Probabilistic graphical models and their application to extreme value statistics

Frank Röttger (University of Twente)

44th Finnish Summer School on Probability and Statistics
May 25-29, 2026

UNIVERSITY
OF TWENTE.

Structure of the Mini-Course

- **Part 1:** Undirected graphical models
- **Part 2:** Directed graphical models
- **Part 3:** Graphical models in extremes

Intended learning goals

- Get an introduction to the vast area of graphical models.
- In particular, learn about graphical models in extremes.

Part 2: Directed graphical models

Part 2: Table of contents

Motivation

- Undirected graphical models encode conditional independence, but not direction.
- In many applications, dependence is naturally asymmetric/ one-directional:

Example

Rain may impose Wet Ground and Umbrella use, but not vice versa

- Directed models provide a natural language for functional relationships, time ordering, and causality.

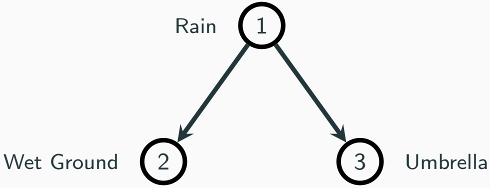


Figure 1: Directed graph

Directed graphical models

Graph terminology 2

- A node $i \in V$ is called a parent of a node $j \in V$ when $(i,j) \in E$ and $(j,i) \notin E$.
- Equivalently, a node $i \in V$ is called a child of a node $j \in V$ when $(j,i) \in E$ and $(i,j) \notin E$.
- We denote the set of parents of a node $i \in V$ as PA_i and the set of children as CH_i.
- The nodes i and j are called adjacent when either $(i,j) \in E$ or $(j,i) \in E$.
- We say that G is fully connected when all pairs of nodes are adjacent.
- A graph G is called directed when all its edges are directed.

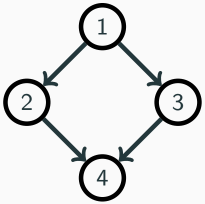


Figure 2: Diamond graph: $V = [4]$ and $E = \{(1,2),(1,3),(2,4),(3,4)\}$

$$\mathrm{PA}_4 = \{2,3\}$$
$$\mathrm{CH}_2 = \{4\}$$
$$\mathrm{PA}_1 = \emptyset$$

Graph terminology 2

- Three nodes, where one node is the child of the other two which are not adjacent, are called an immorality or v-structure.
- The skeleton of a directed graph is the undirected graph constructed by ignoring the directions.

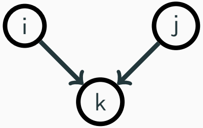


Figure 3: v-structure

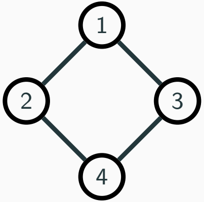


Figure 9: Skeleton of the diamond graph

Graph terminology 2

- A path is a sequence of distinct vertices $i_1, \ldots, i_m$, where an edge connects i_k and i_{k+1} for all $k \in [m-1]$.
- When a path contains three nodes with $i_{k-1} \to i_k$ and $i_{k+1} \to i_k$, we say that the path contains a collider.
- When $i_k \to i_{k+1}$ for all $k \in [m-1]$, we speak of a directed path.
- If there exists a directed path between i_j and i_ℓ, we call i_j an ancestor of i_ℓ and i_ℓ a descendant of i_j.
- A directed acyclic graph (DAG) is a directed graph without directed cycles, i.e. there is no pair i, k of vertices with a directed path from i to k and k to i.

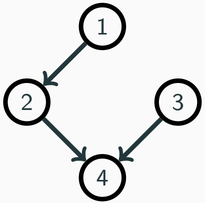


Figure 10: A path from 1 to 3 in the diamond graph

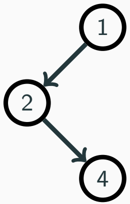


Figure 7: Directed path

Graph terminology 2

- A path is a sequence of distinct vertices $i_1, \ldots, i_m$, where an edge connects i_k and i_{k+1} for all $k \in [m-1]$.
- When a path contains three nodes with $i_{k-1} \to i_k$ and $i_{k+1} \to i_k$, we say that the path contains a collider.
- When $i_k \to i_{k+1}$ for all $k \in [m-1]$, we speak of a directed path.
- If there exists a directed path between i_j and i_ℓ, we call i_j an ancestor of i_ℓ and i_ℓ a descendant of i_j.
- A directed acyclic graph (DAG) is a directed graph without directed cycles, i.e. there is no pair i, k of vertices with a directed path from i to k and k to i.

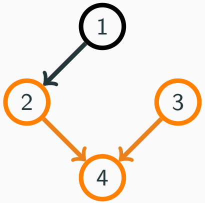


Figure 10: A path from 1 to 3 in the diamond graph with a collider

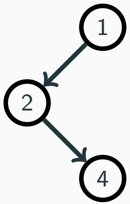


Figure 7: Directed path

Graph terminology 2

- The collection of ancestors of i is denoted as AN_i, not including i.
- The collection of descendants of i we call DE_i, it does not contain i.
- The collection of non-descendants is denoted as ND_i, it does not contain i. Note that $\mathrm{PA}_i \subseteq \mathrm{ND}_i$.
- A node without parents is called a source node, while a node without descendants is called a sink node.

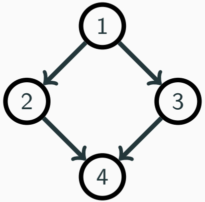


Figure 8: Diamond graph, an example for a DAG

$$\mathrm{AN}_4 = \{1, 2, 3\}$$
$$\mathrm{DE}_1 = \{2, 3, 4\}$$
$$\mathrm{ND}_2 = \{1, 3\}$$

d-separation

Definition

A path $i_1, i_2, \dots, i_m$ is blocked by a set S (where S contains neither i_1 nor i_m), whenever there is a node i_k such that one of the following holds:

1. $i_k \in S$ and

$$
\begin{aligned}
& i_{k-1} \rightarrow i_k \rightarrow i_{k+1}, \\
\text{or} \quad & i_{k-1} \leftarrow i_k \leftarrow i_{k+1}, \\
\text{or} \quad & i_{k-1} \leftarrow i_k \rightarrow i_{k+1},
\end{aligned}
$$

2. neither i_k nor any of its descendants is in S and

$$i_{k-1} \rightarrow i_k \leftarrow i_{k+1}.$$

Two disjoint sets $A, B \subset V$ are d-separated by a disjoint set $S \subset V$ when each path between nodes in A and B is blocked by S. This is denoted as

$$A \perp\!\!\!\perp_G B | S.$$

Exercise

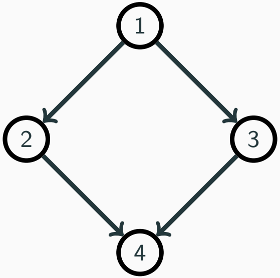


Figure 9: Diamond graph

- **Which set separates** 1 **and** 4**?**
- **Which set separates** 2 **and** 3**?**

d-separation: Example

- We observe that the set $\{2, 3\}$ blocks both paths from 1 to 4. Hence, we have

$$1 \perp\!\!\!\perp_G 4 | \{2, 3\}.$$

- Furthermore, the singleton 1 blocks all paths between 2 and 3 because of the collider $2 \rightarrow 4 \leftarrow 3$. Hence,

$$2 \perp\!\!\!\perp_G 3 | 1.$$

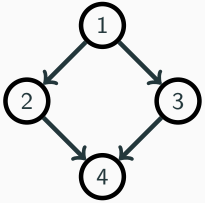


Figure 10: Diamond graph

Directed Markov properties

Definition

Let $G = (V, E)$ be a DAG and $X = (X_i)_{i \in V}$ a random vector. We say that $\mathbb{P}_X$ satisfies

1. the global Markov property with respect to G when

$$A \perp\!\!\!\perp_G B | C \Longrightarrow A \perp\!\!\!\perp B | C$$

   for all disjoint $A, B, C \subset V$,

2. the local Markov property with respect to G when $i \perp\!\!\!\perp (\mathrm{ND}_i \setminus \mathrm{PA}_i) \,|\, \mathrm{PA}_i$ for each $i \in V$,
3. the Markov factorization property with respect to G when

$$p(\mathrm{x}) = \prod_{j=1}^{d} p(x_j | \mathrm{x}_{\mathrm{PA}_j}),$$

   where we assume that $\mathbb{P}_X$ has a density $p(\mathrm{x})$.

Theorem

If $\mathbb{P}_X$ has a density, the directed global, local and factorization properties are equivalent.

Example

- The diamond graph to the right satisfies

$$1 \perp\!\!\!\perp_G 4 | \{2, 3\}, \quad 2 \perp\!\!\!\perp_G 3 | 1.$$

- A distribution $\mathbb{P}_X$ is global Markov with respect to the diamond when

$$X_1 \perp\!\!\!\perp X_4 | \{X_2, X_3\}, \quad X_2 \perp\!\!\!\perp X_3 | X_1.$$

- Note that the local Markov property gives precisely the same.
- Finally, we have a factorization

$$p(x) = p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_4|x_2, x_3).$$

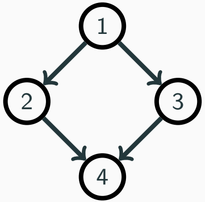


Figure 11: Diamond graph

Markov equivalence

Definition

The set of all distributions Markov to a DAG G is defined as

$$\mathcal{M}(G) := \{\mathbb{P}_X : \mathbb{P}_X \text{ satisfies the global Markov property with respect to } G\}.$$

Two DAGs G_1 and G_2 are Markov equivalent when $\mathcal{M}(G_1) = \mathcal{M}(G_2)$.

- Two DAGs are Markov equivalent if they allow the same set of d-separations, which via the global Markov property implies that they encode the same conditional independence statements.
- The set of all DAGs that are Markov equivalent to the same DAG G is called the Markov equivalence class of G.

Lemma

Two DAGs G_1 and G_2 are Markov equivalent if and only if they have the same skeleton and the same v-structures.

Example: The diamond graph rooted in 1 satisfies

$$1 \perp\!\!\!\perp_G 4 | \{2,3\}, \quad 2 \perp\!\!\!\perp_G 3 | 1.$$

The diamond graph with flipped edge $3 \to 1$ satisfies precisely the same d-separations. Note that the skeleton and the v-structures are identical!

The diamond graph rooted in 4 satisfies

$$1 \perp\!\!\!\perp_G 4 | \{2,3\}, \quad 2 \perp\!\!\!\perp_G 3 | 4.$$

Note that the skeleton is identical, but the v-structures are different!

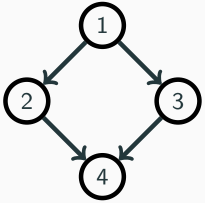


Figure 12: Diamond graph

Markov equivalence

Lemma

Two DAGs G_1 and G_2 are Markov equivalent if and only if they have the same skeleton and the same v-structures.

Example: The diamond graph rooted in 1 satisfies

$$1 \perp\!\!\!\perp_G 4 | \{2, 3\}, \quad 2 \perp\!\!\!\perp_G 3 | 1.$$

The diamond graph with flipped edge $3 \to 1$ satisfies precisely the same d-separations. Note that the skeleton and the v-structures are identical!

The diamond graph rooted in 4 satisfies

$$1 \perp\!\!\!\perp_G 4 | \{2, 3\}, \quad 2 \perp\!\!\!\perp_G 3 | 4.$$

Note that the skeleton is identical, but the v-structures are different!

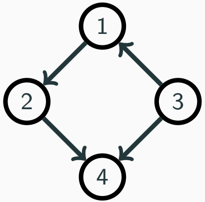


Figure 3: Diamond graph with flipped edge $3 \to 1$

Lemma

Two DAGs G_1 and G_2 are Markov equivalent if and only if they have the same skeleton and the same v-structures.

Example: The diamond graph rooted in 1 satisfies

$$1 \perp\!\!\!\perp_G 4 | \{2, 3\}, \quad 2 \perp\!\!\!\perp_G 3 | 1.$$

The diamond graph with flipped edge $3 \to 1$ satisfies precisely the same d-separations. Note that the skeleton and the v-structures are identical!

The diamond graph rooted in 4 satisfies

$$1 \perp\!\!\!\perp_G 4 | \{2, 3\}, \quad 2 \perp\!\!\!\perp_G 3 | 4.$$

Note that the skeleton is identical, but the v-structures are different!

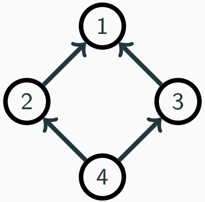


Figure 4: Example for a diamond graph rooted in 4

Partially directed acyclic graph (PDAG)

- A graph is called partially directed acyclic graph (PDAG) if there is no directed cycle in the graph.
- Note that a PDAG is a DAG when all edges are directed.
- A Markov equivalence class $\mathcal{M}(G)$ can be represented by a completed PDAG (CPDAG), which contains an edge $(i, j) \in V \times V$ if and only if there exists a DAG $\bar{G} \in \mathcal{M}(G)$ such that $(i, j) \in E(\bar{G})$.
- The CPDAG allows a graphical representation of the Markov equivalence class!

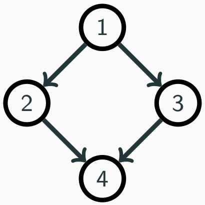


Figure 5: Diamond graph

Partially directed acyclic graph (PDAG)

- A graph is called partially directed acyclic graph (PDAG) if there is no directed cycle in the graph.
- Note that a PDAG is a DAG when all edges are directed.
- A Markov equivalence class $\mathcal{M}(G)$ can be represented by a completed PDAG (CPDAG), which contains an edge $(i, j) \in V \times V$ if and only if there exists a DAG $\bar{G} \in \mathcal{M}(G)$ such that $(i, j) \in E(\bar{G})$.
- The CPDAG allows a graphical representation of the Markov equivalence class!

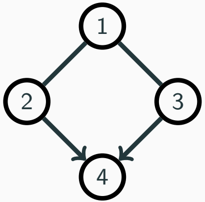


Figure 6: CPDAG of the diamond graph's Markov equivalence class

Linking directed and undirected graphical models: The moral graph

- The moral graph of a directed graph is constructed by adding edges in the skeleton between nodes that are parents in a v-structure.
- A perfect DAG is a DAG where the skeleton equals the moral graph.
- A random vector that is Markov to a DAG satisfies the undirected Markov property with respect to its moral graph, which is also decomposable. For perfect DAGs, the models are equivalent.

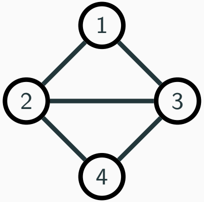


Figure 7: Moral graph of the diamond graph

Structural equation models

Multivariate structural equation models

Definition

A structural equation model (SEM) consists of a collection of d (structural) assignments

$$X_j := f_j(X_{\mathrm{PA}_j}, N_j),$$

where the noise variables $N_1, \dots, N_d$ are jointly independent.

- In causal inference, SEMs are considered as structural causal models (SCMs).
- The parents X_{PA_j} are sometimes called direct causes, and X_j a direct effect of them.
- When all f_j are linear functions, we also speak of a (linear) structural equation model.

Proposition

Let X be an SEM with respect to a DAG G. Then, $\mathbb{P}_X$ is Markov to G.

Linear SEMs: Example

An example for a linear SEM is

$$
\begin{aligned}
X_1 &:= N_1, \\
X_2 &:= b_{12}X_1 + N_2, \\
X_3 &:= b_{13}X_1 + N_3, \\
X_4 &:= b_{24}X_2 + b_{34}X_3 + N_4,
\end{aligned}
$$

where $b_{ij} \in \mathbb{R}$.
We observe that

$$X = B\,X + \mathrm{N},$$

where B is the matrix with entries b_{ij} and zeros for non-edges, and N is the vector of errors. We derive

$$X = (I - B)^{-1}\mathrm{N},$$

when $(I - B)$ is invertible.

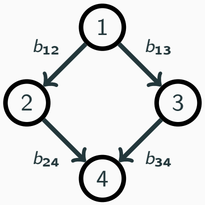


Figure 8: Diamond graph with edge weights of a linear model

If N is jointly Gaussian, we obtain

$$\mathrm{Cov}(X) = (I - B)^{-1}\,\mathrm{Cov}(N)(I - B)^{-T}.$$

Learning directed graphical models

Learning directed graphical models

- Parameter learning: For linear Gaussian SEMs $\Rightarrow$ Linear regression.
  $\rightarrow$ One reason why these models are so popular.
- Structure learning: Many different approaches. We will briefly discuss methods based on conditional independence.
  $\rightarrow$ Structure learning based on CI requires a (reasonable) assumption: faithfulness.

Faithfulness

Definition

A distribution $\mathbb{P}_X$ is faithful to a DAG G when

$$X_A \perp\!\!\!\perp X_B | X_C \Rightarrow A \perp\!\!\!\perp_G B | C$$

for all disjoint index sets A, B, C.

- Faithfulness is the opposite implication of the global Markov condition.
- If the coefficients of a linear SCM are drawn randomly from positive densities, the SCM is faithful to its DAG with probability 1.
- An example for a non-faithful linear SCM:

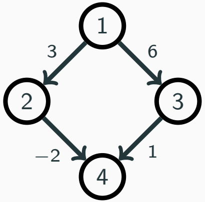


Figure 9: Example for a linear Gaussian SCM that is not faithful

This violates faithfulness because

$$X_1 \perp\!\!\!\perp X_4 \not\Rightarrow 1 \perp\!\!\!\perp_G 4.$$

Independence-based methods

Rough procedure for independence-based structure learning:

- Assume that a random vector is faithful and Markov to an underlying DAG.
- d-separation and conditional independences are one-to-one:
  1. Learn the skeleton.
  2. Learn the v-structures.
  3. Use edge orientation rules to find further directed edges.
- In applications, we need to test conditional independence statements given finite data.
- If we assume that a "CI oracle" provides us with the correct CI statements, this is guaranteed to recover the correct Markov equivalence class.

The following lemma is useful for estimating the skeleton of the CPDAG.

Lemma

(i) Two nodes i, j in a DAG $G = (V, E)$ are adjacent if and only if they cannot be d-separated by any set $S \subseteq V \setminus \{i, j\}$.

(ii) When two nodes i, j in a DAG are not adjacent, then they are d-separated either by PA_i or PA_j.

Consequences for structure learning:

- If two variables are not independent for any conditioning set S, then they must be adjacent.
- The search space for conditioning sets is limited to subsets of neighboring vertices.

Meek's edge orientation rules

- After learning the skeleton and the v-structures, logical reasoning allows to orient further edges.
- It was shown that the following rules, if repeatedly applied, are sufficient to recover the CPDAG, and thus the Markov equivalence class.

Meek's edge orientation rules:

1. Orient $j - k$ into $j \to k$ when there is an arrow $i \to j$ such that i and k are non-adjacent.
2. Orient $i - j$ into $i \to j$ whenever there is a chain $i \to k \to j$.
3. Orient $i - j$ into $i \to j$ whenever there are two chains $i - k \to j$ and $i - \ell \to j$ such that k and ℓ are non-adjacent.
4. Orient $i - j$ into $i \to j$ whenever there are two chains $i - k \to \ell$ and $k \to \ell \to j$ such that k and j are non-adjacent and i and ℓ are adjacent.

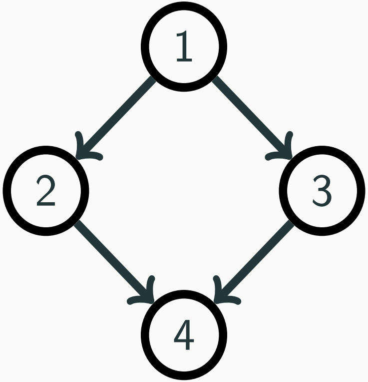
PC algorithm

Input: A CI oracle for a random vector X that is Markov and faithful to a DAG $G = (V, E)$.
Output: A CPDAG $\hat{G}$.

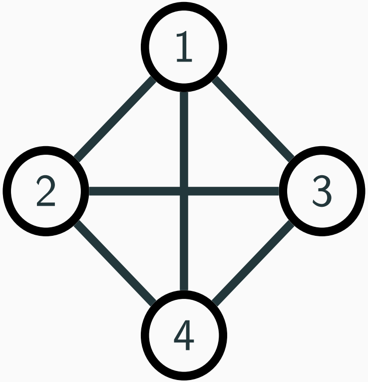
1. Begin with a full graph on V and for each pair $i, j \in V$, remove the corresponding edge when marginal independence holds between them.
2. Remove a remaining edge $i - j$ when a set S_{ij} of cardinality 1 contained in the set of vertices adjacent to i and j exists such that $X_i \perp\!\!\!\perp X_j | X_{S_{ij}}$.
3. Repeat step 2 for growing cardinality.
4. For each non-adjacent pair $i, j \in V$ with a common neighbor k, add a v-structure $i \to k \leftarrow j$ when $k \notin S_{ij}$.
5. Direct edges according to Meek's orientation rules.

Example

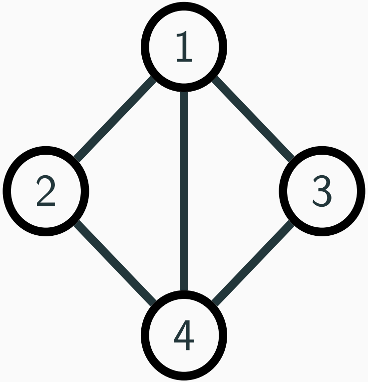
Let X be faithful and Markov to the diamond graph, i.e. we have $X_1 \perp\!\!\!\perp X_4 | \{X_2, X_3\}$, $X_2 \perp\!\!\!\perp X_3 | X_1$.



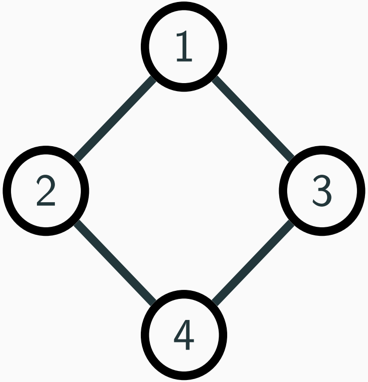
Step 1: Begin with a full graph. Observe that no marginal independence holds.



Step 2: For $S_{23} = 1$, we remove the edge $2 - 3$.

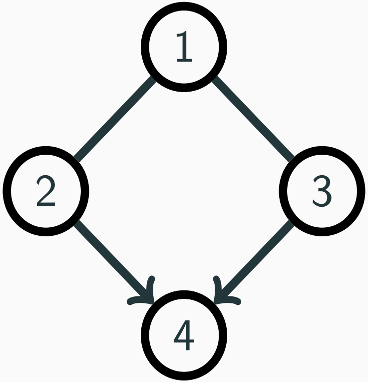


Step 3: For $S_{14} = \{2, 3\}$, we remove the edge $1 - 4$. Note the reduced search space.



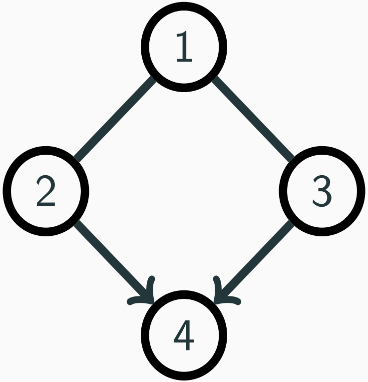
Example

Step 4: Add a v-structure because 2 and 3 are not adjacent and $4 \notin S_{23}$.



Example

Step 5: Check Meek's orientation rules. No further edge orientations are possible.



Score-based structure learning

Score-based structure learning

Instead of testing conditional independences, test whether graph structures fit the data.

Definition (Best scoring graph)

Let $\mathcal{D} = (X^1, \ldots, X^n)$ be an iid sample of a random vector X. Given a scoring function $S(\mathcal{D}, G)$ for each DAG G, the best scoring graph is defined as

$$\hat{G} = \operatorname*{argmax}_G S(\mathcal{D}, G).$$

It is typical to assume a parametric model for X, for example the multivariate Gaussian.

Greedy search techniques

- The number of DAGs grows super-exponentially in the number of variables.
- Searching over all DAGs may be infeasible.
- Greedy algorithms compare the score of a candidate graph with neighboring graphs, and terminate when it achieves the highest score.
- The definition of "neighboring graphs" clearly is essential.

Summary part 2

Summary part 2

- We introduced DAGs, d-separation and directed Markov properties.
- Using conditional independence, we can only determine the Markov equivalence class!
- Structural equation models are Markov to their underlying DAG.
- With a CI oracle, the PC algorithm recovers the Markov equivalence class.

What we did not discuss:

- Stronger restrictions permit DAG identification (e.g., LiNGAM).
- Conditional independence testing is difficult.
- Directed latent variable models.
- Causal interpretation of directed graphical models.

References

Drton, M. and M. H. Maathuis (2017).
Structure learning in graphical modeling.
Annual Review of Statistics and Its Application 4(1), 365–393.

Lauritzen, S. L. (1996).
***Graphical models*, Volume 17 of *Oxford Statistical Science Series*.**
The Clarendon Press, Oxford University Press, New York.
Oxford Science Publications.

Peters, J., D. Janzing, and B. Schölkopf (2017).
***Elements of causal inference: foundations and learning algorithms*.**
The MIT Press.

Thank You!