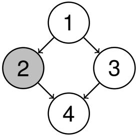
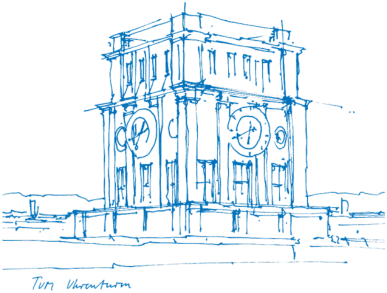
Conditional independence in max-linear Bayesian neworks: impact graphs, source graphs, and $*$-dependence

Claudia Klüppelberg
with Steffen Lauritzen and Ngoc Tran

Technical University of Munich

Applied Probability Days, 9 July, 2021



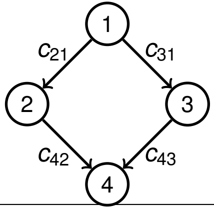


Max-linear Bayesian networks (MLBN) [Gissibl & K. (2018)]

Let $\mathcal{D} = (V, E)$ be a DAG and each node i represent a random variable X_i. Define the MLBN over $\mathcal{D}$ by the recursive ML structural equation system[1]

$$X_i := \bigvee_{k \in \mathrm{pa}(i)} c_{ki} X_k \vee c_{ii} Z_i \quad i = 1, \ldots, d$$

for independent innovations $Z_1, \ldots, Z_d > 0$, continuous, and structural coefficients $c_{ki} > 0$ (I set $c_{ii} = 1$).
In matrix form (max-times semiring): $X = C \odot X \vee Z$, with solution[2]



$$X = C^* \odot Z$$

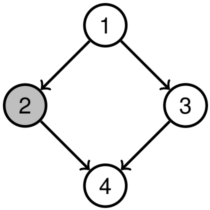
$$\text{and } C^* = \bigvee_{k=0}^{d-1} C^{\odot k} = (I_d \vee C)^{\odot(d-1)}$$

C^* is the Kleene star matrix.

[1] Pearl (2009)
[2] Butkovic (2010)

Separation criteria for conditional independence[3]

CI statements in (linear) Bayesian networks relate to *d*-separation:
If $I, J, K \subseteq V$ and all paths from I to J are blocked by K,
we say that K *d*-separates I from J, and we write $I \perp_d J \mid K$.



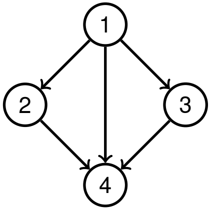
$1 \to 2 \to 4$ is blocked by 2
$1 \to 3 \to 4$ is not blocked by 2

Hence, $1 \not\perp_d 4 \mid 2$ and $X_1 \not\perp\!\!\!\perp X_4 \mid X_2$

This is not the correct separation criterion for a MLBN!

[3] Lauritzen (1996)

Interpretation of C^* by a path analysis[4]

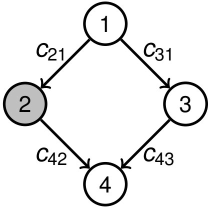


Reachability DAG $\mathcal{D}^*$

- An element c^*_{ij} of C^* is the maximal weight of all paths (weight = product of the coefficients) from j to i. Hence, C^* is a weighted reachability matrix, i.e., supported by the reachability DAG $\mathcal{D}^*$.
- A path in $\mathcal{D}$, which realizes c^*_{ij} is called a critical path. Then
  $$X_i = \bigvee_{j \in \mathrm{an}(i) \cup \{i\}} c^*_{ij} Z_j \quad i = 1, \ldots, d,$$
- We can remove any edge from $\mathcal{D}$, which is not part of a critical path, without changing the distribution of X.

[4]Wang and Stoev (2011), "hitting scenarios" (no graphs)

Diamond DAG



$1 \to 2 \to 4$ is critical $\Leftrightarrow c_{42}c_{21} \geq c_{43}c_{31}$. Then

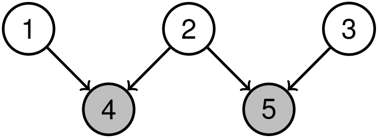
$$
\begin{aligned}
X_1 &= Z_1, \quad X_2 = c_{21}X_1 \vee Z_2, \\
X_4 &= c_{42}X_2 \vee Z_4 \vee c_{43}X_3 \\
&= c_{42}(Z_2 \vee c_{21}Z_1) \vee Z_4 \vee c_{43}(Z_3 \vee c_{31}Z_1) \\
&= c_{42}Z_2 \vee c_{42}c_{21}Z_1 \vee Z_4 \vee c_{43}Z_3 \vee c_{43}c_{31}Z_1 \\
&= c_{42}Z_2 \vee c_{42}c_{21}Z_1 \vee Z_4 \vee c_{43}Z_3 \\
&= c_{42}X_2 \vee Z_4 \vee c_{43}Z_3
\end{aligned}
$$

$X_1 \perp\!\!\!\perp X_4 \,|\, X_2$.

- This 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 CI statement $X_1 \perp\!\!\!\perp X_4 \,|\, X_2$, even if $1 \to 3 \to 4$ were also critical (that is, even if $c_{42}c_{21} = c_{43}c_{31}$).
- *d*-separation does not give all CI statements for a MLBN.
- There are more CI relations in a MLBN than in a (linear) Bayesian network.

Cassiopeia DAG

For the sake of the argument, I set all $c_{ij} = 1$.



$$X_1 = Z_1 \quad X_2 = Z_2 \quad X_3 = Z_3$$
$$X_4 = Z_1 \vee Z_2 \vee Z_4$$
$$X_5 = \quad Z_2 \vee Z_3 \vee Z_5$$

Q: $X_1 \perp\!\!\!\perp X_3 \mid \{X_4 = x_4, X_5 = x_5\}$

$x_4 > x_5$: then the causal ancestor of 4 is 1, and of 5 they are 2,3

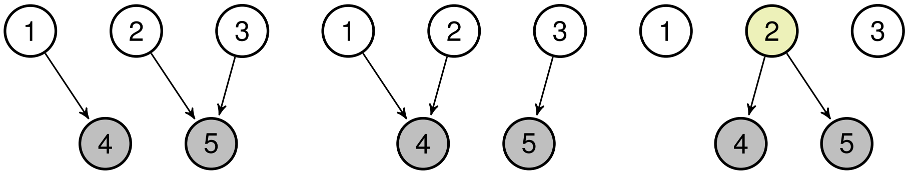
$x_4 < x_5$, then the situation is reversed

$x_4 = x_5$, then $X_2 = x_4 = x_5$ is a fixed node

Note: CI does not follow from the d-separation criterion, since the path from 1 to 3 is d-connecting relative to $\{4, 5\}$.

... continued

The three possible **causal graphs**:



$x_4 > x_5$ $x_4 < x_5$ $x_4 = x_5$

Then in all three situations,

$$X_1 \perp\!\!\!\perp X_3 \mid \{X_4 = x_4, X_5 = x_5\}$$

So the key is to identify causes.

Representation of $X_{\overline{K}} \mid X_K = x_K$

Proposition. Let $K \subset V$ and $\overline{K} = V \backslash K$. Write $X = (X_K, X_{\overline{K}})^\top$ as

$$\begin{pmatrix} X_K \\ X_{\overline{K}} \end{pmatrix} = \begin{pmatrix} C^*_{KK} & C^*_{K\overline{K}} \\ C^*_{\overline{K}K} & C^*_{\overline{K}\overline{K}} \end{pmatrix} \odot \begin{pmatrix} X_K \\ X_{\overline{K}} \end{pmatrix} \vee \begin{pmatrix} Z_K \\ Z_{\overline{K}} \end{pmatrix}$$

The model $X = C^* \odot Z$ with C^* Kleene star of C gives the representation

$$X_{\overline{K}} = C^*_{\overline{K}K} \odot x_K \;\vee\; C^*_{\overline{K}\overline{K}} \odot Z_{\overline{K}}$$

where $(Z_i, i \in \overline{K})$ are independent random variables with conditional distribution given $X_K = x_K$ determined by the restriction:

$$x_K \geq C^*_{K\overline{K}} \odot Z_{\overline{K}}.$$

Taking care of fixed nodes and redundant nodes and edges yields a reduced representation of the conditional distribution of $X_{\overline{K}} \mid X_K = x_K$.

Closer look into the structure of a MLBN: extreme impact

Definition. Let $\mathcal{D} = (V, E)$ be a DAG with coefficient matrix C with support $\mathcal{D}$ and Kleene star C^* with support $\mathcal{D}^*$. The impact graph is a random graph $G = G(Z)$ on V consisting of the following edges:

$$j \to i \quad \Longleftrightarrow \quad X_i = c^*_{ij} Z_j.$$

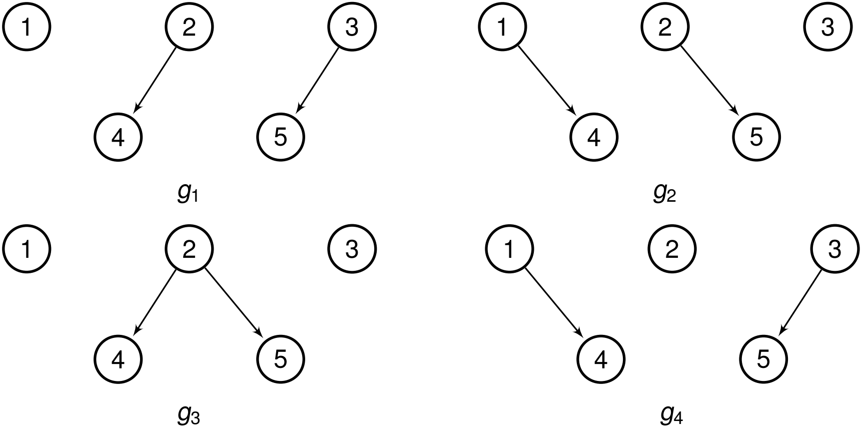
- Note that with probability one, each node has at most one parent, hence, $G(Z)$ is a forest of stars.
- We say that $j \to i$ belongs to g means that X_i is realised by Z_j.

Denote $R(g)$ the root set of the forest g and define

$$(C^*_g)_{ij} = \begin{cases} 1 & \text{if } i = j \in R(g) \\ c^*_{ij} & \text{if } j \to i \in g \\ 0 & \text{otherwise.} \end{cases}$$

Then $\quad X = C^* \odot Z \overset{\text{a.s.}}{=} C^*_G \odot Z = C^*_G Z.$

Impact graphs of the Cassiopeia DAG



The four impact graphs with two edges. If all coefficients are equal to 1, then only g_3 is compatible with $\{X_4 = X_5\}$,
whereas only g_2 and g_4 are compatible with the context $\{X_4 > X_5\}$.

Closer look into the structure of a MLBN: extreme impact continued ...

For a full characterization of all impact graphs of a MLBN we need:

Definition. Let $\mathcal{D} = (V, E)$ be a DAG with coefficient matrix C with support $\mathcal{D}$ and Kleene star C^*. Let g be a forest on V with root set $R = R(g)$. The impact exchange matrix $M = M(g) = M(g, C^*)$ of g with respect to C^* is an $R \times R$ matrix with entries

$$m_{rr} = 0 \quad \text{and} \quad m_{rr'} := \max_{i \in ch_g(r)} \frac{c^*_{ir'}}{c^*_{ir}} \quad \text{for } i \neq i'.$$

Here, $ch_g(r)$ denotes the children of root r in g.

- Intuitively, $m_{rr'}$ is the worst possible relative cost for a node i to be reassigned from root r to r' in g.
- Let $\lambda(M(g))$ be the principal tropical eigenvalue of $M(g)$.

Structure of impact graphs

Denote $\mathfrak{G} = \mathfrak{G}(C)$ all impact graphs of the MLBN.

Theorem. Consider a MLBN 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 $\mathcal{D}^*$.

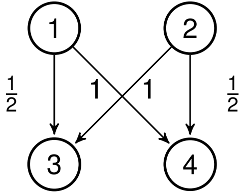
(b) g is a galaxy, i.e. a forest of stars (trees of height at most one).

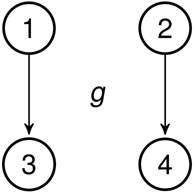
(c) If $j \to i$ in g and $c^*_{ij} = c^*_{ik} c^*_{kj}$ then $k \not\to i$ and $j \to k$ in g.

(d) $\lambda(M(g)) < 1$ (tropical eigenvalue condition).

- This theorem gives complete control of how extreme events from roots in g spread deterministically to other parts of the network.

Ilustration of the eigenvalue condition (d)





The subgraph g on the right cannot be an impact graph for the MLBN to the left, because

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

and $\lambda(M(g)) = 2 > 1$.
Indeed, $1 \to 3$ would imply $X_3 = \frac{1}{2}Z_1 > Z_2$,
but since $2 \to 4$ implies $X_4 = \frac{1}{2}Z_2 > Z_1$, this is inconsistent.

Closer look into the structure of a MLBN: extreme sources

Whereas impact graphs describe how extreme events spread in the network, the source graph $\mathcal{C}(X_K = x_k)$ tracks possible sources (causal ancestors) for a given event $\{X_k = x_k\}$.

The construction of the source graph $\mathcal{C}(X_K = x_k)$

- start from the total impact graph compatible with $\{X_k = x_k\}$:

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

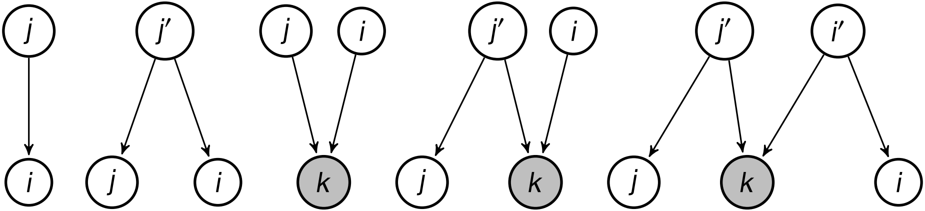
  where $\mathfrak{G}(X_K = x_K) = \{$impact graphs compatile with $\{X_k = x_k\}\}$,
- identify redundant nodes and edges, remove them from $\mathcal{I}(X_K = x_K)$.

Eventually, the source graph $\mathcal{C}(X_K = x_k)$ leads to a reduced representation of the conditional distribution of $X_{\overline{K}} \mid X_K = x_K$ of our Lemma above.

$\mathcal{C}(X_K = x_k)$ yields context-dependent CI relations.

Conditional independence by $*$-separation

An undirected path π between j and i in a DAG is $*$-connecting relative to $K \subset V$ (shaded nodes), if and only if it is one of following:



A path that is not $*$-connecting relative to K is said to be $*$-blocked by K. We also say that I and J are $*$-separated by K if all paths between I and J are $*$-blocked; then we write $I \perp_* J \,|\, K$.

Three conditional independence Theorems

For three different context situations we apply the same $*$-separation criterion, however, on three different DAGs.

Theorem. [Context-dependent, given $\{X_K = x_K\}$]

Let X be a MLBN over a DAG $\mathcal{D} = (V, E)$ with fixed coefficient matrix C. Let $K \subseteq V$ and $\mathcal{C}(X_K = x_K)$ be the source DAG of the possible context $\{X_K = x_K\}$. For all subsets $I, J \subseteq V$,

$$I \perp_* J \,|\, K \text{ in } \mathcal{C}(X_K = x_K) \iff X_I \perp\!\!\!\perp X_J \,|\, X_K = x_K.$$

Theorem. [Context-free, fixed C]

Let X be a MLBN over a DAG $\mathcal{D} = (V, E)$ with fixed coefficient matrix C. For all $I, J, K \subseteq V$, it then holds that

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

Three conditional independence Theorems

Theorem. [Context-free, independent of C]

Let X be a MLBN over a DAG $\mathcal{D} = (V, E)$. Then for all $I, J, K \subseteq V$,

$$I \perp_* J \,|\, K \text{ in } \mathcal{D}_K^* \iff X_I \perp\!\!\!\perp X_J \,|\, X_K \text{ for all } C \text{ with support included in } \mathcal{D}$$

Corollary. Recall that $\perp_d$ denote d-separation, then all following implications are strict:

$$\begin{aligned} I \perp_d J \,|\, K \implies I \perp_* J \,|\, K \text{ in } \mathcal{D}_K^* &\implies I \perp_* J \,|\, K \text{ in } \mathcal{D}_K^*(C) \\ &\implies I \perp_* J \,|\, K \text{ in } \mathcal{C}(X_K = x_K). \end{aligned}$$

Summary and conclusion

- A representation of $X_{\overline{K}} \mid X_K = x_k$ guides us to find a reduced representation of $X_{\overline{K}}$ taking deterministic features of a MLBN into account.
- Impact graphs describe how extreme events spread in the MLBN.
- The union of all impact graphs compatible with $\{X_K = x_K\}$ is the starting point for tracking possible sources of $\{X_K = x_K\}$.
- Cleaning up this union of graphs for fixed and redundant nodes and redundant edges yields the source graph $\mathcal{C}(X_K = x_K)$ giving a compact representation of the condional distribution given $X_K = x_K$.
- Our new $*$-separation criterion is equivalent to CI statements in context-free and context-dependent settings, which we formulate as $*$-separation in different derived DAGs.

References

- Améndola, C., Klüppelberg, C., Lauritzen, S. and Tran, N.M. (2020) Conditional independence in max-linear Bayesian networks. *Annals of Applied Probability*, accepted. arXiv:2002.09233.
- Améndola, C., Hollering, B., Sullivant, S. and Tran, N.M. (2021) Markov equivalence of max-linear Bayesian network. arXiv:2106.08305.
- Buck, J., Klüppelberg, C. and Tran, N.M. (2020) Learning max-linear Bayesian trees under measurement errors. In preparation. `https://media.ed.ac.uk/media/NgocTranEVATalkPreview/1_lfdxj4bb`
- Engelke, S. and Hitz, A.S. (2020) Graphical models for extremes. *JRSS B* **92**, 981-932
- Gissibl, N. and Klüppelberg, C. (2018) Max-linear models on directed acyclic graphs. *Bernoulli* **24**(4A), 2693-2720.
- Gissibl, N., Klüppelberg, C. and Lauritzen, S. (2020) Estimation, identifiability, and structure learning of recursive max-linear models. *Scand. J. Statistics* **47**.
- Hollering, B. and Sullivant, S. (2021) Discrete max-linear Bayesian networks. arXiv:2102.03426
- Lauritzen, S.L. (1996). *Graphical Models.* Oxford University Press.
- Pearl, J. (2009) *Causality: Models, Reasoning, and Inference (2nd Ed.)* Cambridge University Press.