_j$$

and we let $\mathcal{E}(g) = \{z \in \mathbb{R}_{>}^{V} : G(z) = g\}$ denote the event that $\{G = g\}$.

In the following, we let $\mathfrak{G} = \mathfrak{G}(C)$ denote the set of impact graphs that have positive probability for a given coefficient matrix C.

REMARK 1. Since the distributions of the Z_j are atom-free, it holds with probability one that any node i has at most one parent and thus if $\mathbb{P}(\mathcal{E}(g) > 0)$, that is, $g \in \mathfrak{G}$, g will be a forest. We shall only consider configurations of Z that conform with this and we emphasize that we are only ignoring a null-set in $\Omega = \mathbb{R}_{>}^{V}$.

REMARK 2. Define the restricted Kleene star C_g^* as

$$(3.1) \qquad (C_g^*)_{ij} = \begin{cases} 1 & \text{if } i = j \in R(g), \\ c_{ij}^* & \text{if } j \to i \in g, \\ 0 & \text{otherwise,} \end{cases}$$

where $R(g)$ are the set of roots in the forest g. The impact graphs induce a partition of $\mathbb{R}_{>}^{V}$ into regions where the map $Z \to X$ is linear with matrix C_g^*. In other words, we have an alternative representation of X as

$$(3.2) \qquad X = C^* \odot Z \stackrel{\text{a.s.}}{=} C_G^* \odot Z = C_G^* Z,$$

where the product in the rightmost expression is a standard linear matrix product as C_G^* has exactly one positive number in each row. See also Example 3.7 below.

The main result of this section is Theorem 3.3, which gives a precise and complete characterization of all impact graphs $\mathfrak{G}$ in a max-linear Bayesian network (2.2). To establish this characterization, we need to define the impact exchange matrix of a given forest g. Recall that $\mathrm{ch}_g(i)$ denotes the set of children of i in g (Section 2.1).

DEFINITION 3.2. Consider a DAG $\mathcal{D}$ with coefficient matrix C and Kleene star C^* and let g be a forest with root set $R = R(g)$. The *impact exchange matrix* $M(g) = M(g, C^*)$ of g with respect to C^* is an $|R| \times |R|$ matrix with entries defined by $m_{rr} = 0$ for all $r \in R$, and for $r \neq r'$:

$$(3.3) \qquad m_{rr'} := \max_{i \in \mathrm{ch}_g(r)} \frac{c_{ir'}^*}{c_{ir}^*}.$$

Note that $m_{rr'} = 0$ if $\mathrm{ch}_g(r) = \varnothing$. Finally, recall from Section 2.2 that $\mathcal{D}^*$ is the reachability DAG and $\lambda(M(g))$ is the principal eigenvalue of $M(g)$. We now have the following fundamental theorem.

THEOREM 3.3. *Consider a max-linear Bayesian network with coefficient matrix C and Kleene star C^*. Then $g \in \mathfrak{G}$ if and only if the following four conditions hold*:

    (a) *g is a subgraph of $\mathcal{D}^*$.*
    (b) *g is a galaxy, that is, a forest of stars.*
    (c) *If $j \to i$ in g and $c_{ij}^* = c_{ik}^* c_{kj}^*$, then $k \nrightarrow i$ and $j \to k$ in g.*
    (d) *$\lambda(M(g)) < 1$.*

Before we proceed to the proof of this result, some explanation of the elements of the theorem might be appropriate. Theorem 3.3 describes all possible impact scenarios. With probability one, any outcome of the max-linear Bayesian network has a system of (extreme) root variables Z_R and the value at all other nodes will be a.s. constant and appropriate multiples of these, their impact spreading across the network as determined by the galaxy $g \in \mathfrak{G}$.

The conditions (a) and (b) in Theorem 3.3 are necessary, but not sufficient. To understand condition (d), consider the definition of the impact exchange matrix $M(g)$. Intuitively, the
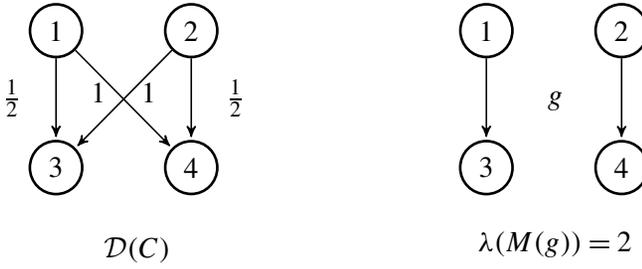
FIG. 5. *Bipartite DAG: The subgraph g to the right is not an impact graph for the weighted DAG $\mathcal{D}(C)$ to the left as it violates the principal eigenvalue condition (d) in Theorem* 3.3.

entry $m_{rr'}$ measures the worst possible relative cost for a node i to be reassigned from root r to root r' in g. The graph induced by positive entries of $M(g)$ may have directed cycles. A directed cycle in this graph starting at a root r creates an inequality involving Z_r. Condition (d) of Theorem 3.3 ensures that this inequality can be satisfied. Example 3.4 below shows that a violation of the condition on the principal eigenvalue $\lambda(M(g))$ of the impact exchange matrix $M(g)$ yields an inconsistency, even if the other conditions are satisfied. The argument in Example 3.4 below illustrates the key step in the proof that establishes the necessity of condition (d).

EXAMPLE 3.4 (Bipartite).   Consider the weighted graph $\mathcal{D}$ with weights given in Figure 5. The subgraph g to the right in Figure 5 satisfies conditions (a)–(c) of Theorem 3.3. However, it fails to satisfy condition (d) because

$$M(g) = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix},$$

since

$$m_{12} = \max_{i \in \mathrm{ch}_g(1)=\{3\}} \frac{c_{i2}^*}{c_{i1}^*} = \frac{c_{32}^*}{c_{31}^*} = \frac{1}{1/2} = 2,$$

$$m_{21} = \max_{i \in \mathrm{ch}_g(2)=\{4\}} \frac{c_{i1}^*}{c_{i2}^*} = \frac{c_{41}^*}{c_{42}^*} = \frac{1}{1/2} = 2.$$

Then $\lambda(M(g)) = \sqrt{2 \cdot 2} = 2 > 1$, so g is not an impact graph. Indeed if it were, $1 \to 3$ would imply $X_3 = \frac{1}{2}Z_1 > Z_2$ but $2 \to 4$ would imply $X_4 = \frac{1}{2}Z_2 > Z_1$, and thus $Z_1 > 2Z_2 > 4Z_1$, which is inconsistent since $Z_1 > 0$.

PROOF OF THEOREM 3.3.   First, we show that all conditions are necessary.

To prove (a), let $g \in \mathfrak{G}$. If $c_{ij}^* = 0$, then $X_i > 0 = c_{ij}^* Z_j$, which means that $j \to i \notin g$. So g is a subgraph of $\mathcal{D}^*$ which proves (a).

To establish (b), we first note that g must be a forest from Remark 1. We shall then argue that any tree in the forest has height at most one. Suppose $j \to i \in g$. Then

$$X_i = c_{ij}^* Z_j > Z_i \quad \text{on } \mathcal{E}(g).$$

Now, for any $k \in V$, either $c_{ki}^* = 0$ so $i \to k \notin g$ by (a), or we have by the idempotency of C^* and (2.5) that

$$X_k \geq c_{ki}^* c_{ij}^* Z_j > c_{ki}^* Z_i \quad \text{on } \mathcal{E}(g)$$

and, therefore, there is no edge $i \to k$ in g which proves (b).

Next, we establish (c). Consider a triple of nodes i, j, k with $j \to i$ and $c_{ij}^* = c_{ik}^* c_{kj}^*$. Since g is a forest and $j \to i$, we must have $k \to i \notin g$. Also, since $j \to i$ we have as before

$$X_i = c_{ij}^* Z_j \quad \text{on } \mathcal{E}(g).$$

Using (2.5) again, we know that $X_i \geq c_{ik}^* X_k$. Then the relation $c_{ij}^* = c_{ik}^* c_{kj}^*$ yields

$$X_k \geq c_{kj}^* Z_j = \frac{c_{ij}^*}{c_{ik}^*} Z_j = \frac{X_i}{c_{ik}^*} \geq X_k \quad \text{on } \mathcal{E}(g),$$

and hence we must have equality so $X_k = c_{kj}^* Z_j$ on $\mathcal{E}(g)$. This proves (c).

For condition (d), we first note that if $\lambda(M) = 0$ then it is certainly less than 1. So assume $\lambda(M) > 0$. Then there exists a critical cycle $r_1 \leftarrow r_2 \cdots \leftarrow r_k \leftarrow r_1$ with $r_1, \ldots, r_k \in R$ such that

$$(3.4) \qquad 0 < \left(\lambda(M)\right)^k = m_{r_1 r_2} m_{r_2 r_3} \ldots m_{r_k r_1}.$$

In particular, this implies each edge in the cycle is not 0, so for each edge, say, $r_2 \to r_1$, there exists a node $i \in V$ that achieves this maximum so that $r_1 \to i$ in g and

$$\frac{c_{ir_2}^*}{c_{ir_1}^*} = m_{r_1 r_2}.$$

Now, since $r_1 \to i$ in g and i has at most one parent, this implies $c_{ir_1}^* Z_{r_1} > c_{ir_2}^* Z_{r_2}$, whereby $Z_{r_1} > m_{r_1 r_2} Z_{r_2}$ by rearranging. Tracing this cycle, we obtain the equation

$$Z_{r_1} > (m_{r_1 r_2} m_{r_2 r_3} \ldots m_{r_k r_1}) Z_{r_1}.$$

Dividing by $Z_{r_1} > 0$, we obtain from (3.4) that $\lambda(M) < 1$. Thus, all four conditions are necessary.

To see that the conditions are sufficient, let g be a graph that satisfies all four conditions. Let $\epsilon > 0$ be an arbitrarily small constant, and

$$(3.5) \qquad \alpha = \max_{i,j,k,\ell: c_{ji}^*, c_{\ell k}^* > 0} \frac{c_{\ell k}^*}{c_{ji}^*}.$$

Let v be a tropical eigenvector of M for $\lambda(M) < 1$. This means

$$(M \odot v)_r = \lambda(M) v_r < v_r,$$

for $v_r > 0$, so that the event

$$\mathcal{E} = \{ v_r > Z_r > (M \odot v)_r \text{ for } r \in R \text{ s.t. } v_r > 0, \; Z_r > \alpha\epsilon \text{ for } r \in R \text{ s.t. } v_r = 0,$$
$$\text{and } Z_j < \epsilon \text{ for } j \notin R \}$$

satisfies $\mathbb{P}(\mathcal{E}) > 0$. We now argue that $\mathcal{E}$ is a subevent of $\mathcal{E}(g)$. Since the collection of events $\{\mathcal{E}(g) : g \in \mathfrak{G}\}$ partitions the innovation space $\mathcal{E}$, the event $\mathcal{E}$ must be partitioned into finitely many events, each with positive probability, which we denote $\mathcal{E} \cap \mathcal{E}(g_1'), \ldots, \mathcal{E} \cap \mathcal{E}(g_s')$. By definition of g, each i belongs to a unique star with root r. Under the event $\mathcal{E}$, for all $r' \in R$, $r' \neq r$,

$$Z_r > \max_{r'} m_{rr'} v_{r'} > \max_{r'} m_{rr'} Z_{r'},$$

therefore, $Z_r > m_{rr'} Z_{r'}$ that is, $c_{ir}^* Z_r > c_{ir'}^* Z_{r'}$ whence $r' \not\to i$ for all $g' \in \{g_1', \ldots, g_s'\}$. Similarly, for any $r' \notin R$ with $c_{ir'}^* > 0$,

$$c_{ir}^* Z_i > c_{ir}^* \alpha \geq c_{ir}^* \frac{c_{ir'}^*}{c_{ir}^*} \epsilon > c_{ir'}^* Z_{r'}$$

and hence $r' \not\to i$ in any of $g' \in \{g_1', \ldots, g_s'\}$. Thus $r \to i$ in any g' and we must have $\mathcal{E} \subseteq \mathcal{E}(g)$, so $\mathbb{P}(\mathcal{E}(g)) \geq \mathbb{P}(\mathcal{E}) > 0$, as needed. $\square$
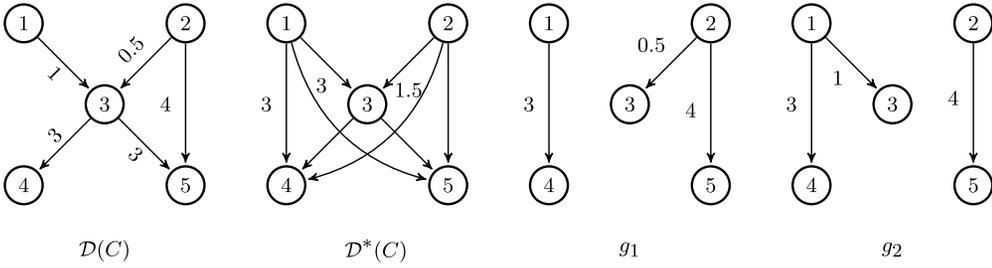
FIG. 6. *The half-butterfly graph $\mathcal{D}(C)$ and its weighted reachability DAG $\mathcal{D}^*(C)$ where only edge weights for additional edges are indicated. The galaxy g_1 is not an impact graph for this DAG as it violated the as it violates the triangle condition (c) in Theorem 3.3, while the galaxy g_2 is.*

EXAMPLE 3.5 (Half-butterfly). Let $\mathcal{D}$ be the weighted DAG given in the leftmost part of Figure 6. Its weighted reachability DAG $\mathcal{D}^*(C)$ is shown to its right. For example, we have $c_{41}^* = c_{43}^* c_{31}^* = 3$. Now consider the two different subgalaxies g_1 and g_2 shown to the right in Figure 6. We shall see that g_2 is an impact graph for the given coefficient matrix C while g_1 is not. Indeed, $g_1 \notin \mathfrak{G}$ as it violates the triangle condition (c) in Theorem 3.3: $1 \to 4 \in g_1$ and $c_{41}^* = c_{43}^* c_{31}^*$ but $1 \to 3 \notin g_1$. On the other hand, $1 \to 3 \in g_2$ as required. Furthermore, g_2 has impact exchange matrix given by

$$M(g_2) = \begin{bmatrix} 0 & \frac{1}{2} \\ \frac{3}{4} & 0 \end{bmatrix},$$

because

$$m_{12} = \max_{i \in \text{ch}_{g_2}(1) = \{3,4\}} \frac{c_{i2}^*}{c_{i1}^*} = \max\left\{ \frac{1/2}{1}, \frac{3/2}{3} \right\} = \frac{1}{2},$$

$$m_{21} = \max_{i \in \text{ch}_{g_2}(2) = \{5\}} \frac{c_{i1}^*}{c_{i2}^*} = \frac{c_{51}^*}{c_{52}^*} = \frac{3}{4}.$$

We have then $\lambda(M(g_2)) = \sqrt{\frac{1}{2} \cdot \frac{3}{4}} = \sqrt{\frac{3}{8}} < 1$ and so condition (d) also holds. We conclude by Theorem 3.3 that $g_2 \in \mathfrak{G}$. A possible realization in terms of Z is given by $Z = (z_1, z_2, z_3, z_4, z_5) = (2, 3, 0.1, 0.4, 0.2)$ leading to $X = (x_1, x_2, x_3, x_4, x_5) = (2, 3, 2, 6, 12)$.

The following simple lemma shall be used in the subsequent analysis.

LEMMA 3.6. *Consider the max-linear Bayesian network* (2.2) *with fixed coefficient matrix C. Let $g \in \mathfrak{G}$ be an impact graph with root set $R = R(g)$. Then it holds for all $z \in \mathcal{E}(g)$ that*

$$(3.6) \qquad\qquad\qquad\qquad M \odot z_R \leq z_R,$$

where $M = M(g, C)$ is the impact exchange matrix of g and $z_R = (z_r)_{r \in R}$ is the truncation of z to the root set.

PROOF. Let $\mathcal{E}' = \{z : z \text{ does not satisfy (3.6)}\}$. Note that $\mathcal{E}'$ decomposes as the union of $R(R-1)$ subevents $\mathcal{E}'_{rr'}$, where
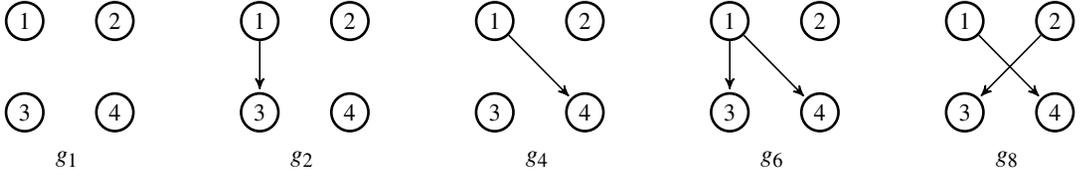
$$\mathcal{E}'_{rr'} = \{m_{rr'} z_{r'} > z_r\}.$$

FIG. 7.   *Impact graphs $\mathfrak{G}$ for the bipartite DAG in Figure 5. There are a total of eight such graphs, the remaining three (g_3, g_5, and g_7) obtained by the reflection $(1, 3) \leftrightarrow (2, 4)$ of g_2, g_4 and g_6.*

We shall next show that $\mathcal{E}(g) \cap \mathcal{E}'_{rr'} = \varnothing$ for each pair $r, r' \in R$ with $r \neq r'$. Suppose for contradiction that there exists some $z \in \mathcal{E}(g) \cap \mathcal{E}'_{rr'}$. By definition, $m_{rr'} = \max_{i \in \mathrm{ch}_g(r)} \frac{c^*_{ir'}}{c^*_{ir}}$. Let $i \in \mathrm{ch}_g(r)$ be a node that achieves this maximum. Then $z \in \mathcal{E}'_{rr'}$ implies

$$c^*_{ir'} z_{r'} > c^*_{ir} z_r.$$

But now the max-linear representation of X implies that on $\mathcal{E}'_{rr'}$,

$$x_i \neq c^*_{ir} z_r$$

which contradicts that $r \to i$ in g. Hence we conclude that $\mathcal{E}(g) \cap \mathcal{E}'_{rr'} = \varnothing$, and thus further that

$$\mathcal{E}(g) \cap \mathcal{E}' = \mathcal{E}(g) \cap \bigcup_{rr'} \mathcal{E}'_{rr'} = \varnothing,$$

as needed. $\square$

3.2. *Impact graphs compatible with a context.*   As mentioned in Remark 2, the impact graphs represent a partition of the innovation space $\mathcal{E} = \mathbb{R}^V_>$ into regions of linearity. We can also represent these as linear maps $L_g : z \in \mathbb{R}^{R(g)}_> \mapsto x \in \mathbb{R}^V_>$ via

$$L_g(z)_r = z_r, \qquad L_g(z)_i = c^*_{ir} z_r \quad \text{iff} \quad r \to i \quad \text{in } g.$$

We shall illustrate this in a small example.

EXAMPLE 3.7 (Bipartite).   Consider again the DAG and coefficient matrix of Example 3.4 as depicted in Figure 5. Figure 7 displays all impact graphs for this DAG, save for their symmetric counterparts.

Of the 16 edge-induced subgraphs of the DAG $\mathcal{D}$, only nine are forests and one of them, displayed to the right in Figure 5, violates the principal eigenvalue condition; so there are eight valid galaxies left, five of which are displayed in Figure 7, and the remaining three obtained by appropriate relabeling. The max-linear map is

$$C^* \odot z = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{1}{2} & 1 & 1 & 0 \\ 1 & \frac{1}{2} & 0 & 1 \end{bmatrix} \odot \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \frac{1}{2} z_1 \vee z_2 \vee z_3 \\ z_1 \vee \frac{1}{2} z_2 \vee z_4 \end{bmatrix}$$

and the corresponding matrices C_g^* for the pieces of linearity are

$$
\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
\quad
\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
\quad
\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}
$$

$$
\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}
\quad
\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}
$$

mapping z respectively into

$$
\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{bmatrix}
\quad
\begin{bmatrix} z_1 \\ z_2 \\ \frac{1}{2}z_1 \\ z_4 \end{bmatrix}
\quad
\begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_1 \end{bmatrix}
\quad
\begin{bmatrix} z_1 \\ z_2 \\ \frac{1}{2}z_1 \\ z_1 \end{bmatrix}
\quad
\begin{bmatrix} z_1 \\ z_2 \\ z_2 \\ z_1 \end{bmatrix} .
$$

These maps can also be considered as maps L_g from the root set to the node set and would then have matrices

$$
\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
\quad
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\quad
\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}
\quad
\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \frac{1}{2} & 0 \\ 1 & 0 \end{bmatrix}
\quad
\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} ,
$$

where the roots $(1, 2, 4)$ in g_2 have been renumbered as $(1, 2, 3)$. Indeed these matrices are simply obtained by removing zero-columns in the first set of matrices. Note that the rank r_g of the maps are all equal to $r_g = |R(g)|$, the number of stars in the galaxy, that is, 4, 3, 3, 2, 2 in these cases.

DEFINITION 3.8. Let $K \subseteq V$ and $\Pi_K(x) = x_K$ be the projection onto coordinates in K. For $x_K \in \mathcal{L}_K^C$ as defined in (2.6), we define:

(a) A graph $g \in \mathfrak{G}$ is *compatible with the context* $\{X_K = x_K\}$ if the following are true:

   (i) $\mathcal{E}(g) \cap \{X_K = x_K\} \neq \varnothing$,
   (ii) the rank of $\Pi_K \circ L_g$ is minimal among those $g \in \mathfrak{G}$ which satisfy (i).

(b) The set of compatible graphs g is called the *impact graphs for the context* $\{X_K = x_K\}$, denoted $\mathfrak{G}(X_K = x_K)$.

(c) We further say that the context $\{X_K = x_K\}$ is *possible* if $\mathfrak{G}(X_K = x_K) \neq \varnothing$ and *possible under* g if $g \in \mathfrak{G}(X_K = x_K)$. Else the context $\{X_K = x_K\}$ is said to be *impossible* or *impossible under* g respectively. For brevity, we shall also use the expression that x_K is *possible*.

Note that although all events of the form $\{X_K = x_K\}$ have probability zero, we are now distinguishing between those that are exceptions from events of the form $\mathcal{E}(g)$ (impossible contexts) and those that are not (possible contexts). In other words, x_K might still satisfy $x_K \in \mathcal{L}_K^C$ without being possible. In the following, we shall only pay attention to possible contexts. Furthermore, this definition also applies to the special case $K = V$ so we now can speak about $\{X = x\}$ being possible or impossible under $g \in \mathfrak{G}$.

The rank condition (a) (ii) ensures that if any subevent $\mathcal{E}(g^*)$ includes x_K and the map $\Pi_K \circ L_{g^*}$ has higher rank than $\Pi_K \circ L_g$, then the entire collection of contexts $\{X_K = x_K\}$ in the image of $\Pi_K \circ L_g$ is a null-set in $\mathcal{E}(g^*)$. Therefore, the set of points in $\mathcal{L}_K^C$ that are not possible has measure zero and can be ignored when discussing conditional distributions.

EXAMPLE 3.9.   Consider the Cassiopeia graph in Example 1.2 with all coefficients equal to one and the event $\{X_4 = X_5 = 2\}$. The impact graphs g_3 and g_4 are the only impact graphs among those in Figure 4 that satisfy condition (i) in Definition 3.8, as the other impact graphs imply strict inequalities between x_4 and x_5. In addition, the empty galaxy, and all galaxies with a single edge satisfy condition (i). However, the rank of $\Pi_K \circ L_{g_3}$ is one, whereas the rank of all other maps $\Pi_K \circ L_g$ is two. Hence only g_3 is compatible with $\{X_4 = X_5 = 2\}$.

DEFINITION 3.10.   Suppose $\{X_K = x_K\}$ is possible. We say that X_j is *constant on* $\{X_K = x_K\}$ if there exists $x_j^* \in \mathbb{R}_>$ such that

$$\{X_{K \cup j} = x_{K \cup j}\} \quad \text{is possible if and only if} \quad x_j = x_j^*.$$

Similarly, X_j is *constant on* $\{X_K = x_K\} \cap \mathcal{E}(g)$ if there exists $x_j^* \in \mathbb{R}_>$ such that

$$\{X_{K \cup j} = x_{K \cup j}\} \quad \text{is possible under } g \text{ if and only if} \quad x_j = x_j^*.$$

Define the set of *constant nodes* on $\{X_K = x_K\}$ as

$$K^* := K^*(X_K = x_K) := \{j \in V : X_j \text{ is constant on } \{X_K = x_K\}\}$$

and nodes that are *constant under g* as

$$K^*(g) := K^*(X_K = x_K, g) := \{j \in V : X_j \text{ is constant on } \{X_K = x_K\} \cap \mathcal{E}(g)\}.$$

Note that $K \subseteq K^* \subseteq K^*(g)$ for specific $g \in \mathfrak{G}$. Often these inclusions can be strict (see Example 3.14). The following lemma characterizes these sets. Recall from Theorem 3.3 that each impact graph $g \in \mathfrak{G}$ is a galaxy.

LEMMA 3.11.   *Suppose $g \in \mathfrak{G}(X_K = x_K)$ and $S = V(\sigma)$ is the node set of a star σ in the galaxy g. Then, either*

(a) $S \cap K \neq \varnothing$, *in which case $S \subseteq K^*(g)$ and we call S a* constant star; *or*
(b) $S \cap K = \varnothing$, *in which case $S \cap K^*(g) = \varnothing$.*

In particular,

$$K^*(g) = \bigcup_{\sigma \in g : S \cap K \neq \varnothing} V(\sigma).$$

PROOF.   We first consider (a): Note that if $j \to i \in g$, then on $\mathcal{E}(g)$, $X_i = c_{ij}^* Z_j$, and thus $X_i = c_{ij}^* X_j$. Therefore, if either X_i or X_j is constant on $\{X_K = x_K\} \cap \mathcal{E}(g)$, then both must be constant. So if one node in S is in K, all nodes in S must be in $K^*(g)$. This proves (a).

Next, we show (b): Let R be the set of root nodes in g, $R^1 \subset R$ be the set of root nodes for all stars S in g such that $S \cap K = \varnothing$, and $R^c = R \setminus R^1$ be the set of roots of the constant stars. By the first statement, X_r is constant for all $r \in R^c$ on $\mathcal{E}(g) \cap \{X_K \in x_K\}$. Indeed, on $\mathcal{E}(g) \cap \{X_K = x_K\}$, $M \odot z_R \leq z_R$ by Lemma 3.6. Furthermore, by the minimal rank condition, we must have strict inequality, that is, $M \odot z_R < z_R$. This equation splits up into the following lower and upper-bounds for z_{R^1} in terms of z_{R^c}:

(3.7) $$M_{R^1 R^c} \odot z_{R^c} < z_{R^1},$$

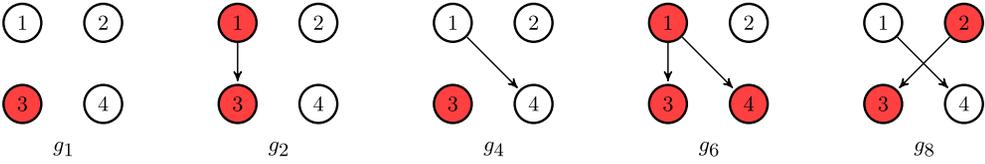(3.8) $$M_{R^c R^1} \odot z_{R^1} < z_{R^c}.$$

FIG. 8.   *Impact graphs in $\mathfrak{G}(X_3 = x_3)$ for the bipartite DAG in Figures 5 and 7. Red nodes are constant under g for the context $\{X_3 = x_3\}$, that is, elements of $K^*(g)$. These are nodes that are in the same star as the node 3.*

Since $g \in \mathfrak{G}(X_K = x_K)$, the upper and lower bounds cannot coincide. In particular, there exist two values of x_r such that $\{X_{K \cup r} = x_{K \cup r}\}$ is possible under g. This implies that $R^1 \cap K^*(g) = \varnothing$, and since R^1 are the roots, none of their children can be in $K^*(g)$. This completes the proof.   □

EXAMPLE 3.12 (Bipartite).   Consider again the DAG and coefficient matrix of Example 3.4 and let $K = \{3\}$. Figure 8 again displays impact graphs for this DAG, now with the constant nodes in red color. Nodes are constant in the context $\{X_3 = x_3\}$ if and only if they belong to the same star as the node 3.

Lemma 3.13 below identifies a crucial property of a compatible impact graph.

LEMMA 3.13.   *Let $g \in \mathfrak{G}$ be an impact graph. If g is compatible with $\{X_K = x_K\}$, we have for all $i, j, h \in V$ that*

$$(3.9) \qquad \frac{x_h}{c^*_{hj}} < \frac{x_i}{c^*_{ij}}, \quad h, i \in K^*(g) \quad \implies \quad j \neq R_g(i),$$

$$(3.10) \qquad \exists j \in V : \frac{x_h}{c^*_{hj}} = \frac{x_i}{c^*_{ij}}, \quad h, i \in K^*(g) \quad \implies \quad R_g(i) = R_g(h).$$

PROOF.   Consider (3.9). Clearly, $j \neq i$; suppose then for contradiction that $j \to i$. Since $i \in K^*(g)$, Lemma 3.11 implies that $j \in K^*(g)$. Then $x_i = c^*_{ij} x_j$, so (3.9) implies that $x_h / c^*_{hj} < x_i / c^*_{ij} = x_j$, so $x_h < c^*_{hj} x_j$. But this contradicts that $x_h \geq c^*_{hj} x_j$. Now consider (3.10). Write $r = R_g(i)$ and $r' = R_g(h)$. Then $x_i = c^*_{ir} x_r$ and $x_h = c^*_{hr'} x_{r'}$. Suppose for contradiction that $r \neq r'$. Substituting into the hypothesis of (3.10) we get

$$\frac{c^*_{hr'}}{c^*_{hj}} x_{r'} = \frac{c^*_{ir}}{c^*_{ij}} x_r,$$

which is a linear relation on the roots x_r and $x_{r'}$ of two different stars in g. But this contradicts that g has minimal rank according to Definition 3.8, and thus $\{X_K = x_K\}$ is not possible under g. Hence (3.10) must hold.   □

EXAMPLE 3.14 (Half-butterfly).   To illustrate Lemma 3.13, we again consider the DAG and coefficient matrix C of Example 3.5, depicted in Figure 6 and the context $\{X_K = x_K\}$ where $K = \{4, 5\}$ and $x_4 = x_5 = 1$. We claim that the set of constant nodes is $K^* = \{3, 4, 5\}$ because the events $\{X_4 = X_5 = 1\}$ and $\{X_4 = X_5 = 1, X_3 = 1/3\}$ are almost surely identical, and that there are exactly two impact graphs compatible with this context, depicted in Figure 9.

To see this, let g be a compatible impact graph in $\mathfrak{G}(X_4 = X_5 = 1)$. Since $x_4 / c^*_{43} = 1/3 = x_5 / c^*_{53}$ we may apply (3.10) with $i = 4$, $h = 5$, $j = 3$ to conclude that 4 and 5 belong to the same star in g, with common root $R_g(4) = R_g(5)$.
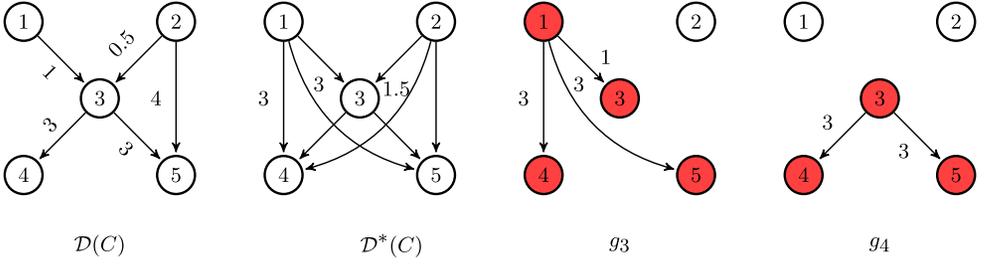
FIG. 9. *Impact graphs g_3 and g_4 for the half-butterfly compatible with the context $\{X_4 = X_5 = 1\}$ are displayed to the right. To the left, the original DAG $\mathcal{D}(C)$ and its weighted reachability DAG $\mathcal{D}^*(C)$ are shown.*

By Theorem 3.3(a), g is a subgraph of $\mathcal{D}^*(C)$, and hence the roots must be parents in this graph, that is, we have $R_g(4) \in \{1, 2, 3, 4\}$, and $R_g(5) \in \{1, 2, 3, 5\}$. So $R_g(4) = R_g(5)$ implies $R_g(4) = R_g(5) \in \{1, 2, 3\}$. On the other hand, it also holds that $x_5/c^*_{52} = 1/4 < 1/1.5 = x_4/c^*_{42}$. Then (3.9) implies that $2 \notin \mathrm{pa}_g(4)$, so $R_g(4) = R_g(5) \in \{1, 3\}$. By Theorem 3.3(b), g is a star so each node can not have more than one parent. So either it must hold that $R_g(4) = R_g(5) = 1$, or that $R_g(4) = R_g(5) = 3$. In the second case, 1 and 2 are left as isolated roots. In the first case, $1 \to 4 \in g$ implies that we also have $1 \to 3 \in g$ by Theorem 3.3(c), that is, 3 must belong to the same star as 4 with 1 as a root. This gives the two impact graphs to the right in Figure 9. In both cases, there is at most one nonisolated root, so $M(g)$ has no cycles, and thus $\lambda(M(g)) = 0 < 1$ required for g to be an impact graph. Thus both graphs are in $\mathfrak{G}(X_4 = X_5 = 1)$. By Definition 3.10, the constant nodes are those that are in the same star as 4, 5 in the compatible impact graphs, so $K^* = \{3, 4, 5\}$ as we then have $X_3 = 1/3$ on $\{X_4 = X_5 = 1\}$, so $\{X_4 = X_5 = 1\} = \{X_4 = X_5 = 1, X_3 = 1/3\}$.

We can double-check that $\mathfrak{G}(X_4 = X_5 = 1) = \mathfrak{G}(X_4 = X_5 = 1, X_3 = 1/3)$ by computing the latter set of graphs directly. Let $g \in \mathfrak{G}(X_4 = X_5 = 1, X_3 = 1/3)$. Now we first apply (3.10) with $i = j = 3$ and $h = 5$ to conclude that $R_g(3) = R_g(5)$. Thus again we have by Theorem 3.3(a) that $R_g(3) \in \{1, 2, 3\}$ and $R_g(5) \in \{1, 2, 3, 5\}$. Therefore, $R_g(5) \neq 5$. Apply again (3.9) with $j = 2$, $i = 3$ and $h = 5$ to conclude that $2 \neq R_g(3)$. So $R_g(3) \in \{1, 3\}$. The two cases $R_g(3) = 1$ and $R_g(3) = 3$ yield the two impact graphs to the right in Figure 9 as expected.

REMARK 3. Although in Definition 3.1, we have defined the impact graph $G = G(Z)$ (for almost all Z), G can also be expressed in terms of X, as we indeed have for any $g \in \mathfrak{G}$ which is compatible with $\{X = x\}$ that

$$(3.11) \qquad j \to i \in g \quad \Longrightarrow \quad x_i/x_j = c^*_{ij} \quad \text{on } \mathcal{E}(g)$$

since on $\mathcal{E}(g)$ we must have $j \in R(g)$, and thus $X_j = Z_j$. Hence with probability one there is a unique $g \in \mathfrak{G}$ that is compatible with $\{X = x\}$. Another way of expressing this is to say that the map $z \to g$ is almost surely $\sigma(X)$-measurable, where $\sigma(X)$ is the σ-algebra generated by the max-linear map $z \to x$ given by $x = C^* \odot z$.

3.3. *The source DAG.* Impact graphs describe how extreme events at their roots spread deterministically to other nodes. In this section, we shall capitalize on this, but from the perspective of identifying which are the possible sources of extreme values responsible for a given possible context of the form $\{X_K = x_K\}$ (see Definition 3.8(c)). This will eventually make it possible for us to answer interesting queries concerning conditional independence.

We first let $\mathcal{I}(X_K = x_K)$ denote the union of impact graphs which are compatible with the context:

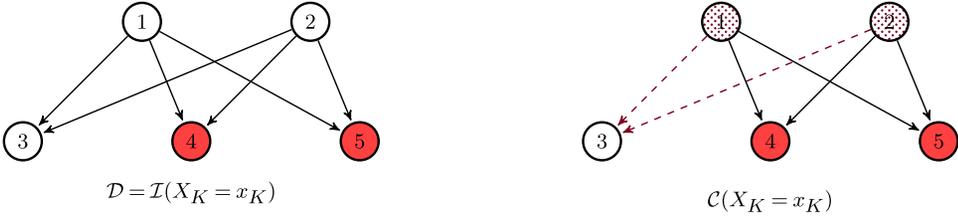$$(3.12) \qquad \mathcal{I}(X_K = x_K) = \bigcup_{g \in \mathfrak{G}(X_K = x_K)} g$$

FIG. 10.   *Tent graph: To the left, this displays $\mathcal{D}(C) = \mathcal{D}^*(C) = \mathcal{I}(X_K = x_K)$ when all coefficients are equal to 1. To the right, the source DAG $\mathcal{C}(X_K = x_K)$ for $K = \{4, 5\}$ and $x_4 = x_5 = 2$ is obtained by removing the dashed edges, which are redundant as each of the dotted nodes is constant in any impact graph containing the edge.*

and we shall refer to this as the *total impact graph* and note that it is a subgraph of the reachability DAG $\mathcal{D}^*$. In other words, this graph yields all possible ways that impact could have spread across the network in a way that conforms with the observation $\{X_K = x_K\}$.

DEFINITION 3.15.   Let $K \subset V$ and $K^*(g)$ be the set of constant nodes under g as in Definition 3.10. An edge $j \to i$ in $\mathcal{I}(X_K = x_K)$ is *redundant in the context* $\{X_K = x_K\}$ if either:

- $j \in K^*$, or
- $i \notin K^*$ and X_j is constant under all $g \in \mathfrak{G}(X_K = x_K)$ that contain the edge.

The set of redundant edges is denoted $E^- = E^-(X_K = x_K)$.

DEFINITION 3.16.   The *source DAG* $\mathcal{C}(X_K = x_K)$ of a possible context $\{X_K = x_K\}$ is the graph obtained from $\mathcal{I}(X_K = x_K)$ by removing redundant edges; see Definition 3.15.

EXAMPLE 3.17 (Tent).   Consider the DAG $\mathcal{D}$ to the left in Figure 10 with all edge weights $c_{ij} = 1$. Let $K = \{4, 5\}$ and $x_4 = x_5 = 2$. Note that $\mathcal{I}(X_K = x_K) = \mathcal{D}(C)$. However, $\mathcal{C}(X_K = x_K)$ is the strict subgraph of $\mathcal{I}(X_K = x_K)$ obtained by removing the dashed edges from the graph to the right in Figure 10.

The edge $1 \to 3$ is in E^- since 1 is constant under all impact graphs containing this edge, and similarly with the edge $2 \to 3$. To see this, note that $1 \to 3 \in g \in \mathfrak{G}(X_4 = X_5 = 2)$ if and only if $1 \to 4$, $1 \to 5 \in g$, which then implies $1 \in K^*(g)$. Therefore, $1 \to 3$ is a redundant edge and not included in $\mathcal{C}(X_4 = X_5 = 2)$. A similar argument applies to the edge $2 \to 3$. From the node partition of Proposition 3.18 below, we see that the active nodes are $A = \{1, 2, 3\}$ and the constant nodes $K^* = K = L = \{4, 5\}$.

We first prove some results on the structure of the source DAG before linking it up to probabilistic statements. In particular, we establish that the source DAG admits a nice partition structure; see Figure 11 for an illustration.

PROPOSITION 3.18.   *Fix a possible context $\{X_K = x_K\}$, let $\mathcal{I} = \mathcal{I}(X_K = x_K)$ and $\mathcal{C} = \mathcal{C}(X_K = x_K)$ be the corresponding total impact graph and source DAG, respectively. Then for either of these graphs, its node set V can be partitioned into disjoint sets $A \cup U \cup H \cup L$, where*

(a) A: $a \in A \iff a \notin K^*$ *is the set of* active nodes (*nonconstant*);
(b) U: $u \in U \iff u \in K^*$ *and* $\exists k \in K^*$, $k \neq u$ *such that* $x_u = c_{uk}^* x_k$
(c) H: $h \in H \iff h \in K^*$ *and* $\exists g \in \mathfrak{G}(X_K = x_K)$ *such that* $h \in R(g)$, *and*
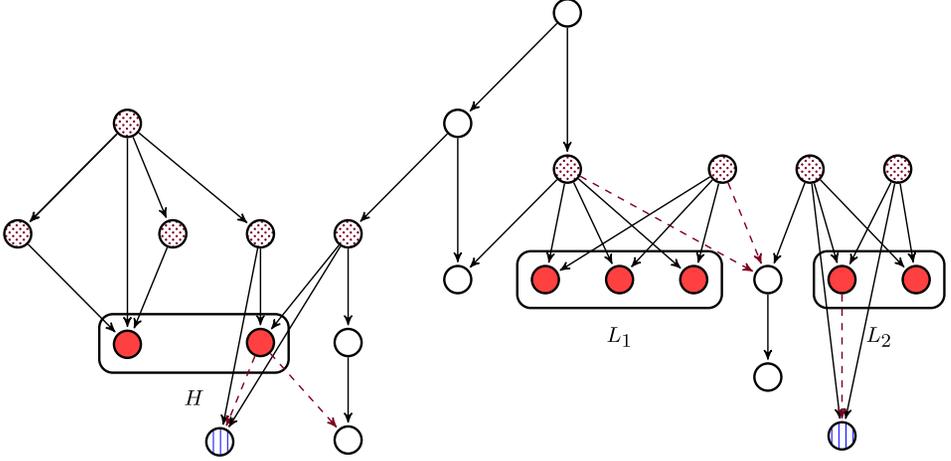(d) L: $\ell \in L \iff \ell \in K^* \setminus (H \cup U)$

FIG. 11. *Illustration of the partition in Proposition* 3.18. *Nodes in H and L are red, nodes in U are filled with blue lines. Active nodes A are white or dotted burgundy, where the dotted nodes are parents of constant nodes in $K = H \cup L \cup U$. All edges, including the dashed edges, are in $\mathcal{I}(X_K = x_K)$, while only the solid edges are in $\mathcal{C}(X_K = x_K)$.*

In addition, we have the following:

(e) *For all $k \in H \cup L$, $\mathrm{pa}_{\mathcal{C}}(k) = \mathrm{pa}_{\mathcal{I}}(k)$.*

(f) *For $k, k' \in H \cup L$, $k \neq k'$, either $\mathrm{pa}_{\mathcal{C}}(k) \cap \mathrm{pa}_{\mathcal{C}}(k') = \varnothing$ or $\mathrm{pa}_{\mathcal{C}}(k) = \mathrm{pa}_{\mathcal{C}}(k') \neq \varnothing$.*

(g) *If $h \in H$, then $\mathrm{pa}_{\mathcal{C}}(h) \cap \mathrm{pa}_{\mathcal{C}}(k) = \varnothing$ for all $k \in H \cup L$, $k \neq h$.*

(h) *The set $H \cup L$ can be partitioned into equivalence classes where $k \equiv k' \iff \mathrm{pa}_{\mathcal{C}}(k) = \mathrm{pa}_{\mathcal{C}}(k')$. Under this equivalence relation, elements of H are singletons, and L is partitioned into disjoint subsets $L = L_1 \cup \cdots \cup L_m$.*

(i) *Any $\ell \in L$ has at least two parents.*

(j) *For all $a \in A$ and $\ell \in L$, there exists some $i \in \mathrm{pa}_{\mathcal{C}}(\ell)$ such that $i \notin \mathrm{pa}_{\mathcal{C}}(a)$.*

PROOF. By definition, $V = A \cup U \cup H \cup L$, and all pairs are mutually disjoint except for possibly U and H. Indeed, suppose $u \in U$. Let $k \in K^*$ be such that $x_u = c^*_{uk} x_k$, $k \neq u$. Let $g \in \mathfrak{G}(X_K = x_K)$. By (3.10), $R_g(u) = R_g(k)$. But $\mathcal{D}$ is a DAG, so $c^*_{uk} > 0$ implies $c^*_{ku} = 0$, so in particular, $u \neq R_g(k)$, hence, u cannot be a root in g. Thus $u \notin H$, so $U \cap H = \varnothing$. This proves (a) to (d). Consider (e). By definition, $\mathrm{pa}_{\mathcal{C}}(k) \subseteq \mathrm{pa}_{\mathcal{I}}(k)$. Suppose for contradiction that the containment is strict, that is, there exists some $i \in V$ such that $i \to k \in \mathcal{I}$ but $i \to k \notin \mathcal{C}$. Then $i \to k \in E^-$. Since $k \in K^*$, we must have $i \in K^*$. Then $k \in U$, so $k \notin H \cup L$, a contradiction. This proves (e). Consider (f). Let k, k' be two such nodes. Let $i \in \mathrm{pa}_{\mathcal{C}}(k) \cap \mathrm{pa}_{\mathcal{C}}(k')$. If this set is empty, then we are done. Otherwise, consider x_k / c^*_{ki} and $x_{k'} / c^*_{k'i}$. If one of these two quantities are bigger, then either $i \to k$ or $i \to k'$ is not in g for all $g \in \mathfrak{G}(X_K = x_k)$ by (3.9), so $i \notin \mathrm{pa}_{\mathcal{C}}(k) \cap \mathrm{pa}_{\mathcal{C}}(k')$. So these two quantities must be equal. By (3.10), for all $g \in \mathfrak{G}(X_K = x_K)$, $\mathrm{pa}_g(k) = \mathrm{pa}_g(k')$. Thus $\mathrm{pa}_{\mathcal{I}}(k) = \mathrm{pa}_{\mathcal{I}}(k')$. Since $k, k' \in H \cup L$, (e) then implies (f). Now consider (g). Suppose for contradiction that there exists some $k \in H \cup L$ such that $\mathrm{pa}_{\mathcal{C}}(h) = \mathrm{pa}_{\mathcal{C}}(k)$. As argued previously, this implies h and k cannot be the root of any $g \in \mathfrak{G}(X_K = x_K)$. So in particular, $h \notin H$, and we obtain the desired contradiction. Statement (h) follows immediately from (f) and (g). Now we prove (i). Suppose for contradiction that $\ell \in L$ has only one parent $i \in V$. Since $i \to \ell \notin E^-$, $i \to \ell \in \mathcal{I}(X_K = x_K)$. In other words, for all $g \in \mathfrak{G}(X_K = x_K)$, $R_g(\ell) = i$. By Lemma 3.11(a), this implies $i \in K^*$, so $\ell \in U$, a contradiction, as desired. Now we prove (j). Suppose for contradiction that there exists some $a \in A$ and $\ell \in L$ such that $\mathrm{pa}_{\mathcal{C}}(\ell) \subseteq \mathrm{pa}_{\mathcal{C}}(a)$. Let $r \in \mathrm{pa}_{\mathcal{C}}(\ell)$ be a node with smallest coefficient

$c^*_{aj} x_\ell / c^*_{\ell j}$ among $j \in \mathrm{pa}_\mathcal{C}(\ell)$, that is,

$$c^*_{ar} \frac{x_\ell}{c^*_{\ell r}} \leq c^*_{aj} \frac{x_\ell}{c^*_{\ell j}} \quad \text{for all } j \in \mathrm{pa}_\mathcal{C}(\ell).$$

Since $r \to a \in \mathcal{C}(X_K = x_K)$, there exists some $g \in \mathfrak{G}(X_K = x_K)$ such that $r \to a \in g$ and $r \notin K^*(g)$. Thus this implies $r \to \ell \notin g$, so there exists some $j \in \mathrm{pa}_\mathcal{C}(\ell)$ with $j \to \ell \in g$. Since g is a galaxy, $j \to a \notin g$. Then by definition, on the event $\mathcal{E}(g)$, $j \to \ell \in g$ and $r \to \ell \notin g$ together imply

$$x_\ell = c^*_{\ell j} Z_j > c^*_{\ell r} Z_r.$$

Rearranging gives

$$c^*_{aj} Z_j = c^*_{aj} \frac{x_\ell}{c^*_{\ell j}} \geq c^*_{ar} \frac{x_\ell}{c^*_{\ell r}} > c^*_{ar} Z_r,$$

but this contradicts the fact that $j \to a \notin g$ and $r \to a \notin g$, since these two imply

$$X_a = c^*_{aj'} Z_{j'} < c^*_{ar} Z_r.$$

So we have a contradiction, as needed. $\square$

The nodes in U have no direct effect on the conditional distribution, as their effect is mitigated through their (constant) parents. Proposition 3.18 is illustrated in Figure 11.

COROLLARY 3.19. *Let $\mathcal{C}(X_K = x_K)$ be the source DAG of a possible context $\{X_K = x_K\}$ and consider the node partition $V = A \cup K^* = A \cup H \cup L \cup U$ as given by Proposition 3.18. If $j \to i \in g$ for some $g \in \mathfrak{G}(X_K = x_K)$, $j, i \notin K^*$ and $j \in K^*(g)$, then $j \to h \in g$ for some $h \in H \cup L$.*

PROOF. Let $S = V(\sigma) \subseteq V$ be the set of nodes in the star σ with root j in the galaxy g. Since $\mathfrak{G}(X_K = x_K) = \mathfrak{G}(X_{K^*} = x_{K^*})$, apply Lemma 3.11(a) to $\mathfrak{G}(X_{K^*} = x_{K^*})$ giving $S \cap K^* \neq \varnothing$. Let $u \in S \cap K^*$. If $u \notin U$, then take $h = u$ and we are done. Else, by Proposition 3.18, there exist some $h \neq u$, $h \in K^*$ such that $x_u = c^*_{uh} x_h$. By Theorem 3.3(c), $j \to h \in g$, so $h \in S \cap K^*$. If $h \notin U$, then we are done, else we repeat the above argument once more to find another node in $S \cap K^*$. Since $\mathcal{D}$ is a DAG, every time we repeat this argument we obtain a new node. Since the graph is finite, this procedure eventually terminates and returns some node $h \in S \cap K^*$ and $h \notin U$. By Proposition 3.18, $h \in K \cup L$. $\square$

4. Representing the conditional distribution. Before we derive conditional independence results, we need to have a good control of conditional distributions in a max-linear Bayesian network. We first derive a basic representation in Section 4.1 and subsequently a more compact representation without redundancy in Section 4.2.

4.1. *Basic representation.* Let $K \subset V$ and $\bar{K} = V \setminus K$. The conditional distribution of $X_{\bar{K}} \mid X_K = x_K$ can be represented by a system of max-linear equations in the $Z_{\bar{K}}$ variables; more precisely, we have:

PROPOSITION 4.1. *The following is a representation for $X \mid X_K = x_K$ with respect to the innovations Z,*

$$(4.1) \qquad X_{\overline{K}} = C^*_{\overline{K} K} \odot x_K \vee C^*_{\overline{K} \overline{K}} \odot Z_{\overline{K}},$$

where the distribution of Z is that of independent components, conditioned to satisfy

$$(4.2) \qquad x_K = C^*_{KK} \odot Z_K \vee C^*_{K \overline{K}} \odot Z_{\overline{K}}.$$

PROOF.   By (2.5), we have $X = C^* \odot X$ so

$$X \geq \begin{bmatrix} 0 & 0 \\ C^*_{\overline{K}K} & 0 \end{bmatrix} \odot \begin{bmatrix} X_K \\ X_{\overline{K}} \end{bmatrix}.$$

Now, $X = C^* \odot Z$, therefore,

$$X = \begin{bmatrix} 0 & 0 \\ C^*_{\overline{K}K} & 0 \end{bmatrix} \odot \begin{bmatrix} X_K \\ X_{\overline{K}} \end{bmatrix} \vee C^* \odot Z \quad \text{and} \quad X = \begin{bmatrix} C^*_{KK} & C^*_{K\overline{K}} \\ C^*_{\overline{K}K} & C^*_{\overline{K}\overline{K}} \end{bmatrix} \odot \begin{bmatrix} Z_K \\ Z_{\overline{K}} \end{bmatrix}.$$

Writing out these equations, we obtain

(4.3) $$X_{\overline{K}} = C^*_{\overline{K}K} \odot x_K \vee C^*_{\overline{K}K} \odot Z_K \vee C^*_{\overline{K}\overline{K}} \odot Z_{\overline{K}},$$

(4.4) $$x_K = C^*_{KK} \odot Z_K \vee C^*_{K\overline{K}} \odot Z_{\overline{K}}.$$

The second equation is (4.2). For the first equation, note that $c^*_{ii} = 1$ for all i, so $x_K \geq Z_K$. Thus $C^*_{\overline{K}K} \odot x_K \vee C^*_{\overline{K}K} \odot Z_K = C^*_{\overline{K}K} \odot x_K$, so (4.3) is equivalent to (4.1). Thus the context $\{X_K = x_K\}$ is equal to the conjunction of the events (4.2) and (4.4). The result follows.   □

EXAMPLE 4.2.   We shall illustrate Proposition 4.1 for the Cassiopeia graph of Figure 2 in Example 1.2. Assume that we have $c_{ji} = c^*_{ji} = 1$ for all edges in this DAG and let $x_K = (x_4, x_5)$. Then (4.1) becomes

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \odot \begin{bmatrix} x_4 \\ x_5 \end{bmatrix} \vee I_3 \odot \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix},$$

whereas (4.2) becomes

$$\begin{bmatrix} x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \odot \begin{bmatrix} Z_4 \\ Z_5 \end{bmatrix} \vee \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \odot \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix}.$$

This means that $x_4 \geq Z_4$, $x_5 \geq Z_5$ and

$$\begin{bmatrix} x_4 \\ x_5 \end{bmatrix} \geq \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \odot \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} = \begin{bmatrix} Z_1 \vee Z_2 \\ Z_2 \vee Z_3 \end{bmatrix}.$$

Depending on whether $x_4 < x_5$, $x_4 > x_5$, or $x_4 = x_5$, these inequalities are a.s. equivalent to respectively

$$\begin{bmatrix} x_4 \\ x_5 \end{bmatrix} \geq \begin{bmatrix} Z_1 \vee Z_2 \\ Z_3 \end{bmatrix}, \qquad \begin{bmatrix} x_4 \\ x_5 \end{bmatrix} \geq \begin{bmatrix} Z_1 \\ Z_2 \vee Z_3 \end{bmatrix}, \qquad \begin{bmatrix} x_4 \\ x_5 \end{bmatrix} \geq \begin{bmatrix} Z_1 \\ Z_3 \end{bmatrix} \quad \text{and} \quad Z_2 = x_4 = x_5.$$

Thus the conditioning under this restriction renders Z_i bounded in all cases, and in the third case Z_2 becomes constant. Note also that in these reduced inequalities, Z_1 and Z_3 never occur together in any inequality, rendering $X_1 \perp\!\!\!\perp X_3 \mid X_{\{4,5\}}$.

4.2. *Compact representation.*   The main result of this section, Theorem 4.3, states that the source DAG gives a representation of the active nodes $X_A \mid X_K = x_K$ with respect to the innovations Z. Compared to the representation of the conditional distribution in Proposition 4.1, this is a representation with fewer terms. Most importantly, we shall show below that the system of equations involving Z can be separated into blocks where no terms are redundant.

THEOREM 4.3.   *Let $\mathcal{C}(X_K = x_K)$ be the source DAG of a possible context $\{X_K = x_K\}$, with node partition $V = A \cup H \cup L \cup U = A \cup H \cup (L_1 \cup \cdots \cup L_m) \cup U$ as in Proposition* 3.18. *For each $t = 1, \ldots, m$, select a node $\ell_t \in L_t$. Then the following system of equations yields a representation for $X_A \mid X_K = x_K$ with respect to Z:*

$$(4.5) \qquad X_a = \alpha_a \vee Z_a \vee \bigvee_{j \in \mathrm{pa}_{\mathcal{C}}(a)} c^*_{aj} Z_j, \quad a \in A,$$

where the constants α_a are given by

$$(4.6) \qquad \alpha_a = \bigvee_{k \in K^*} c^*_{ak} x_k \vee \left(\bigvee_{\substack{j \in A \\ j \to a \in E^-}} \bigvee_{k \in \mathrm{ch}_{\mathcal{D}}(j) \cap (H \cup L)} c^*_{aj} \frac{x_k}{c^*_{kj}} \right), \quad a \in A,$$

and the distribution of Z is that of independent components, conditioned to satisfy the bounds

$$(4.7) \qquad Z_i \leq \bigwedge_{k \in K^*: c^*_{ki} > 0} \frac{x_k}{c^*_{ki}}, \quad i \in V,$$

as well as the equations

$$(4.8) \qquad x_h = Z_h \vee \bigvee_{j \in \mathrm{pa}_{\mathcal{C}}(h)} c^*_{hj} Z_j, \quad h \in H,$$

$$(4.9) \qquad x_{\ell_t} = \bigvee_{j \in \mathrm{pa}_{\mathcal{C}}(\ell_t)} c^*_{\ell_t j} Z_j, \quad t = 1, \ldots, m.$$

Furthermore, each innovation term on the right-hand side of (4.5), (4.8) *and* (4.9) *has positive probability of being the term that achieves equality.*

PROOF.   We shall begin with the representation of $X \mid X_K = x_K$ given by Proposition 4.1 and then simplify terms until we obtain the representation above. The contexts $\{X_K = x_K\}$ and $\{X_{K^*} = x_{K^*}\}$ are clearly equivalent, so we may without loss of generality assume that $K = K^*$ and $A = \bar{K}$. This gives for (4.1) and (4.2) the representations

$$(4.10) \qquad X_A = C^*_{AK^*} \odot x_{K^*} \vee C^*_{AA} \odot Z_A,$$

$$(4.11) \qquad x_{K^*} = C^*_{K^*K^*} \odot Z_{K^*} \vee C^*_{K^*A} \odot Z_A.$$

First, we simplify (4.11). With $K^* = H \cup L \cup U$, we expand this system of equations as follows:

$$(4.12) \qquad x_H = C^*_{HK^*} \odot Z_{K^*} \vee C^*_{HA} \odot Z_A$$

$$(4.13) \qquad x_L = C^*_{LK^*} \odot Z_{K^*} \vee C^*_{LA} \odot Z_A$$

$$(4.14) \qquad x_U = C^*_{UK^*} \odot Z_{K^*} \vee C^*_{UA} \odot Z_A.$$

For each $i \in V$ and each $k \in K^*$, all inequalities on Z_i implied by (4.11) are

$$x_k \geq c^*_{ki} Z_i$$

whenever $c^*_{ki} > 0$, and in particular this is equivalent to (4.7).

Next, we keep track of the equalities. For $u \in U$, by Proposition 3.18, $x_u = c^*_{uk} x_k$ for some $k \in K^*$, $k \neq u$. We have

$$\begin{aligned} x_u &= c^*_{uk} x_k \\ &= c^*_{uk} C^*_{kK^*} \odot Z_{K^*} \vee c^*_{uk} C^*_{kA} \odot Z_A \text{ by (4.11)} \\ &\leq C^*_{uK^*} \odot Z_{K^*} \vee C^*_{uA} \odot Z_A \\ &= x_u \quad \text{by (4.11).} \end{aligned}$$

Thus we conclude

$$C^*_{uK^*} \odot Z_{K^*} \vee C^*_{uA} \odot Z_A = c^*_{uk}(C^*_{kK^*} \odot Z_{K^*} \vee C^*_{kA} \odot Z_A), \quad k \in K^*, k \neq u,$$

whence the constraint imposed upon Z by x_u is identical to the constraint imposed upon Z by x_k. Therefore, all equations in (4.14) are redundant. So (4.11) is equivalent to the conjunction of (4.12) and (4.13).

Next, we simplify the terms that appear on the right-hand side of (4.12) and (4.13). Fix $k \in H \cup L$. By definition of the source DAG, we keep a term $c^*_{ki} Z_i$ for $i \neq k$ if and only if $i \to k \in g$ for some $g \in \mathfrak{G}(X_K = x_K)$, and we keep the term $c^*_{kk} Z_k = Z_k$ if and only if $k \in R(g)$ for some $g \in \mathfrak{G}(X_K = x_K)$. Since each g is a galaxy, each remaining term has a positive probability of achieving the maximum. Since each event under $\mathfrak{G}(X_K = x_K)$ with positive probability must correspond to some g, there is always one that achieves the maximum among the remaining terms. Since $k \in K^*$, it holds that $i \to k \in g$ for some $g \in \mathfrak{G}(X_K = x_K)$ if and only if we also have $i \to k \in \mathcal{C}(X_K = x_K)$. Therefore, by Proposition 3.18, (4.12) simplifies to (4.8) and (4.13) simplifies to

$$(4.15) \qquad\qquad x_\ell = \bigvee_{j \in \mathrm{pa}_\mathcal{C}(\ell)} c^*_{\ell j} Z_j.$$

Write $L = L_1 \cup \cdots \cup L_m$ as given by Proposition 3.18. If $\ell, \ell' \in L_t$ for some $t = 1, \ldots, m$, then they share the same set of parents. By Lemma 3.13, this implies

$$\frac{x_\ell}{c^*_{\ell j}} = \frac{x_{\ell'}}{c^*_{\ell' j}}$$

for all $j \in \mathrm{pa}_\mathcal{C}(\ell) = \mathrm{pa}_\mathcal{C}(\ell')$. So (4.15) for x_ℓ and $x_{\ell'}$ are constant multiples of each other. So for each $t = 1, \ldots, m$ (4.15) for all $\ell \in L_t$ is equivalent to the single equation (4.9).

Now we simplify (4.10). Like in the previous step, we can for $a, j \in A$ drop terms $c^*_{aj} Z_j$ where $j \to a \notin g$ for every $g \in \mathfrak{G}(X_K = x_K)$. This gives

$$X_a = C^*_{aK^*} \odot x_{K^*} \vee \left(\bigvee_{j \in A, j \to a \in E^-} c^*_{aj} Z_j \right) \vee Z_a \vee \bigvee_{j \in \mathrm{pa}_\mathcal{C}(a)} c^*_{aj} Z_j, \quad a \in A.$$

Now we argue that each term in $\bigvee_{j \in A, j \to a \in E^-} c^*_{aj} Z_j$ can be replaced by an appropriate constant. Let $j \in A$ be a node with $j \to a \in E^-$. Let $\mathcal{E}$ be the subevent of $\mathfrak{G}(X_K = x_K)$ where j is the root of a, that is,

$$\mathcal{E} = \bigcup \{ \mathcal{E}(g) : g \in \mathfrak{G}(X_K = x_K), j \to a \in g \}.$$

By definition, for each $g \in \mathfrak{G}(X_K = x_K)$ such that $j \to a \in g$, we have that $a, j \in K^*(g)$. Since $a, j \notin K^*(g)$, it follows that there exists some $k = k(g) \in K^*(g)$ such that $j \to k \in g$. Thus, on the event $\mathcal{E}(g) \cap \{X_K = x_K\}$, we have

$$X_a = c^*_{aj} Z_j = c^*_{aj} \frac{x_{k(g)}}{c^*_{k(g)j}}, \quad a \in A.$$

By Corollary 3.19, we may assume that $k(g) \in H \cup L$. Therefore, on $\mathcal{E}$,

$$(4.16) \qquad\qquad X_a = \bigvee_{k \in \mathrm{ch}_\mathcal{D}(j) \cap (H \cup L)} c^*_{aj} \frac{x_k}{c^*_{kj}}, \quad a \in A.$$

By definition of $\mathcal{E}$, on the complement $\{X_K = x_K\} \setminus \mathcal{E}$, $X_a > c^*_{aj} Z_j$. Therefore, the term $c^*_{aj} Z_j$ can be dropped from the representation of X_a and be replaced by the right-hand side of (4.16). This gives

$$X_a = C^*_{aK^*} \odot x_{K^*} \vee \left(\bigvee_{j \in A, j \to a \in E^-} \bigvee_{k \in \mathrm{ch}_\mathcal{D}(j) \cap (H \cup L)} c^*_{aj} \frac{x_k}{c^*_{kj}} \right) \vee Z_a \vee \bigvee_{j \in \mathrm{pa}_\mathcal{C}(a)} c^*_{aj} Z_j, \quad a \in A.$$
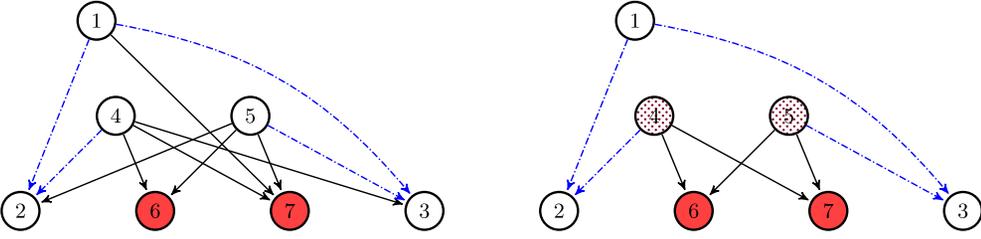
FIG. 12. *The umbrella graph: To the left, $\mathcal{D}(C) = \mathcal{D}^*(C) = \mathcal{D}_K^*(C)$; to the right: the source DAG $\mathcal{C}(X_K = x_K)$ for $K = \{6, 7\}$ and $x_6 = x_7 = 3$. Black edges have weights 1, dash-dotted blue edges have weights 2. The nodes are colored as in Fig. 11.*

This is (4.5), with α_a equal to the constant terms in the equation above, which is the formula in (4.6). Finally, by definition of $\mathcal{C}(X_K = x_K)$, for each $j \in \mathrm{pa}_C(i)$ there exists some $g \in \mathfrak{G}(X_K = x_K)$ such that $j \to i$ and $j \notin K^*(g)$. Since g is a galaxy and X_i is *not* constant on the event $\mathcal{E}(g) \cap \{X_K = x_K\}$, on this event, $c_{ij}^* Z_j$ is the unique term that achieves the maximum in (4.5). □

REMARK 4. We note that in (4.7), only the bounds for the variables $Z_{A \cup H}$ are directly relevant for the conditional distribution of X_A given $X_K = x_K$, as the variables $Z_{L \cup U}$ do not enter into any of the equations (4.5), (4.6), (4.8) or (4.9). However, we have included these in Theorem 4.3 to provide a full description of the conditional distribution of Z given $X_K = x_K$, which may be of interest for other purposes.

The following example of an umbrella graph illustrates some aspects of this representation.

EXAMPLE 4.4 (Umbrella). Consider the graph to the left in Figure 12 where solid black edges have weights 1 and dash-dotted blue edges have weights 2. The partition in Theorem 4.3 yields nodes $A = V \setminus K^* = V \setminus K = \{1, 2, 3, 4, 5\}$, and $K = L = \{6, 7\}$. The nonzero constants (4.6) in the representation of the active variables are $\alpha_2 = \alpha_3 = 3$, calculated as

$$\alpha_2 = 0 \vee \bigvee_{5:5 \to 2} \bigvee_{k \in \{6,7\}} c_{25}^* \frac{x_k}{c_{k5}^*} = \max(3, 3) = 3,$$

since $c_{2k}^* = 0$ for $k = 6, 7$ and there is only one edge $5 \to 2 \in E^-$ pointing to 2. A similar calculation yields $\alpha_3 = 3$. The full representation (4.5) then becomes

$$X_1 = Z_1, \qquad X_4 = Z_4, \qquad X_5 = Z_5,$$
$$X_2 = 3 \vee Z_2 \vee 2Z_1 \vee 2Z_4,$$
$$X_3 = 3 \vee Z_3 \vee 2Z_1 \vee 2Z_5$$

with inequalities from (4.7) yielding the bounds $Z_1, Z_4, Z_5 \leq 3$ (and $Z_6, Z_7 \leq 3$). Further, we have the equality (4.9) yielding

$$(4.17) \qquad\qquad x_6 = x_7 = 3 = Z_4 \vee Z_5$$

whereas (4.8) is void.

For Z_1, we claim that it cannot simultaneously achieve the bound in both of the equations for X_2 and X_3, illustrating Proposition 3.18(g). In fact, if $X_2 = 2Z_1$ then $3 < 2Z_1 \leq 6$, so that $2Z_4 < 2Z_1 \leq 6$. Then (4.17) yields that $Z_4 < 3$ and thus $Z_5 = 3$, but then $X_3 = 6 > 2Z_1$. Moreover, since Z_4 and Z_5 both enter into the equation (4.17) we conclude that $X_2 \not\perp\!\!\!\perp X_3 \mid X_{\{6,7\}} = (3, 3)$.

We next present some important consequences of Theorem 4.3, enabling us to identify conditional independencies.

COROLLARY 4.5.   *For each pair $i, j \in V$, either Z_i, Z_j appear together in exactly one equation among those in* (4.8) *and* (4.9), *or they do not appear together in any of those equations. In the first case, they are conditionally dependent, that is, $Z_i \not\perp\!\!\!\perp Z_j \mid X_K = x_K$. In the second case, they are conditionally independent, that is, $Z_i \perp\!\!\!\perp Z_j \mid X_K = x_K$.*

PROOF.   By Theorem 4.3, the distribution of $Z \mid X_K = x_K$ is the distribution of Z given the events defined by (4.7), (4.8) and (4.9). The bounds (4.7) only involve one variable at a time, and thus play no role for independence issues. Groups of Z's that appear in different equations in (4.8) and (4.9) are independent. It remains to show that, if Z_i, Z_j appear in the same equation, then they are dependent. Indeed, suppose that Z_i, Z_j appear in (4.8) for some $h \in H$, with coefficients $a_i, a_j > 0$. The event $\mathcal{E}_h$ defined by this equation can be rewritten as

$$Z_h \leq x_h, \quad a_i Z_i \leq x_h, \quad a_j Z_j \leq x_h, \quad \text{and} \quad a_{j'} Z_{j'} \leq x_h \quad \text{for all other } j' \in \mathrm{pa}_\mathcal{C}(h),$$

and exactly one of these terms achieves equality. Further, each term has a positive probability of achieving equality. That is,

$$\mathbb{P}(a_i Z_i = x_h \mid \mathcal{E}_h) > 0 \quad \text{and} \quad \mathbb{P}(a_j Z_j = x_h \mid \mathcal{E}_h) > 0,$$

but

$$\mathbb{P}(a_j Z_j = x_h \mid a_i Z_i = x_h, \mathcal{E}_h) = 0.$$

Therefore, $Z_i \not\perp\!\!\!\perp Z_j \mid \mathcal{E}_h$. A similar argument applies for the equation (4.9).   □

COROLLARY 4.6.   *For $i, j \in A$, $j \to i \in \mathcal{C}(X_K = x_K)$, suppose that $j \to k \in \mathcal{C}(X_K = x_K)$ for some $k \in H \cup L$. Then there exists $g \in \mathfrak{G}(X_K = x_K)$ such that j is a parent of both i and j in g. In other words,*

(4.18) $$\mathbb{P}\big(X_i = c^*_{ij} Z_j, Z_j = x_k / c^*_{kj} \mid X_K = x_K\big) > 0$$

PROOF.   Since $j \to i \in \mathcal{C}(X_K = x_K)$, there exists some $g \in \mathfrak{G}(X_K = x_K)$ such that $j \to i \in g$ and $j \notin K^*(g)$, so in particular, $j \to k \notin g$. Let $\mathcal{E}(g)$ as usual denote the Z-values corresponding to the impact graph g. Since $j \to k \notin g$, there exists some other $j' \neq j$ such that $z_{j'} = x_k / c^*_{kj'}$ and $z_j < x_k / c^*_{kj}$ for all $z \in \mathcal{E}(g)$. Transform the region $\mathcal{E}(g)$ to another region $\phi(\mathcal{E}(g))$ via the following linear map ϕ, where

$$\phi(z)_j = x_k / c^*_{kj}, \qquad \phi(z)_{j'} = z_j, \quad \text{and} \quad \phi(z)_{j''} = z_{j''} \quad \text{for all } j'' \neq j, j'.$$

Since this is an invertible map and $\mathbb{P}(\mathcal{E}(g) \mid X_K = x_K) > 0$, we have $\mathbb{P}(\phi(\mathcal{E}(g)) \mid X_K = x_K) > 0$. Thus there exists some $g' \in \mathfrak{G}(X_K = x_K)$ such that $\mathbb{P}(\mathcal{E}(g') \cap \{X_K = x_K\} \cap \phi(\mathcal{E}(g))) > 0$. By definition of ϕ, for such g' we must have $j \to i, j \to k \in g'$. This concludes the proof.   □

COROLLARY 4.7.   *For each $a \in A$, the atomic component of the distribution of X_a is supported precisely on the following points*:

   (a) α_a *defined by* (4.6) *if $\alpha_a > 0$;*
   (b) $c^*_{aj} x_k / c^*_{kj}$ *for each $j \in \mathrm{pa}_\mathcal{C}(a) \cap \mathrm{pa}_\mathcal{C}(k)$ for some $k \in H \cup L$.*

PROOF.   Suppose X_a has an atomic component at some $c \in \mathbb{R}_>$. This happens if and only if there exists some $g \in \mathfrak{G}(X_K = x_K)$ such that $X_a = c$ on $\mathcal{E}(g) \cap \{X_K = x_K\}$. In particular, we must have $a \in K^*(g)$. Since $a \notin K^*$, $a \in K^*(g)$ if and only if $j \to a \in g$ for some $j \in V$, and either $j \in K^*$ or $\mathrm{ch}_g(j) \cap K^* \neq \varnothing$. We consider these two cases separately.

(a) Suppose $j \in K^*$. Then $c = c_{aj}^* x_j \leq \alpha_a$ by (4.6). If $c < \alpha_a$, then $\mathbb{P}(X_a = c \mid X_K = x_K) = 0$ by (4.5), a contradiction. So $c = \alpha_a$.

(b) Suppose $j \notin K^*$. By Corollary 3.19, there exists some $k \in \mathrm{ch}_g(j) \cap H \cap L$. By Corollary 4.6, $\mathbb{P}(X_a = c_{aj}^* Z_j, Z_j = x_k/c_{jk}^* \mid X_K = x_K) > 0$, so the distribution of X_a has an atom at $c_{aj}^* x_k/c_{kj}^*$.

So all the atomic components of the distribution of X_a must be of the form given.   □

5. Markov properties of max-linear Bayesian networks.   In this section, we introduce the relevant separation criteria and state and prove the three conditional independence theorems. We first consider the most difficult context-specific case and then use the results for this case to derive the more generic results which are valid in all contexts.

5.1. *Graphs and separation.*   In addition to the source DAG as defined in Definition 3.16, we shall need the following graphs to identify Markov properties of a max-linear Bayesian network.

DEFINITION 5.1.   Fix a DAG $\mathcal{D}$ on V and $K \subset V$. Say that a directed path π from j to i *factors through* K if there exists a node $k \in \pi$, $k \neq i, j$ such that $k \in K$. The *conditional reachability DAG* $\mathcal{D}_K^*$ is the graph on V consisting of the following edges: $j \to i \in \mathcal{D}_K^*$ if and only if there exists a directed path from j to i that does *not* factor through K.

DEFINITION 5.2.   Fix a DAG $\mathcal{D}$ on V, $K \subset V$ and a coefficient matrix C supported by $\mathcal{D}$. The *critical DAG* $\mathcal{D}_K^*(C)$ is the graph on V consisting of the following edges: $j \to i \in \mathcal{D}_K^*(C)$ if and only if $c_{ij}^* > 0$ and no critical directed path from j to i factors through K.

Note that in contrast to Definition 5.1 the existence of a *single* critical path through K removes the corresponding edge in the critical DAG $\mathcal{D}_K^*(C)$; this conforms with Example 1.1 in the Introduction where it is sufficient to block a single critical path to obtain conditional independence.

When $K = \varnothing$, we write $\mathcal{D}^* = \mathcal{D}_\varnothing^*$ for the reachability DAG of $\mathcal{D}$, and $\mathcal{D}^*(C) = \mathcal{D}_\varnothing^*(C)$ if the support of C is $\mathcal{D}$. The source DAG $\mathcal{C}(X_K = x_K)$ for $K = \varnothing$ does not have a direct meaning, but by convention we let this be $\mathcal{C}(X_\varnothing = x_\varnothing) = \mathcal{D}^*$.

LEMMA 5.3.   *Let C be a coefficient matrix with support $\mathcal{D}$. Furthermore, let $K \subset V$ and let $\{X_K = x_K\}$ be a possible context. Then*

$$(5.1) \qquad\qquad \mathcal{D}_K^* \supseteq \mathcal{D}_K^*(C) \supseteq \mathcal{C}(X_K = x_K).$$

PROOF.   First, we prove that $\mathcal{D}_K^*(C) \subseteq \mathcal{D}_K^*$. Let $j \to i \in \mathcal{D}_K^*(C)$. Since $c_{ji}^* > 0$ and no critical directed paths from j to i factor through K, there exists at least one critical directed path from j to i that does not factor through K. Therefore, $j \to i \in \mathcal{D}_K^*$. Now we prove that $\mathcal{C}(X_K = x_K) \subseteq \mathcal{D}_K^*(C)$. Suppose $j \to i \in \mathcal{C}(X_K = x_K)$. Clearly, we must have $c_{ji}^* > 0$. Suppose for contradiction that $j \to k \to i$ is critical for some $k \in K$. Then on $\{X_K = x_K\}$,

$$c_{ij}^* Z_j = c_{ik}^* c_{kj}^* Z_j \leq c_{ik}^* x_k.$$

So $j \to k \notin \mathcal{C}(X_K = x_K)$, a contradiction. Therefore, all critical paths from j to i do not factor through K, so $j \to i \in \mathcal{D}_K^*(C)$ by definition.   □
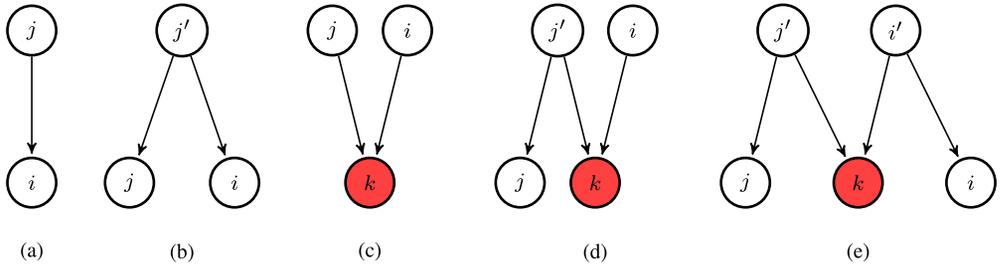
FIG. 13.   *Types of ∗-connecting paths between i and j. Nodes that are colored red are in K.*

DEFINITION 5.4.   An undirected path π between j and i in a DAG is *∗-connecting* relative to K if and only if it is one of the paths in Figure 13.

We shall consider ∗-connecting paths in the conditional reachability DAG $\mathcal{D}_K^*$, in the critical DAG $\mathcal{D}_K^*(C)$, and in the source DAG $\mathcal{C}(X_K = x_K)$. Edges in these DAGs represent directed paths in the original DAG $\mathcal{D}$ and each of the paths in Figure 13 may represent longer paths in the original DAG $\mathcal{D}$. Note also that any ∗-connecting path in a derived DAG corresponds to a d-connecting path in $\mathcal{D}$, but not vice versa, as illustrated in Example 5.8 below.

We now define three independence models by applying ∗-separation to $\mathcal{D}_K^*$, $\mathcal{D}_K^*(C)$ and the source DAG $\mathcal{C}(X_K = x_K)$, respectively.

DEFINITION 5.5.   For three disjoint subsets I, J and K of the node set V we say that I and J are $\mathcal{D}^*$-*separated* by K in $\mathcal{D}$ if there are no ∗-connecting paths from I to J in $\mathcal{D}_K^*$ and we then write $I \perp_{\mathcal{D}^*} J \mid K$ or $I \perp_* J \mid K$ in $\mathcal{D}_K^*$.

DEFINITION 5.6.   For three disjoint subsets I, J and K of the node set V we say that I and J are *critically separated* by K in $\mathcal{D}$ if there is no ∗-connecting path π from I to J in $\mathcal{D}_K^*(C)$. We then write $I \perp_{C^*} J \mid K$ or $I \perp_* J \mid K$ in $\mathcal{D}_K^*(C)$.

DEFINITION 5.7.   For three disjoint subsets I, J and K of the node set V we say that I and J are *source separated* by $X_K = x_K$ in $\mathcal{D}$ if there are no ∗-connecting paths from I to J in $\mathcal{C}(X_K = x_K)$. We then write $I \perp_{(C^*, x_K)} J \mid K$ or $I \perp_* J \mid K$ in $\mathcal{C}(X_k = x_k)$.

EXAMPLE 5.8 (Cassiopeia).   Example 1.2 illustrates that $\mathcal{D}^*$-separation is strictly weaker than d-separation. Here, $\mathcal{D} = \mathcal{D}_K^* = \mathcal{D}_K^*(C)$ for any C with support $\mathcal{D}$. The path between 1 and 3 is d-connecting, but it is not ∗-connecting.

We emphasize that our separation criteria follow the form of the moralization procedure in [25], which is not stated in a directly path-based form. Rather, we check for separation by constructing derived graphs—$\mathcal{D}_K^*$, $\mathcal{D}_K^*(C)$, and $\mathcal{C}(X_k = x_k)$—and then use a single common separation criteria for all of these. This formulation shall simplify some of the proofs. As a consequence of Lemma 5.3, we get the following.

COROLLARY 5.9.   *For I, J, K disjoint subsets of V and any possible context $\{X_K = x_K\}$, it holds that*

$$I \perp_{\mathcal{D}} J \mid K \quad \Longrightarrow \quad I \perp_{\mathcal{D}^*} J \mid K \quad \Longrightarrow \quad I \perp_{C^*} J \mid K \quad \Longrightarrow \quad I \perp_{(C^*, x_K)} J \mid K,$$
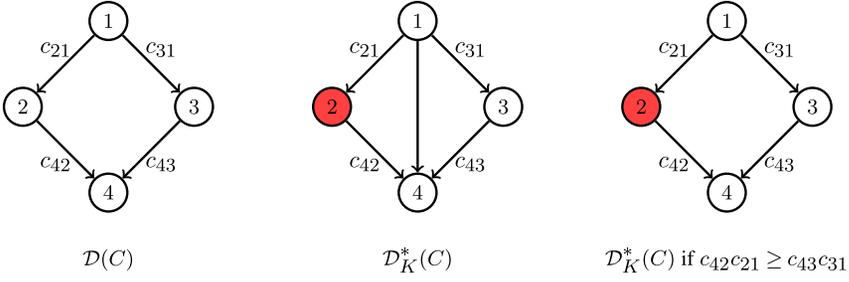
where $\perp_{\mathcal{D}}$ denotes d-separation.

FIG. 14.   *Diamond graph with $K = \{2\}$. The conditional reachability DAG (middle figure) is equal to the reachability DAG $\mathcal{D}^*$, whereas the edge $1 \to 4$ is missing in the critical DAG (right-hand figure) since the path $1 \to 2 \to 4$ is critical and factors through K. Note that the path $1 \to 3 \to 4$ in $\mathcal{D}_K^*(C)$ is not $*$-connecting as it is not one of the configurations in Figure 13.*

We note that these implications are strict, as illustrated in the next example and other examples further below.

EXAMPLE 5.10 (Diamond).   Consider the DAG in Figure 14 in a situation where the path $1 \to 2 \to 4$ is critical: $c_{42}c_{21} \geq c_{43}c_{31}$. It then holds that $1 \perp\!\!\!\perp 4 \mid 2$ even though there is a d-connecting path $1 \to 3 \to 4$. By Definition 5.4, this path is not $*$-connecting in $\mathcal{D}_K^*(C)$ so $1 \perp_{C^*} 4 \mid 2$. Note also that $\perp_{\mathcal{D}^*}$ is strictly weaker than $\perp_{C^*}$, as $1 \perp_{C^*} 4 \mid 2$ if $c_{21}c_{42} \geq c_{31}c_{43}$, but it holds that $\neg(1 \perp_{\mathcal{D}^*} 4 \mid 2)$ since $1 \to 3 \to 4$ is $*$-connecting in $\mathcal{D}_K^*$.

5.2. *The context-specific case.*   We first consider the case of a specific context and a fixed coefficient matrix C. To prove our main Theorem 5.12, we need the following lemma.

LEMMA 5.11.   *Let $\mathcal{C}(X_K = x_K)$ be the source DAG of a possible context $\{X_K = x_K\}$. Then $i, j \in A$ (the active nodes of Proposition 3.18) are source separated if and only if*

$$(5.2) \qquad\qquad (\{i\} \cup \mathrm{pa}_C(i)) \cap (\{j\} \cup \mathrm{pa}_C(j)) = \varnothing,$$

and that there is no triple of nodes i', j', k such that

$$(5.3) \quad i' \in (\{i\} \cup \mathrm{pa}_C(i)), \qquad j' \in (\{j\} \cup \mathrm{pa}_C(j)), \qquad k \in H \cup L \quad and \quad i', j' \in \mathrm{pa}_C(k).$$

PROOF.   The nodes i and j are $*$-connected if and only if there is a path π in $\mathcal{C}(X_K = x_K)$ that matches one of the five configurations in Figure 13. One can choose a path π of type (a) or (b) if and only if (5.2) does not hold. For types (c) to (e), let $j' = j$, $i' = i$ for case (c), $j' = j$ for case (d), and j', i' be as-is for case (e). By definition of $\mathcal{C}(X_K = x_K)$, it holds that $\pi \subset \mathcal{C}(X_K = x_K)$ if and only if (5.3) holds for this particular triple of nodes i', j', k.   □

We are now ready for the proof of the main theorem of this section.

THEOREM 5.12 (Context-specific, fixed C).   *Let X be a max-linear Bayesian network over a directed acyclic graph $\mathcal{D} = (V, E)$ with fixed coefficient matrix C. Let $K \subseteq V$ and $\mathcal{C}(X_K = x_K)$ be the source DAG of the possible context $\{X_K = x_K\}$. Then for all $I, J \subseteq V$,*

$$I \perp_* J \mid K \quad in \ \mathcal{C}(X_K = x_K) \quad \Longrightarrow \quad X_I \perp\!\!\!\perp X_J \mid X_K = x_K$$

PROOF.   Suppose that I and J are source separated by $\{X_K = x_K\}$. By Lemma 5.11, this implies that

$$(I \cup \mathrm{pa}_C(I)) \cap (J \cup \mathrm{pa}_C(J)) = \varnothing,$$

and that there are no pairs $i' \in I \cup \mathrm{pa}_{\mathcal{C}}(I)$, $j' \in J \cup \mathrm{pa}_{\mathcal{C}}(J)$ that simultaneously appear in the same equation among those in (4.8) and (4.9). By Corollary 4.5, this implies

$$\{Z_i : i \in I \cup \mathrm{pa}_{\mathcal{C}}(I)\} \perp\!\!\!\perp \{Z_j : j \in J \cup \mathrm{pa}_{\mathcal{C}}(J)\} \mid X_K = x_K$$

and by the representation (4.5), this implies $X_I \perp\!\!\!\perp X_J \mid X_K = x_K$. □

EXAMPLE 5.13 (Tent).   Applying Theorem 5.12 to the source DAG in Figure 10 of Example 3.17 yields the conditional independence statement $X_3 \perp\!\!\!\perp (X_1, X_2) \mid X_4 = X_5 = 2$, as also stated in the Introduction; see Example 1.2.

5.3. *The context-free cases.*   In the previous subsection, we identified sufficient conditions for conditional independence given a specific possible context $\{X_K = x_K\}$. We now exploit this result to derive conditions for independence that are valid in any context.

THEOREM 5.14 (Context-free, fixed C).   *Let X be a max-linear Bayesian network over a directed acyclic graph $\mathcal{D} = (V, E)$ with fixed coefficient matrix C. Then for all $I, J, K \subseteq V$,*

$$I \perp_* J \mid K \quad \text{in } \mathcal{D}_K^*(C) \quad \Longrightarrow \quad X_I \perp\!\!\!\perp X_J \mid X_K.$$

PROOF.   It is enough to prove the result for $K \neq \varnothing$. Suppose that there are no $*$-connecting paths in $\mathcal{D}_K^*(C)$. For any possible context $\{X_K = x_K\}$, by Lemma 5.3, $\mathcal{D}_K^*(C) \supseteq \mathcal{C}(X_K = x_K)$, therefore, there is no $*$-connecting path in $\mathcal{C}(X_K = x_K)$. Thus we have $X_i \perp\!\!\!\perp X_j \mid X_K$ by Theorem 5.12. □

Finally, we can give the generic Markov condition which does not involve knowledge of the coefficient matrix C.

THEOREM 5.15 (Context-free, independent of C).   *Let X be a max-linear Bayesian network over a directed acyclic graph $\mathcal{D} = (V, E)$. Then for all $I, J, K \subseteq V$,*

$$I \perp_* J \mid K \quad \text{in } \mathcal{D}_K^* \quad \Longrightarrow \quad X_I \perp\!\!\!\perp X_J \mid X_K \quad \text{for all } C \text{ with support included in } \mathcal{D}.$$

PROOF.   By Lemma 5.3, $\mathcal{D}_K^* \supseteq \mathcal{D}_K^*(C)$, so if there are no $*$-connecting paths in $\mathcal{D}_K^*$, then there are also no $*$-connecting paths in $\mathcal{D}_K^*(C)$ for all C supported by $\mathcal{D}$. Thus $X_i \perp\!\!\!\perp X_j \mid X_K$ for all such C by Theorem 5.14. □

REMARK 5.   We note that Theorem 5.15 is corresponding to what is known as the *global Markov property* for Bayesian networks, that is, it establishes that separation in a suitable graph always implies conditional independence simultaneously for all possible values of the conditioning variables, and this statement holds for any choice of coefficient matrix C.

6. Completeness.   In this section, we shall investigate to what extent the separation criteria developed in Section 5 are complete for conditional independence in max-linear Bayesian networks, that is, yield all conditional independence relations that are valid. As before, we divide the discussion into the context-specific and context-free cases.

6.1. *The context-specific case.*   We first establish the converse to Theorem 5.12 in the context-specific case. The next lemma is used several times in the proof.

LEMMA 6.1.   *Suppose there is a $*$-connecting path between i and j in $\mathcal{C}(X_K = x_K)$ of types* (a) *or* (b) *in Figure 13. Suppose further for type* (b) *that there exists some $g \in \mathfrak{G}(X_K = x_K)$ such that $j' \to i$, $j' \to j \in g$ and $j' \notin K^*(g)$. Then $X_i \not\perp\!\!\!\perp X_j \mid X_K = x_K$.*

PROOF.   By Corollary 4.7(b), type (b) implies

$$\mathbb{P}\left\{\frac{X_i}{c_{ij'}^*} = \frac{X_j}{c_{jj'}^*} \neq \text{ an atomic value of } X_i \text{ or } X_j \,\Big|\, X_K = x_K\right\} \geq \mathbb{P}(g \mid X_K = x_K) > 0,$$

and in type (a), we have the same inequality with $j = j'$. In either case, $X_i \not\!\perp\!\!\!\perp X_j \mid X_K = x_K$, as claimed.   □

The main difficulty in proving Theorem 6.2 below is that having two edges $j' \to i$, $j' \to j \in \mathcal{C}(X_K = x_K)$ does not in general (as in the above Lemma 6.1) imply that there exists a compatible impact graph $g \in \mathfrak{G}(X_K = x_K)$, where both of these edges appear simultaneously. Indeed, Example 4.4 above shows that this need not be the case, whereas Corollary 4.6 establishes this fact in a specific case.

THEOREM 6.2 (Context-specific completeness).   *Let X be a max-linear Bayesian network over a directed acyclic graph $\mathcal{D} = (V, E)$ with fixed coefficient matrix C. Let $K \subseteq V$ and $\mathcal{C}(X_K = x_K)$ be the source DAG of a possible context $\{X_K = x_K\}$. For all subsets $I, J \subseteq V$, it holds that*

$$X_I \perp\!\!\!\perp X_J \mid X_K = x_K \quad \Longrightarrow \quad I \perp_* J \mid K \quad \text{in } \mathcal{C}(X_K = x_K)$$

PROOF.   To prove this, we separately consider the five different types of $*$-connectivity in Figure 13 and in each of them establish that the variables are dependent, using the node partition $V \setminus U = A \cup H \cup L$ and the representation as in Theorem 4.3 and its corollaries.

First, we claim that it is sufficient to consider the case where $I = \{i\}$ and $J = \{j\}$ are singletons with $i, j \in A$. For if I and J are $*$-connected, there must be $i \in I, j \in J$ such that i and j are $*$-connected, so if i and j are dependent, so are I and J.

Suppose then that i and j are $*$-connected, that is, there is a path π of the types shown in Figure 13. The proof considers the five different cases (a)–(e) of this figure in turn and gives an appropriate event for each one to establish conditional dependence.

Case (a): This follows directly from Lemma 6.1.

Case (b): For each $t = 1, \ldots, m$, let

$$\tilde{I}_t = \mathrm{pa}_{\mathcal{C}}(\ell_t) \cap \mathrm{pa}_{\mathcal{C}}(i), \qquad \tilde{J}_t = \mathrm{pa}_{\mathcal{C}}(\ell_t) \cap \mathrm{pa}_{\mathcal{C}}(j).$$

There are two mutually exclusive subcases.

6.1.0.1. *Case (b)I.   For each $t = 1, \ldots, m$, we have $\tilde{I}_t \cup \tilde{J}_t \subsetneq \mathrm{pa}_{\mathcal{C}}(\ell_t)$.*
In particular, for each such t, there exists some $r_t \in \mathrm{pa}_{\mathcal{C}}(\ell_t)$ such that

(6.1) $$r_t \notin \mathrm{pa}_{\mathcal{C}}(i) \cup \mathrm{pa}_{\mathcal{C}}(j).$$

Our goal is to construct an appropriate g and appeal to Lemma 6.1. Apply (4.7) to j', let $\beta_{j'}$ be the constant on the right-hand side of this inequality. For a sufficiently small constant $\epsilon > 0$, consider the event $\mathcal{E}$ defined by:

- $\beta_{j'} - \epsilon < Z_{j'} < \beta_{j'}$ ($Z_{j'}$ is only slightly smaller than the largest value possible in the context $\{X_K = x_K\}$)
- $Z_r < \epsilon$ for all $r \in \mathrm{pa}_{\mathcal{C}}(i) \cup \mathrm{pa}_{\mathcal{C}}(j) \setminus \{j'\}$ (any other parent of i or j, except j', has very small Z-value)
- $Z_i, Z_j < \epsilon$ (i and j also have very small Z values)
- for each $h \in H$, set $Z_{r'} < \epsilon$ for all $r' \in \mathrm{pa}_{\mathcal{C}}(h) \setminus \{j'\}$, and $Z_h = x_h$ (any node in h realizes itself: its parents have small Z-values, and its own Z-value is x_h.)

- for each ℓ_t for $t = 1, \ldots, m$, let r_t satisfy (6.1), and set it to achieve the maximum in (4.9). (Each block L_t gets a parent whose Z-value is not already constrained by the previous conditions.)

In the above, the only nodes that were mentioned but did not get set to be less than ϵ are $Z_{j'}$, Z_{r_t} for $t = 1, \ldots, m$ and Z_h for $h \in H$. By Proposition 3.18 and (6.1), these nodes are all distinct, so the event $\mathcal{E}$ is well defined. Furthermore, by Proposition 3.18 and Corollary 4.5, $\{Z_{r_t}, Z_h : t = 1, \ldots, m, h \in H\}$ are independent, and either $Z_{j'}$ is independent of $\{Z_{r_t}, Z_h : t = 1, \ldots, m, h \in H\}$, or it is independent of all but exactly one of them, say, Z_u for $u \in \{r_t : t = 1, \ldots, m\} \cup H$. In both cases, by Theorem 4.3 and Corollary 4.5, $\mathbb{P}(\mathcal{E} \mid X_K = x_K) > 0$. So there exists at least one $g \in \mathcal{C}(X_K = x_K)$ such that $\mathbb{P}(\mathcal{E}(g) \cap \mathcal{E} \mid X_K = x_K) > 0$. By construction of this event, $j' \to i$, $j' \to j \in g$ and $j' \notin K^*(g)$. Hence $X_i \not\perp\!\!\!\perp X_j \mid X_K = x_K$ by Lemma 6.1.

6.1.0.2. *Case (b)II.* *There exists at least one $t = 1, \ldots, m$ such that*

$$(6.2) \qquad\qquad \tilde{I}_t \cup \tilde{J}_t = \mathrm{pa}_{\mathcal{C}}(\ell_t).$$

Fix such a t. Define

$$\mathcal{E}_1 = \left\{ X_i < \min_{r \in \tilde{I}_t} \frac{c^*_{ir} x_{\ell_t}}{c^*_{\ell_t r}} \right\} \quad \text{and} \quad \mathcal{E}_2 = \left\{ X_j < \min_{r \in \tilde{J}_t} \frac{c^*_{jr} x_{\ell_t}}{c^*_{\ell_t r}} \right\}.$$

By Proposition 3.18(j), $\tilde{I}_t, \tilde{J}_t \neq \varnothing$, so the above events are well defined. Let r_0 denote a node $r \in \tilde{J}_t$ that achieves the minimum in $\mathcal{E}_2$ above. Since $r_0 \to j \in \mathcal{C}(X_K = x_K)$, there exists some $g \in \mathfrak{G}(X_K = x_K)$ such that $r_0 \to j$, $r_0 \notin K^*(g)$. This implies that on $\mathcal{E}(g)$,

$$Z_{r_0} < x_{\ell_t} / c^*_{\ell_t r_0}, X_j = c^*_{j r_0} Z_{r_0}.$$

Together these imply that on $\mathcal{E}(g)$,

$$X_j < \frac{c^*_{j r_0} x_{\ell_t}}{c^*_{\ell_t r_0}} = \min_{r \in \tilde{J}_t} \frac{c^*_{jr} x_{\ell_t}}{c^*_{\ell_t r}}.$$

So $\mathcal{E}(g) \subseteq \mathcal{E}_2$. Therefore, $\mathbb{P}(\mathcal{E}_2 \mid X_K = x_K) > 0$, and by symmetry, $\mathbb{P}(\mathcal{E}_1 \mid X_K = x_K) > 0$.

By (4.9) in Theorem 4.3, for each $g \in \mathfrak{G}(X_K = x_K)$, $\mathrm{pa}_g(\ell_t) \in \mathrm{pa}_{\mathcal{C}}(\ell_t)$. By (6.2), either $\mathrm{pa}_g(\ell_t) \in \tilde{I}_t$ or $\mathrm{pa}_g(\ell_t) \in \tilde{J}_t$; note that both can occur simultaneously as we are *not* claiming that $\tilde{I}_t \cap \tilde{J}_t = \varnothing$. Consider all g such that $\mathrm{pa}_g(\ell_t) \in \tilde{J}_t$. Let $r = \mathrm{pa}_g(\ell_t)$. By definition of the max-linear model,

$$X_j \geq c^*_{jr} Z_r = \frac{c^*_{jr} x_{\ell_t}}{c^*_{\ell_t r}} \quad \text{on } \mathcal{E}(g) \text{ for any } g \text{ such that } r = \mathrm{pa}_g(\ell_t) \in \tilde{J}_t.$$

In particular, for any g such that $\mathrm{pa}_g(\ell_t) \in \tilde{J}_t$,

$$\mathbb{P}(\mathcal{E}(g) \cap \mathcal{E}_2 | X_K = x_K) = 0.$$

By the same argument, for any g such that $\mathrm{pa}_g(\ell_t) \in \tilde{I}_t$,

$$\mathbb{P}(\mathcal{E}(g) \cap \mathcal{E}_1 | X_K = x_K) = 0.$$

But $\mathrm{pa}_g(\ell_t) \in \tilde{I}_t \cup \tilde{J}_t$ for all $g \in \mathfrak{G}(X_K = x_K)$ as mentioned above. Therefore, there is no $g \in \mathfrak{G}(X_K = x_K)$ such that $\mathcal{E}(g) \subseteq \mathcal{E}_1 \cap \mathcal{E}_2$. That is,

$$\mathbb{P}(\mathcal{E}_1 \cap \mathcal{E}_2 | X_K = x_K) = 0.$$

But $\mathbb{P}(\mathcal{E}_1 | X_K = x_K) > 0$, $\mathbb{P}(\mathcal{E}_2 | X_K = x_K) > 0$, so the events $\mathcal{E}_1$ and $\mathcal{E}_2$ are not independent conditioned on $\{X_K = x_K\}$. Therefore, $X_i \not\perp\!\!\!\perp X_j \mid X_K = x_K$.

Cases (c), (d) and (e): We may assume that cases (a) and (b) do not apply. In particular, (5.2) holds. For case (c), let

$$\mathcal{E}_1 = \left\{ X_i = \frac{x_k}{c^*_{ki}}, X_K = x_K \right\}, \qquad \mathcal{E}_2 = \left\{ X_j = \frac{x_k}{c^*_{kj}}, X_K = x_K \right\}.$$

For case (d), let

$$\mathcal{E}_1 = \left\{ X_i = \frac{x_k}{c^*_{ki}}, X_K = x_K \right\}, \qquad \mathcal{E}_2 = \left\{ X_j = \frac{c^*_{jj'} x_k}{c^*_{kj'}}, X_K = x_K \right\}.$$

For case (e), let

$$\mathcal{E}_1 = \left\{ X_i = \frac{c^*_{ii'} x_k}{c^*_{ki'}}, X_K = x_K \right\}, \qquad \mathcal{E}_2 = \left\{ X_j = \frac{c^*_{jj'} x_k}{c^*_{kj'}}, X_K = x_K \right\}.$$

We now claim that in all three cases we have

$$\mathbb{P}(\mathcal{E}_1 \mid X_K = x_K) > 0 \quad \text{and} \quad \mathbb{P}(\mathcal{E}_2 \mid X_K = x_K) > 0.$$

Indeed, these follow in case (c) from $i \to k$, $j \to k \in \mathcal{C}(X_K = x_K)$, and in cases (d) and (e) from Corollary 4.6 applied to the triples $k \leftarrow j' \to j$ and $k \leftarrow i' \to i$. By (3.10) in Lemma 3.13, any $g \in \mathfrak{G}(\mathcal{E}_1)$ must have $R_g(k) = R_g(i)$. Similarly, any $g \in \mathfrak{G}(\mathcal{E}_2)$ must have $R_g(k) = R_g(j)$. But (5.2) implies $R_g(i) \neq R_g(j)$ for all $g \in \mathfrak{G}(X_K = x_K)$. Therefore,

$$\mathbb{P}(\mathcal{E}_1 \mid \mathcal{E}_2, X_K = x_K) = \mathbb{P}(\mathcal{E}_2 \mid \mathcal{E}_1, X_K = x_K) = 0.$$

So $X_i \not\perp\!\!\!\perp X_j \mid X_K = x_K$ in each of the three cases, as needed. Since all cases have been considered, this concludes the proof. $\square$

6.2. *The context-free cases.* Next, we consider the context-free case for a given coefficient matrix C. We begin by showing that the direct converse to Theorem 5.14 is false, as demonstrated in the following example.

EXAMPLE 6.3. Consider the graph in Figure 15 with all edge weights equal to one. We have

$$X_5 = Z_5 \vee Z_1,$$
$$X_4 = Z_4 \vee Z_1 \vee Z_3,$$
$$X_2 = Z_2 \vee Z_3 \vee Z_5.$$

The important feature of this example is that $c^*_{21} = c^*_{25} c^*_{51}$, that is, there is a critical directed path from 1 to 2 that factors through K, so $1 \to 2 \notin \mathcal{D}^*_K(C)$ and $1 \to 2 \notin \mathcal{D}^*_K$. On the other hand, $\pi = 1 \to 4 \leftarrow 3 \to 2$ is a $*$-connecting path. Nevertheless, we claim below that $X_1 \perp\!\!\!\perp X_2 \mid X_{4,5}$.
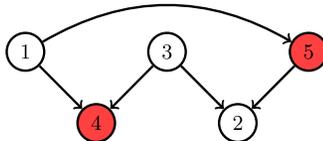
FIG. 16.   *The source DAG $\mathcal{C}(x_4, x_5)$ in a context satisfying $\{x_5 \geq x_4\}$ is a subgraph of the graph to the left and of the graph to the right if $\{x_5 < x_4\}$.*

Indeed, if $x_5 \geq x_4$, then also $x_5 \geq Z_3$ so $\mathcal{C}(x_4, x_5)$ is a subgraph of the graph to the left in Figure 16 On the other hand, if $x_5 < x_4$, then $1 \to 4 \notin \mathcal{C}(x_4, x_5)$ so that $\mathcal{C}(x_4, x_5)$ is a subgraph of the graph to the right in Figure 16.

In both cases, there is no $*$-connecting path between 1 and 2, hence by Theorem 5.12 we have $X_1 \perp\!\!\!\perp X_2 \mid X_{\{4,5\}}$.

We note that the phenomenon here has some similarity to "path cancellations" in standard linear Bayesian networks, where specific values of the coefficients may allow dependence relations to cancel and yield conditional independences, which do not follow from the separation criterion. Below we shall discuss and resolve this type of problem. In contrast to the standard case, the set of coefficients corresponding to such a phenomenon may have positive Lebesgue measure, which makes further discussion necessary. The key concept is that of an effective edge or path, as described below.

6.2.1. *Effective edges and paths.*   To obtain converses for the context-free cases, we wish to construct a possible context $\{X_K = x_K\}$ that violates the context-specific Markov condition. However, Example 6.3 above shows that this is not always possible. We need to ensure that no inequalities along $*$-connecting paths imply further equalities and to control this we need the following concept.

DEFINITION 6.4.   Let X be a max-linear Bayesian network over a directed acyclic graph $\mathcal{D} = (V, E)$ with fixed coefficient matrix C and $K \subset V$. For an edge $j \to i \in \mathcal{D}_K^*(C)$, the *substitution matrix* Ξ_K^{ij} of this edge relative to K is a $|K| \times |K|$ matrix with the following nonzero entries:

(6.3) $(\xi_K^{ij})_{k\ell} = \dfrac{c_{kj}^* c_{i\ell}^*}{c_{ij}^*}$ for $k \in K \cap \mathrm{ch}_{\mathcal{D}^*}(j), \ell \in K \cap (\mathrm{pa}_{\mathcal{D}^*}(i) \cup \{i\}), k \neq \ell.$

If π is a $*$-connecting path between i and j, then its *substitution matrix* Ξ_K^π relative to K is defined as

$$\Xi_K^\pi = \bigvee_{v \to u \in \pi} \Xi_K^{uv}.$$

REMARK 6.   Example 6.3 above features a path $\pi = 1 \to 4 \leftarrow 3 \to 2$ in $\mathcal{D}_K^*(C)$, but there is no x_K such that $\pi \subset \mathcal{C}(X_K = x_K)$. More importantly, as we show in Proposition 6.14 below, existence of such an x_K is equivalent to the additional condition (6.4) ensuring that the path is effective, as defined below.

DEFINITION 6.5.   A $*$-connecting path π from I to J in $\mathcal{D}_K^*(C)$ is said to be *effective* if it satisfies the tropical eigenvalue condition

(6.4) $\lambda(\Gamma_{KK} \vee \Xi_K^\pi) < 1,$

where Ξ_K^π is the substitution matrix of π with respect to K and Γ_{KK} is the restriction of the weak transitive closure $\Gamma(C)$ as in (2.3) to the components in K.

EXAMPLE 6.6. The condition (6.4) is necessary in general. In Example 6.3, we have a single $*$-connecting path π in $\mathcal{D}_K^*(C)$ between 1 and 2 and for this path (6.4) fails, as we shall now show. The substitution matrix for the path $\pi = 1 \to 4 \leftarrow 3 \to 2$ is

$$\Xi_K^\pi = \Xi_K^{41} \vee \Xi_K^{43} \vee \Xi_K^{23}.$$

We find positive entries

$$b_{54}^{41} = \frac{c_{51}^*}{c_{41}^*} = 1 \quad \text{and} \quad b_{45}^{23} = \frac{c_{43}^* c_{25}^*}{c_{23}^*} = 1,$$

so

$$\Gamma_{KK} \vee \Xi_K^\pi = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \vee \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and hence we get

$$\lambda\big(\Gamma_{KK} \vee \Xi_K^\pi\big) = 1,$$

which violates (6.4). Here, as noticed in Example 6.3, $X_i \perp\!\!\!\perp X_j \mid X_K$ despite the existence of a $*$-connecting path.

It turns out that condition (6.4) is often automatically satisfied. As we shall study effective edges in a specific context, we need the following concept.

DEFINITION 6.7. The *completion* of the coefficient matrix C with respect to a possible context $\{X_K = x_K\}$ is the $|V| \times |V|$ coefficient matrix $\bar{C}$, with

$$\bar{c}_{ij} = \begin{cases} x_i/x_j & \text{if } i, j \in K^*, \\ c_{ij} & \text{else.} \end{cases}$$

We write $\bar{C}^* = (\bar{c}_{kh}^*)$ for the Kleene star of $\bar{C}$ and note that all cycles in $\mathcal{D}(\bar{C})$ that only involve nodes in K^* have weight equal to one:

$$\bar{c}_{i_1 i_2} \bar{c}_{i_2 i_3} \cdots \bar{c}_{i_k i_1} = \frac{x_{i_1}}{x_{i_2}} \frac{x_{i_2}}{x_{i_3}} \cdots \frac{x_{i_k}}{x_{i_1}} = 1.$$

LEMMA 6.8. *Let $\bar{C}$ be the completion of C with respect to a possible context $\{X_K = x_K\}$. Then $\lambda(\bar{C}) = 1$.*

PROOF. If $|K| = 1$, this is obviously true for a self-loop. Assume that $|K| \geq 2$. Since $\mathcal{D}(C)$ is acyclic and all cycles in $\mathcal{D}(\bar{C})$ involving only nodes in K^* have length 1, it is sufficient to consider simple cycles $\pi = 1 \to 2 \cdots \to r \to 1$, with $1, r \in K^*$ and other nodes not in K^*. Write $\bar{c}(\pi)$ for the product of the edge weights of this cycle in $\bar{C}$. We claim that $\bar{c}(\pi) \leq 1$. Indeed,

$$\bar{c}(\pi) \leq c_{r2}^* c_{21}^* \bar{c}_{1r} = c_{r2}^* c_{21}^* \frac{x_1}{x_r} \leq \frac{c_{r1}^* x_1}{x_r} \leq 1,$$

where we have used that $c_{r1}^* \geq c_{21}^* c_{r2}^*$ and the context $\{X_K = x_K\}$ is possible, so $x_r \geq c_{r1}^* x_1$. Hence the maximum cycle mean is $\lambda(\bar{C}) = 1$, as desired. $\square$

COROLLARY 6.9. *For $k, h \in K^*$, we have that $\bar{c}_{kh}^* = \bar{c}_{kh} = x_k/x_h$.*

PROOF. If $h = k$, this is obviously true. Now assume $h \neq k$. By definition of the Kleene star, $\bar{c}_{kh}^* \geq \bar{c}_{kh}$ and since $\lambda(\bar{C}) = 1$, we also have $\lambda(\bar{C}^*) = 1$. Since $k \to h$ and $h \to k$ are edges in $\bar{C}$, we may consider the cycle $k \to h \to k$ and get

$$1 = \lambda(\bar{C}^*) \geq \bar{c}_{kh}^* \bar{c}_{hk}^* \geq \bar{c}_{kh} \bar{c}_{hk} = \frac{x_k}{x_h} \frac{x_h}{x_k} = 1.$$

Thus we must have equalities, that is, $\bar{c}_{kh}^* = \bar{c}_{kh}$ and $\bar{c}_{hk}^* = \bar{c}_{hk}$. $\quad\square$

DEFINITION 6.10. Say that an edge $j \to i$ in $\mathcal{D}_K^*(C)$ is *effective* in the possible context $\{X_K = x_K\}$ if $j \notin K^*$, no critical directed paths from j to i factor through K^*, and $c_{ij}^* = \bar{c}_{ij}^*$. Let $E^+(X_K = x_K)$ denote the set of effective edges in the context $\{X_K = x_K\}$. Edges in $\mathcal{D}_K^*(C)$ which are not effective are *ineffective*. Finally, a path π is *effective* in a context if all its edges are.

Now we give an algebraic characterization of edges that are effective in a context.

LEMMA 6.11. *Let* $j \to i \in \mathcal{D}_K^*(C)$ *and consider a possible context* $\{X_K = x_K\}$. *Then* $j \to i \in E^+(X_K = x_K)$ *if and only if for all* $k \in K^* \cap \mathrm{ch}_{\mathcal{D}^*}(j)$, $\ell \in K^* \cap (\mathrm{pa}_{\mathcal{D}^*}(i) \cup \{i\})$, *it holds that*

(6.5) $$(\xi_{K^*}^{ij})_{k\ell} x_\ell < x_k$$

for $\Xi_{K^*}^{ij}$ *being the substitution matrix relative to* K^* *as defined in* (6.3).

PROOF. Suppose $j \to i \in E^+(X_K = x_K)$. For each $k \in K^* \cap \mathrm{ch}_{\mathcal{D}^*}(j)$ and $\ell \in K^* \cap (\mathrm{pa}_{\mathcal{D}^*}(i) \cup \{i\})$, the path $j \to k \to \ell \to i$ (or $j \to k \to i$ if $i = \ell$) has $\bar{C}$-weight

$$c_{i\ell}^* \frac{x_\ell}{x_k} c_{kj}^*.$$

Since this path factors through K^*, it is not critical, so

$$c_{i\ell}^* \frac{x_\ell}{x_k} c_{kj}^* < c_{ij}^*.$$

Rearranging gives (6.5). Conversely, suppose that (6.5) holds. Let π be a path from j to i in $\mathcal{D}(\bar{C})$ that factors through K^*. If it only goes through one node of K^*, then it is also a path in C that factors through K^*, so $\bar{c}^*(\pi) = c^*(\pi)$. Since $j \to i \in \mathcal{D}_K^*(C)$, by definition of $\mathcal{D}_K^*(C)$, we have

$$\bar{c}^*(\pi) = c^*(\pi) < c_{ij}^*.$$

If π goes through two or more nodes of K^*, then without loss of generality we can assume

$$\pi = j \to \cdots \to k_1 \to \cdots \to k_2 \to \cdots \to k_r \to \cdots \to i,$$

where $r \geq 2$, $k_1, \ldots, k_r \in K^*$, and $\to \cdots \to$ are sequences of critical edges that do not go through K^*. By this criticality assumption, we get the equality

$$\bar{c}^*(\pi) = c_{k_1 j}^* \bar{c}_{k_2 k_1}^* \cdots \bar{c}_{k_r k_{r-1}}^* c_{i k_r}^*.$$

By Corollary 6.9,

$$\bar{c}_{k_2 k_1}^* \cdots \bar{c}_{k_r k_{r-1}}^* = \bar{c}_{k_r k_1}^* = \frac{x_{k_1}}{x_{k_r}}.$$

Note that $k_1 \in \mathrm{ch}_{\mathcal{D}^*}(j)$ and $k_r \in \mathrm{pa}_{\mathcal{D}^*}(i)$. Apply (6.5) with $k_1 = k$ and $k_r = \ell$ yields

$$\frac{c_{kj}^* c_{i\ell}^*}{c_{ij}^*} x_\ell < x_k.$$

Rearranging, we get

$$\bar{c}^*(\pi) = c_{kj}^* \frac{x_\ell}{x_r} c_{i\ell}^* < c_{ij}^*.$$

This shows that any critical path π that factors through K^* has weight strictly less than c_{ij}^*, as desired. □

A simple corollary is the following, showing that all edges in the source DAG for a given context are indeed effective in that context.

COROLLARY 6.12.   *If $j \to i$ is an edge in $\mathcal{C}(X_K = x_K)$, then $j \to i \in E^+(X_K = x_K)$.*

PROOF.   Assume that $j \to i \in \mathcal{C}(X_K = x_K)$ so we have $j \notin K^*$ and $c_{ij}^* > 0$. First, we claim that $j \to i \in \mathcal{D}_K^*(C)$. Indeed, suppose for contradiction that a critical path from j to i in $\mathcal{D}$ factors through some node $k \in K$, then

$$c_{ij}^* Z_j = c_{ik}^* c_{kj}^* Z_j \le c_{ik}^* x_k.$$

Since

$$X_i = c_{ik}^* x_k \vee c_{ij}^* Z_j \vee \dots,$$

this implies that $j \to i \notin \mathcal{C}(X_K = x_K)$, a contradiction as needed. Hence $j \to i \in \mathcal{D}_K^*(C)$.

Next, suppose for contradiction that $j \to i \notin E^+(X_K = x_K)$. By Lemma 6.11, this implies for some $k \in K^* \cap \mathrm{ch}_{\mathcal{D}^*}(j)$ and $\ell \in K^* \cap (\mathrm{pa}_{\mathcal{D}^*}(i) \cup \{i\})$,

$$(\xi_{K^*}^{ij})_{k\ell} x_\ell \ge x_k.$$

We apply the definition of the substitution matrix $\Xi_{K^*}^{ij}$ in (6.3) and rearrange; then we get

$$c_{i\ell}^* x_\ell \ge \frac{c_{ij}^*}{c_{kj}^*} x_k.$$

Since $j \to k \in \mathcal{D}_K^*(C)$, it holds that $x_k \ge c_{kj}^* Z_j$ so

$$\frac{c_{ij}^*}{c_{kj}^*} x_k \ge c_{ij}^* Z_j.$$

Since

$$X_i = c_{i\ell}^* x_\ell \vee c_{ij}^* Z_j \vee \dots,$$

it follows that $j \to i \notin \mathcal{C}(X_K = x_K)$, a contradiction as needed. The proof is complete. □

EXAMPLE 6.13.   Consider the graph to the left in Figure 17. Here, $C = C^*$. In this case, we have $X_1 \not\perp\!\!\!\perp X_4 \mid X_2, X_3$ although this is not true in all contexts. We first show that the
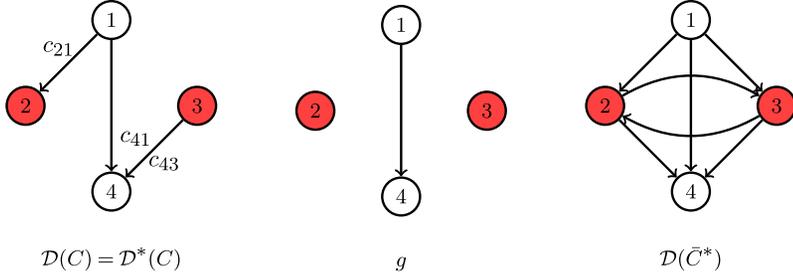
FIG. 17. *The graph to the left displays $\mathcal{D}(C) = \mathcal{D}(C^*)$ with coefficients. The impact graph g in the middle is not compatible with $\{X_{23} = x_{23}\}$ if $c_{43}c_{21}x_3 \geq c_{41}x_2$, as shown in Example 6.13, thus rendering the edge $1 \to 4$ ineffective. The graph to the right is the completion $\mathcal{D}(\bar{C}^*)$.*

graph g in the middle is not compatible with $\{X_K = x_K\}$ if $c_{43}c_{21}x_3 \geq c_{41}x_2$. We thus write out the max-linear model:

$$X_1 = Z_1,$$

$$x_3 = Z_3,$$

(6.6) $$x_2 = c_{21}Z_1 \vee Z_2,$$

(6.7) $$X_4 = c_{43}x_3 \vee c_{41}Z_1 \vee Z_4.$$

From (6.6), we have that $Z_1 \leq x_2/c_{21}$, so $c_{41}Z_1 \leq c_{41}x_2/c_{21}$. Thus, if $c_{41}x_2/c_{21} \leq c_{43}x_3$, or equivalently, $c_{41}x_2 \leq c_{43}c_{21}x_3$, we also have $c_{41}Z_1 < c_{43}x_3$, and hence (x_2, x_3) is not in the image of L_g so g is not compatible with the context.

Further, the support of $\bar{C}^*$ is shown to the right of Figure 17. With the addition of the edges $\bar{c}_{23} = x_2/x_3$ and $\bar{c}_{32} = x_3/x_2$, we have

$$\bar{c}_{41}^* = c_{41} \vee c_{43}\bar{c}_{32}c_{21} = c_{41} \vee c_{43}\frac{x_3}{x_2}c_{21}.$$

In particular, $1 \to 4$ is not effective w.r.t. $\{X_K = x_K\}$ if $c_{41} < c_{43}(x_3/x_2)c_{21}$.

REMARK 7. By definition of $\bar{C}$ and the critical graph $\mathcal{D}_{K^*}^*(C)$, if $j \to i \in E^+(X_K = x_K)$, then it holds that $j \to i \in \mathcal{D}_{K^*}^*(C) \subseteq \mathcal{D}_K^*(C)$. But the converse fails. That is, $E^+(X_K = x_K)$ can be a strictly smaller set of edges than those in $\mathcal{D}_K^*(C)$ or $\mathcal{D}_{K^*}^*(C)$.

The following says that if a path is effective in a context, it is effective in the sense of Definition 6.5. Note that, crucially, Definition 6.5 refers to the original set of conditioned variables K, while being effective in a given context $\{X_K = x_K\}$ is a property that involves the potentially bigger set K^* of variables which are constant in this context.

PROPOSITION 6.14. *Let π be a $*$-connecting path in $\mathcal{D}_K^*(C)$. If π is effective in a possible context $\{X_K = x_K\}$, then $\lambda(\Gamma_{KK} \vee \Xi_K^\pi) < 1$.*

PROOF. For each edge $j \to i \in \pi$, let Ξ_K^{ij} be the substitution matrix of this edge with respect to K (cf. Definition 6.4). Since $K \subseteq K^*(X_K = x_K)$, by Lemma 6.11,

$$\Xi_K^{ij} \odot x_K < x_K.$$

Thus

$$\left(\bigvee_{j \to i \in \pi} \Xi_K^{ij}\right) \odot x_K = \Xi_K^\pi \odot x_K \leq x_K.$$

Since x_K satisfies the max-linear model, we have

$$x_K = (C^* \odot x)_K \geq \Gamma_{KK} \odot x_K.$$

So

$$\left(\Gamma_{KK} \vee \Xi_K^\pi\right) \odot x_K = \Gamma_{KK} \odot x_K \vee \Xi_K^\pi \odot x_K \leq x_K.$$

By Proposition 2.1(a), this implies $\lambda(\Gamma_{KK} \vee \Xi^\pi) \leq 1$. Now we want to argue that this eigenvalue must be strictly less than 1. By Proposition 2.1(b), it is sufficient to show that there does not exist a cycle in $\mathcal{D}(\Gamma_{KK} \vee \Xi_K^\pi)$ with weight 1.

Suppose then for contradiction that there exists a cycle σ with weight $w(\sigma) = 1$, let S be its support, and let $A_{SS} = (\Gamma_{KK} \vee \Xi_K^\pi)_{SS}$. Since $\mathcal{D}(\Gamma_{KK})$ is a DAG and Ξ_K^π has zero diagonal, we must have $|S| \geq 2$. Again by Proposition 2.1(b), we have

$$A_{SS} \odot x_S = x_S.$$

Now consider an edge $v \to u \in \sigma$; by definition of A_{SS}, we have

$$a_{uv} = c_{uv}^* \vee \bigvee_{j \to i \in \pi, i \notin K} (\xi_K^{ij})_{uv} \vee \bigvee_{j \to i \in \pi, i \in K} (\xi_K^{ij})_{uv}.$$

By Lemma 2.2, we further have $a_{uv}x_v = x_u$ and by (6.3),

$$\bigvee_{j \to i \in \pi, i \notin K} (\xi_K^{ij})_{uv}x_v < x_u.$$

Thus

$$a_{uv} = c_{uv}^* \vee \bigvee_{j \to i \in \pi, i \in K} (\xi_K^{ij})_{uv}$$

for all edges $v \to u \in \sigma$. By definition, for $i \in K$, $(\xi_K^{ij})_{uv} > 0$ if and only if $v = i$. In other words, for each edge $v \to u \in \sigma$ such that $a_{uv} > c_{uv}^*$ it holds that $v \in K \cap \pi$. Since π is a $*$-connecting path, $|K \cap \pi| \leq 1$, there is at most one edge $v \to u$ of σ where $a_{uv} = (\xi_K^{vj})_{uv} > c_{uv}^*$, while for all other edges $v' \to u'$ of σ, $a_{u'v'} = c_{u'v'}^*$. Since $\mathcal{D}(C)$ is a DAG, there must be exactly one such edge. Therefore,

$$w(\sigma) = (\xi_K^{vj})_{uv}c_{vu_1}^*c_{u_1u_2}^* \cdots c_{u_ru}^* = (\xi_K^{vj})_{uv}c_{vu}^* = \frac{c_{uj}^*c_{vu}^*}{c_{vj}^*},$$

with $v, u \in K$, $c_{vu}^* > 0$ and $j \to v$, $j \to u \in \mathcal{D}_K^*(C)$. But $j \to u \to v$ is a path from j to v that factors through K. Since $j \to v \in \mathcal{D}_K^*(C)$, we have

$$c_{vu}^*c_{uj}^* < c_{vj}^*.$$

Rearranging gives $w(\sigma) < 1$, which is our desired contradiction. $\square$

6.2.2. *Context-free completeness.* To establish context-free completeness, we must understand the geometry of the set $\mathcal{L}_K^C$ defined in (2.6). For an edge $j \to i \in \mathcal{D}_K^*(C)$, we let $x_K(j \to i)$ be the set of x_K such that $j \to i$ is an effective edge in the possible context $\{X_K = x_K\}$. That is, we abbreviate

$$x_K(j \to i) = \{x_K : j \to i \in E^+(X_K = x_K)\}$$

and for a path π we similarly write

$$x_K(\pi) = \bigcap_{v \to u \in \pi} x_K(v \to u).$$

Now consider a $*$-connecting path π in $\mathcal{D}_K^*(C)$ and let

$$\Sigma(\pi) = \{x_K \in \mathcal{L}_K^C : (\Gamma_{KK} \vee \Xi_K^\pi) \odot x_K < x_K\}.$$

LEMMA 6.15. *Let π be a $*$-connecting path in $\mathcal{D}_K^*(C)$. Suppose there exists x_K such that all edges of π are effective in the possible context $\{X_K = x_K\}$. Then $\Sigma(\pi)$ is a nonempty full-dimensional subset of $\mathbb{R}_>^K$.*

PROOF. By Proposition 6.14, $\lambda(\Gamma_{KK} \vee \Xi_K^\pi) < 1$ so $\Sigma(\pi) \neq \varnothing$ by Proposition 2.1(c). Now, the smooth and invertible map $x \mapsto \log(x)$ has a regular total derivative and maps $\Sigma(\pi)$ to the relative interior of a classical polyhedron Q defined by strict inequalities of the form

$$y \in Q \quad \Longleftrightarrow \quad \text{for all } u, v \in V : y_v - y_u > \log\big((\Gamma_{KK} \vee \Xi_K^\pi)_{uv}\big),$$

where we have let $\log(0) = -\infty$. Thus, Q is an intersection of finitely many open half-spaces. Since $\Sigma(\pi) \neq \varnothing$ we have $Q \neq \varnothing$ and Q is open and full-dimensional. So $\Sigma(\pi)$ is open and full-dimensional. $\square$

PROPOSITION 6.16. *Consider a $*$-connecting path π in $\mathcal{D}_K^*(C)$ with $\Sigma(\pi) \neq \varnothing$. Then there exists some $x_K \in \Sigma(\pi)$ such that in the possible context $\{X_K = x_K\}$ with corresponding node partition $V = A \cup H \cup L \cup U$, we have $L = \varnothing$, $K^* = K$, and all edges of π are effective with respect to $\{X_K = x_K\}$.*

PROOF. For each $v \in V$ and each pair $h, k \in K$ such that $c_{hv}^*, c_{kv}^* > 0$, let

$$\mathcal{L}_{hkv} = \left\{ x_K : \frac{x_h}{c_{hi}^*} = \frac{x_k}{c_{ki}^*} \text{ for some } i \in V \right\} \quad \text{and} \quad \mathcal{L} = \bigcup_{h,k,v:\mathcal{L}_{hkv} \neq \varnothing} \mathcal{L}_{hkv}.$$

Note that $\mathcal{L}$ is a finite union of subspaces, each of codimension 1 in $\mathbb{R}_>^V$. By Lemma 6.15, $\Sigma(\pi)$ is full-dimensional and nonempty, so $\Sigma(\pi) \setminus (\mathcal{L} \cap \Sigma(\pi))$ is nonempty. Let x_K be in this set. Write $V = A \cup H \cup L \cup U$ w.r.t. the context $\{X_K = x_K\}$. Since $x_K \in \Sigma(\pi)$,

$$\Gamma_{KK} \odot x_K \le \big(\Gamma_{KK} \vee \Xi_K^\pi\big) \odot x_K < x_K.$$

Thus there are no pairs $h, k \in K$ with $h \neq k$ such that $x_h = c_{hk}^* x_k$. So $U = \varnothing$. In addition, $x_K \notin \mathcal{L}$. Thus by definition, $L = \varnothing$. So $K^*(X_K = x_K) = H$. Define the event $\mathcal{E} \subset \mathbb{R}_>^V$ via:

- $Z_k = x_k$ for all $k \in K$
- $\max_{k \in K: c_{ik}^* > 0} c_{ik}^* x_k < Z_i < \min_{k \in K: c_{ki}^* > 0} x_k / c_{ki}^*$ for all remaining $i \in V$ (where the minimum over an empty set is 0 and the maximum over an empty set is ∞).

Since $U = \varnothing$, for each i, $\max_{k \in K: c_{ik}^* > 0} c_{ik}^* x_k < \min_{k \in K: c_{ki}^* > 0} x_k / c_{ki}^*$. Thus $\mathcal{E}$ is well defined. By construction, $\mathcal{E} \subset \{X_K = x_K\}$ and it is full-dimensional w.r.t. this set. Thus there exists at least one $g \in \mathfrak{G}(X_K = x_K)$ such that $\mathcal{E}(g) \cap \mathcal{E} \neq \varnothing$. For this g, the only constant stars of g have roots in K and have no children. Thus $K^*(g) = K$. Since $K \subseteq K^* \subseteq K^*(g)$, it follows that $K^* = K$. Finally, since $x_K \in \Sigma(\pi)$,

$$\Xi_K^\pi \odot x_K \le \big(\Gamma_{KK} \vee \Xi_K^\pi\big) \odot x_K < x_K.$$

So in particular, for each edge $j \to i \in \pi$,

$$\Xi_K^{ij} \odot x_K < x_K.$$

Since $K = K^*(X_K = x_K)$, (6.5) holds. Lemma 6.11 then implies $j \to i \in E^+(X_K = x_K)$. $\square$

LEMMA 6.17. *Let π be a $*$-connecting path in $\mathcal{D}_K^*(C)$ with $\Sigma(\pi) \neq \varnothing$. Let $x_K \in \Sigma(\pi)$ that satisfies the conclusion of Proposition 6.16. Then $\pi \subseteq \mathcal{C}(X_K = x_K)$.*

PROOF.   Fix an edge $j \to i$ of π and x_K as above. Since π is $*$-connecting, $j \notin K$. There are two cases.

Case 1. $i \notin K$. We shall show that there exists some $g \in \mathfrak{G}(X_K = x_K)$ that contains the edge $j \to i$, and that this edge is not part of a constant star of g. To do this, we construct a region $\mathcal{E}$ in the manner similar to the proof of Theorem 6.2, case b(I). Apply (4.7) to j and let β_j be the constant on the right-hand side of this inequality. That is,

$$\beta_j = \min_{k \in K : k \in \mathrm{ch}_K^*(j)} \frac{x_k}{c_{kj}^*}.$$

Let

$$\gamma_j = \frac{1}{c_{ij}^*} \max_{\ell \in \mathrm{pa}_K^*(i)} x_\ell c_{i\ell}^*.$$

Since $j \to i \in E^+(X_K = x_K)$ and $i \notin K$, by (6.5) we have

$$(\xi_K^{ij})_{k\ell} x_\ell < x_k$$

for all $k \in \mathrm{ch}_{\mathcal{D}^*}(j)$ and $\ell \in \mathrm{pa}_{\mathcal{D}^*}(i)$. Rearranging gives

$$\frac{x_k}{c_{kj}^*} > \frac{x_\ell c_{i\ell}^*}{c_{ij}^*} \quad \text{for all } k \in \mathrm{ch}_K^*(j), \ell \in \mathrm{pa}_K^*(i),$$

or equivalently, $\beta_j > \gamma_j$. For a sufficiently small constant $\epsilon > 0$, consider the region $\mathcal{E}$ defined by:

- $\gamma_j < Z_j < \beta_j$
- for each $h \in H$, set $Z_h = x_h$
- $Z_r < \epsilon$ for all other nodes.

In the above, the only nodes that were mentioned but did not get set to be less than ϵ are Z_j and Z_h for $h \in H$. By Proposition 3.18 and (6.1), these nodes are all distinct. Since $\beta_j > \gamma_j$, Z_j is well defined. Since $L = \varnothing$, $\mathcal{E}$ is a nonempty polyhedron in $\mathbb{R}_>^V$, $\mathcal{E} \subseteq \{X_K = x_K\}$, and $\mathcal{E}$ is full-dimensional relative to the region $\{X_K = x_K\}$. Therefore, there exists at least one $g \in \mathfrak{G}(X_K = x_K)$ with $\mathcal{E}(g) \cap \mathcal{E} \neq \varnothing$. Now suppose $Z \in \mathcal{E}$. Then $Z_j > \gamma_j$ implies

$$c_{ij}^* Z_j > c_{i\ell}^* x_\ell$$

for all $\ell \in K$. In addition, $Z_j \gg \epsilon > Z_r, Z_i$ implies

$$c_{ij}^* Z_j > c_{ir}^* x_r$$

for all $r \neq j, r \notin K$ such that $c_{ir}^* > 0$. Thus $R_g(i) = j$, so in particular, $j \to i \in g$.

Since $Z_j < \beta_j$, it follows that

$$c_{ki}^* Z_j < x_k$$

for all $k \in K \cap \mathrm{ch}_K^*(j)$. Since $K^* = K$, $j \notin K^*(g)$, so $j \to i \notin E^-(X_K = x_K)$. Thus $j \to i \in \mathcal{C}(X_K = x_K)$. We are done.

Case 2. $i \in K$. Since $j \to i \in E^+(X_K = x_K)$, we can rearrange (6.5) to obtain

(6.8) $$\min_{k \in \mathrm{ch}_G(j), k \neq i} \frac{x_k}{c_{kj}^*} \geq \frac{x_i}{c_{ij}^*}.$$

As $L = \varnothing$, by definition of L, x_K satisfies the stronger inequality

(6.9) $$\min_{k \in \mathrm{ch}_G(j), k \neq i} \frac{x_k}{c_{kj}^*} > \frac{x_i}{c_{ij}^*}.$$

Since $i \in K$, it is sufficient to show that $j \to i \in \mathcal{I}(X_K = x_K)$. That is, we need to construct $g \in \mathfrak{G}(X_K = x_K)$ such that $j \to i \in g$. For a very small constant $\epsilon > 0$, consider the region $\mathcal{E}_1$ defined by:

- for each $h \in H$, $h \neq i$, set $Z_h = x_h$
- $Z_j = x_j / c_{ij}^*$
- $Z_r < \epsilon$ for all other nodes.

Since $L = \varnothing$, $\mathcal{E}$ is well defined and is full-dimensional relative to the region $\{X_K = x_K\}$. By (6.9), Z_j satisfies (4.7), so $\mathcal{E} \subset \{X_K = x_K\}$. Therefore, there exists some $g \in \mathfrak{G}(X_K = x_K)$ such that $\mathcal{E}(g) \cap \mathcal{E}_1 \neq \varnothing$. On $\mathcal{E}_1$, by construction, $j \to i \in g$ and the proof is complete. $\square$

For a given matrix C, we now have the following result. This implies that $\perp_{C^*}$ is not fully complete w.r.t. conditional independence for a given matrix C, as not all $*$-connecting critical paths may be effective, as seen in Example 6.3.

THEOREM 6.18 (Context-free completeness). *Let X be a max-linear Bayesian network over a directed acyclic graph $\mathcal{D} = (V, E)$ with fixed coefficient matrix C. It then holds that*

$$X_I \not\perp\!\!\!\perp X_J \mid X_K$$

if and only if there is an effective $*$-connecting path in the critical DAG $\mathcal{D}_K^*(C)$.

PROOF.   By Theorem 5.14 and Theorem 6.2, $X_I \not\perp\!\!\!\perp X_J \mid X_K$ if and only if there exists some $i \in I$, $j \in J$, some possible x_K, and some $*$-connecting path π between i and j such that $\pi \subseteq \mathcal{C}(X_K = x_K)$. Thus the statement in the proposition is equivalent to the claim that (6.4) holds if and only if there exists x_K such that $\pi \subseteq \mathcal{C}(X_K = x_K)$.

So suppose (6.4) holds. By Proposition 2.1(c), $\Sigma(\pi) \neq \varnothing$. By Proposition 6.16, we can pick a special x_K. Applying Lemma 6.17 to this special x_K, we conclude that $\pi \subseteq \mathcal{C}(X_K = x_K)$.

For the converse, suppose x_K is such that $\pi \subseteq \mathcal{C}(X_K = x_K)$. By Corollary 6.12, each edge of π is in $E^+(X_K = x_K)$. By Proposition 6.14, this implies that π_K is effective. $\square$

REMARK 8.   We note that we now could define yet another form of separation which we could name *effective $*$-separation* and by Theorem 6.18 this would now be strongly complete for conditional independence for a given C or, in other words, *any* max-linear Bayesian network would be faithful to this criterion. However, whereas checking critical separation is a simple task, checking effective $*$-separation would involve calculating $\lambda(\Gamma_{KK} \vee \Xi_K^\pi)$ for all $*$-connecting paths π, so we have preferred not to introduce this variant of separation in this article.

Finally we are able to establish full completeness of $\perp_{\mathcal{D}^*}$-separation for an unspecified coefficient matrix C.

THEOREM 6.19 (Completeness of $\perp_{\mathcal{D}^*}$-separation).   *Let X be a max-linear Bayesian network over a directed acyclic graph $\mathcal{D} = (V, E)$ and assume there is a $*$-connecting path in $\mathcal{D}_K^*$ between I and J. Then there is a coefficient matrix C with support included in $\mathcal{D}$ such that the corresponding max-linear Bayesian network satisfies*

$$X_I \not\perp\!\!\!\perp X_J \mid X_K.$$

PROOF.   Let π be a $*$-connecting path in $\mathcal{D}_K^*$ between I and J. For each of the five types, our goal is to construct a C such that $\pi \subset \mathcal{D}_K^*(C)$ and that (6.4) holds, that is, the path π is effective.

For each edge $v \to u \in \pi$, let $\pi_{uv} \subset \mathcal{D}$ be a path in $\mathcal{D}$ from v to u that does not factor through K. Define $C = C(\pi)$ as follows:

- If $a \to b \in \bigcup_{v \to u \in \pi} \pi_{uv}$, set $c_{ba} = 1$
- Otherwise, set c_{ba} to be some constant such that $c_{ba} < 1$.

First, we claim that for this choice of C, $\pi \subset \mathcal{D}_K^*(C)$. That is, for each edge $a \to b \in \pi$, no critical paths from a to b on C factor through K. Indeed, fix such an edge $a \to b \in \pi$. Let π'_{ba} be another path in $\mathcal{D}$. Then either π'_{ba} contains an edge not in $\bigcup_{u \to v \in \pi} \pi_{vu}$, in which case

$$c(\pi'_{ba}) < c(\pi_{ba}) = 1,$$

or that it only uses edges in $\bigcup_{v \to u \in \pi} \pi_{uv}$ and

$$c(\pi'_{ba}) = c(\pi_{ba}) = 1.$$

But in this case, since none of the paths π_{vu} factor through K, π'_{ba} does not factor through K. This establishes our first claim.

We now prove (6.4). Note that all relevant substitution matrices are formed by combining substitution matrices for single edges Ξ_K^{ij} for $j \notin K$ and we now claim that each entry of such a matrix is strictly less than 1.

As shown above, we must have $c_{ij}^* = 1$. Let $k \in K \cap \mathrm{ch}_{\mathcal{D}^*}(j)$, $\ell \in K \cap (\mathrm{pa}_{\mathcal{D}^*}(i) \cup \{i\})$, $k \neq \ell$ so we again have $c_{i\ell}^* = 1$. Since $k \notin \pi$, any path in $\mathcal{D}$ from j to k must utilize an edge of C whose weight is strictly less than 1 with the choice of C made above. Thus $c_{kj}^* < 1$. By (6.3),

$$\left(\Xi_K^{ij}\right)_{k\ell} = \frac{c_{kj}^* c_{i\ell}^*}{c_{ij}^*} = c_{kj}^* c_{i\ell}^* = c_{i\ell}^* < 1.$$

So each entry of Ξ_K^{ij} is strictly less than 1, as claimed, and hence this also holds for Ξ_K^π. Since $c_{uv} \leq 1$ for all edges $v \to u \in \mathcal{D}$, we have $\gamma_{uv} = c_{uv}^* \leq 1$ for all edges $v \to u \in \mathcal{D}^*$. Thus

$$\lambda\left(\Gamma_{KK} \vee \Xi_K^\pi\right) \leq 1.$$

Suppose now for contradiction that $\lambda(\Gamma_{KK} \vee \Xi_K^\pi) = 1$. Since all entries of Ξ_K^π are strictly less than 1, a critical cycle of $\Gamma_{KK} \vee \Xi_K^\pi$ must only involve edges in Γ_{KK}. But $\mathcal{D}$ is a DAG, so $\mathcal{D}(\Gamma_{KK})$ is cycle-free, yielding a contradiction. This concludes the proof. $\square$

7. Outlook.

7.1. *Properties of max-linear independence.* In the previous section, we have defined two abstract independence models $\perp_{\mathcal{D}^*}$ and $\perp_{C^*}$ in the sense of [27] and showed that they are sound and the former of them is also complete, whereas the latter needs additional conditions for completeness.

One can show without too much effort that these are both *compositional graphoids* (we refrain from giving the details) as also holds for most other graphical separation criteria (see, e.g., [23]). However, we should emphasize that $\perp_{\mathcal{D}^*}$ is *not strongly complete* as the Diamond example shows: in this example, the classical d-separation $\perp_{\mathcal{D}}$ and $\perp_{\mathcal{D}^*}$ coincide and there is no single coefficient matrix C such that the corresponding max-linear Bayesian network is faithful to $\perp_{\mathcal{D}^*}$, that is, in that case $\perp_{\mathcal{D}^*}$ is strictly weaker than critical separation $\perp_{C^*}$, and the same will happen for DAGs with more than a single directed path between any two points. But even in this case, the context-specific analysis typically yields further valid conditional independence statements.

Generally, the study of properties of conditional independence for max-linear models opens up several new avenues: concerning, for example, Markov equivalence as in [12, 28], or the algebraic properties of *maxoids* as an analogue of Gaussoids; see, for example, [6].

7.2. *Extensions and special cases.* We have so far in this article not discussed identifiability, estimation, or any other statistical issues associated with these models. These have been briefly considered in [16]; see also [14]. This work was extended to a recursive max-linear model with propagating noise in [7], but we are not considering models with noise in this article.

Extreme value models often rely on regular variation and several publications have combined Bayesian networks with such heavy-tailed innovations. In [17] and [20], algorithms have been proposed for statistically learning the model based on the estimated tail dependence matrix and on a scaling method, respectively. In [11], for undirected graphs the authors apply a peaks-over-threshold approach giving a multivariate generalized Pareto distribution for exceedances such that a density exists. For a decomposable graph, this density factorizes into lower dimensional marginal densities, whereas [9] deals with conditional densities.

Natural extensions in the framework of recursive max-linear models are based on making dependent innovations $(Z_1, \ldots, Z_d)$, thus defining the analogue of classical path analysis ([30, 31]), or *recursive causal models*; see [19]. The models introduced by [11] could be interesting candidates for this.

An alternative for an appropriate model may originate from multivariate max-stable Fréchet distributions with distribution function (see, e.g., [10], Section 6.1.4 and in particular Remark 6.1.16 with parametrization given in Theorem 1.1.3),

$$F(z) = \exp\left\{-\int_{\mathcal{S}^{d-1}} \bigvee_{1 \leq i \leq d} \frac{\omega_i}{z_i} \Theta(d\omega)\right\}, \quad z = (z_1, \ldots, z_d),$$

where $\omega = (\omega_1, \ldots, \omega_d) \in \mathcal{S}^{d-1}$, the unit sphere in $\mathbb{R}_+^d$ (with respect to any norm), and Θ is a finite measure on S^{d-1}, called the spectral measure. Then the innovation vector $(Z_1, \ldots, Z_d)$ has Fréchet margins with algebraically decreasing tails. If the spectral measure has a Lebesgue density, then the above integral becomes a Lebesgue integral. Then a large jump can happen in every direction with some probability. The Bayesian network introduces additional dependence into the model, which directs the large jumps in special directions.

7.3. *Some open problems.* Proposition 3.18 gives some necessary conditions for a graph to be the source DAG for some context $\{X_K = x_K\}$. It would be of interest to know whether these are also sufficient. Formally this is stated as Problem 1 below.

PROBLEM 1. Fix $\mathcal{D}$. Find a characterization for all possible source DAGs.

Further, even though we have a full characterization of situations with conditional independence, there is still an issue about how to verify conditional independence from a computational point of view. Formally, we state this as the following.

PROBLEM 2. Give an efficient algorithm to compute the source DAG $\mathcal{C}(X_K = x_K)$ and analyze its complexity.

Critical directed paths in a graph can be computed with tropical matrix multiplication [8], Section 3, and thus $\mathcal{D}_K^*$ and $\mathcal{D}_K^*(C)$ can both easily be computed in time at most $O(d^4)$. However, computing the source DAG is harder. A straightforward algorithm using the characterization of the impact graphs $\mathfrak{G}(X_K = x_K)$ in Lemma 3.13 goes as follows:

1. Enumerate all elements in $\mathfrak{G}(X_K = x_K)$ using the system of equations and inequalities given in Theorem 3.3 and Lemma 3.13, with $K^*(g)$ characterized by Lemma 3.11.
2. Compute $K^* = K^*(X_K = x_K)$ from $\mathfrak{G}(X_K = x_K)$ via Definition 3.10.

3. Compute the source DAG via Theorem 4.3.

Of these steps, step 1 is the most computationally intensive. The set $\mathfrak{G}(X_K = x_K)$ represents all possible hitting scenarios in [29]. For general C (not necessarily supported on a DAG), [29] noted that enumerating $\mathfrak{G}(X_K = x_K)$ is related to the NP-hard set covering problem. For our case, C is a DAG, so we were able to characterize $\mathfrak{G}(X_K = x_K)$ in much greater detail than [29]. However, it is unclear what is the complexity of enumerating this set. The difficulty is that the inequalities corresponding to (3.3), (3.9) and (3.10) depend on g. So while it is easy to check whether $g \in \mathfrak{G}(X_K = x_K)$ for a given g, there are exponentially many impact graphs g one needs to consider.

We remark that Problem 2 can be seen as finding the tropical analogue of Gaussian elimination. While there has been work on the tropical Fourier–Motzkin elimination [1], we are not aware of algorithms to solve tropical Gaussian elimination. The geometric relative of this problem is to find minimal external representations of tropical polyhedra, to which algorithms and characterizations in terms of hypergraphs have been developed, for example, in [2–4]. It would be interesting to deepen these connections between extreme value theory and tropical convex geometry. A related problem is the following.

PROBLEM 3.   Give an efficient algorithm to simulate from the conditional distribution of X, given a context $\{X_K = x_K\}$.

This problem was also considered by [29] and has particular interest for Bayesian inference about the unknown parameters of a max-linear Bayesian network. Most Markov chain Monte Carlo (MCMC) algorithms will have such a simulation step built in at some point. In addition, this could be of interest if an unknown source for an observed extreme event should be identified, potentially of interest in environmental science.

REFERENCES

[1] ALLAMIGEON, X., FAHRENBERG, U., GAUBERT, S., KATZ, R. D. and LEGAY, A. (2014). Tropical Fourier–Motzkin elimination, with an application to real-time verification. *Internat. J. Algebra Comput.* **24** 569–607. MR3254715 https://doi.org/10.1142/S0218196714500258

[2] ALLAMIGEON, X., GAUBERT, S. and GOUBAULT, É. (2013). Computing the vertices of tropical polyhedra using directed hypergraphs. *Discrete Comput. Geom.* **49** 247–279. MR3017909 https://doi.org/10.1007/s00454-012-9469-6

[3] ALLAMIGEON, X., GAUBERT, S. and KATZ, R. D. (2011). Tropical polar cones, hypergraph transversals, and mean payoff games. *Linear Algebra Appl.* **435** 1549–1574. MR2810655 https://doi.org/10.1016/j.laa.2011.02.004

[4] ALLAMIGEON, X. and KATZ, R. D. (2013). Minimal external representations of tropical polyhedra. *J. Combin. Theory Ser. A* **120** 907–940. MR3022620 https://doi.org/10.1016/j.jcta.2013.01.011

[5] BACCELLI, F. L., COHEN, G., OLSDER, G. J. and QUADRAT, J.-P. (1992). *Synchronization and Linearity*: *An Algebra for Discrete Event Systems. Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. Wiley, Chichester. MR1204266

[6] BOEGE, T., D'ALÌ, A., KAHLE, T. and STURMFELS, B. (2019). The geometry of gaussoids. *Found. Comput. Math.* **19** 775–812. MR3989713 https://doi.org/10.1007/s10208-018-9396-x

[7] BUCK, J. and KLÜPPELBERG, C. (2020). Recursive max-linear models with propagating noise. Available at arXiv:2003.00362.

[8] BUTKOVIČ, P. (2010). *Max-Linear Systems*: *Theory and Algorithms. Springer Monographs in Mathematics*. Springer, London. MR2681232 https://doi.org/10.1007/978-1-84996-299-5

[9] COOLEY, D., DAVIS, R. A. and NAVEAU, P. (2012). Approximating the conditional density given large observed values via a multivariate extremes framework, with application to environmental data. *Ann. Appl. Stat.* **6** 1406–1429. MR3058669 https://doi.org/10.1214/12-AOAS554

[10] DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory*: *An Introduction*. *Springer Series in Operations Research and Financial Engineering*. Springer, New York. MR2234156 https://doi.org/10.1007/0-387-34471-3

[11] ENGELKE, S. and HITZ, A. S. (2020). Graphical models for extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 871–932. MR4136498

[12] FRYDENBERG, M. (1990). The chain graph Markov property. *Scand. J. Stat.* **17** 333–353. MR1096723

[13] GEIGER, D., VERMA, T. S. and PEARL, J. (1990). Identifying independence in Bayesian networks. *Networks* **20** 507–534.

[14] GISSIBL, N. (2018). Graphical Modeling of Extremes: Max-linear models on directed acyclic graphs Ph.D. thesis, Technical University of Munich.

[15] GISSIBL, N. and KLÜPPELBERG, C. (2018). Max-linear models on directed acyclic graphs. *Bernoulli* **24** 2693–2720. MR3779699 https://doi.org/10.3150/17-BEJ941

[16] GISSIBL, N., KLÜPPELBERG, C. and LAURITZEN, S. L. (2021). Estimation, identifiability, and structure learning of recursive max-linear models. *Scand. J. Stat.* **48** 188–211. https://doi.org/10.1111/sjos.12446

[17] GISSIBL, N., KLÜPPELBERG, C. and OTTO, M. (2018). Tail dependence of recursive max-linear models with regularly varying noise variables. *Econom. Stat.* **6** 149–167. MR3797980 https://doi.org/10.1016/j.ecosta.2018.02.003

[18] JOSWIG, M. (2020). *Essentials of Tropical Combinatorics*. Springer, Heidelberg.

[19] KIIVERI, H., SPEED, T. P. and CARLIN, J. B. (1984). Recursive causal models. *J. Austral. Math. Soc. Ser. A* **36** 30–52. MR0719999

[20] KLÜPPELBERG, C. and KRALI, M. (2021). Estimating an extreme Bayesian network via scalings. *J. Multivariate Anal.* **181** 104672. MR4157685 https://doi.org/10.1016/j.jmva.2020.104672

[21] KLÜPPELBERG, C. and LAURITZEN, S. (2019). Bayesian networks for max-linear models. In *Network Science—an Aerial View* (F. Biagini, G. Kauermann and T. Meyer-Brandis, eds.) 79–97. Springer Nature, Cham.

[22] KOLLER, D. and FRIEDMAN, N. (2009). *Probabilistic Graphical Models*: *Principles and Techniques*. *Adaptive Computation and Machine Learning*. MIT Press, Cambridge, MA. MR2778120

[23] LAURITZEN, S. and SADEGHI, K. (2018). Unifying Markov properties for graphical models. *Ann. Statist.* **46** 2251–2278. MR3845017 https://doi.org/10.1214/17-AOS1618

[24] LAURITZEN, S. L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. The Clarendon Press, Oxford. MR1419991

[25] LAURITZEN, S. L., DAWID, A. P., LARSEN, B. N. and LEIMER, H-G. (1990). Independence properties of directed Markov fields. *Networks* **20** 491–505.

[26] MACLAGAN, D. and STURMFELS, B. (2015). *Introduction to Tropical Geometry*. *Graduate Studies in Mathematics* **161**. Amer. Math. Soc., Providence, RI. MR3287221 https://doi.org/10.1090/gsm/161

[27] STUDENÝ, M. (2005). *Probabilistic Conditional Independence Structures*. *Information Science and Statistics*. Springer, London. MR3183760

[28] VERMA, T. and PEARL, J. (1990). Equivalence and synthesis of causal models. In *Proceedings of the 6th Conference on Uncertainty in Artifical Intelligence* 220–227. MIT Press, Cambridge, MA.

[29] WANG, Y. and STOEV, S. A. (2011). Conditional sampling for spectrally discrete max-stable random fields. *Adv. in Appl. Probab.* **43** 461–483. MR2848386 https://doi.org/10.1239/aap/1308662488

[30] WRIGHT, S. (1921). Correlation and causation. *J. Agric. Res.* **20** 557–585.

[31] WRIGHT, S. (1934). The method of path coefficients. *Ann. Math. Stat.* **5** 161–215.