Title: Graphical models: fundamentals, origins, and beyond

Speakers: Steffen Lauritzen

Collection: Causal Inference & Quantum Foundations Workshop

Date: April 17, 2023 - 11:30 AM

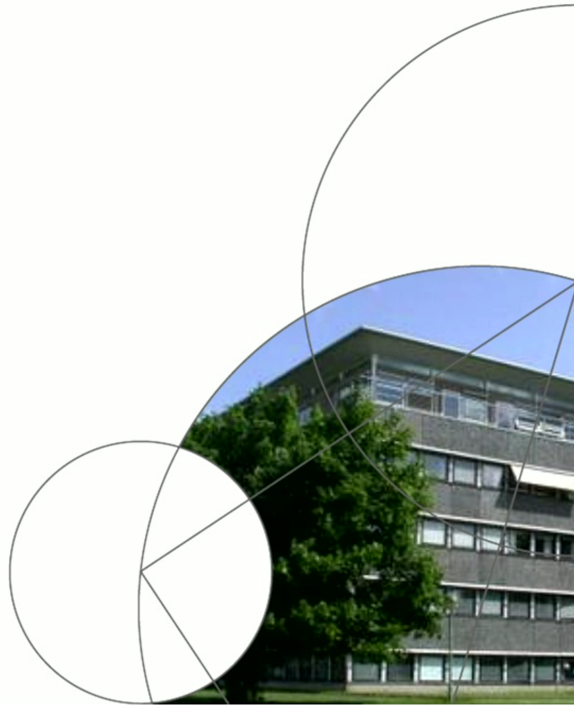
URL: https://pirsa.org/23040104

Abstract: The lecture will give a brief introduction to graphical models, their origins in Physics, Genetics, and Econometrics, their modern usages, and some future perspectives.

Department of Mathematical Sciences

Graphical Models

Steffen Lauritzen



Perimeter Institute, Waterloo, Ontario, April 2023

1. Precursors of graphical models
2. Modern examples
3. Important milestones
4. Recent developments

Lenz (1920); Ising (1925)

884 REVIEWS OF MODERN PHYSICS · OCTOBER 1967

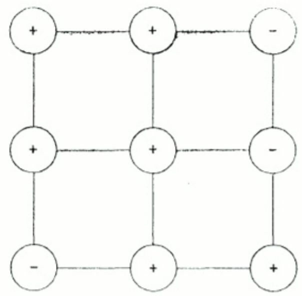


FIG. 1. A possible configuration of a finite square lattice. The energy of this configuration is $E=-2U+3\mu H$.

Lenz (1920); Ising (1925)

884 REVIEWS OF MODERN PHYSICS · OCTOBER 1967

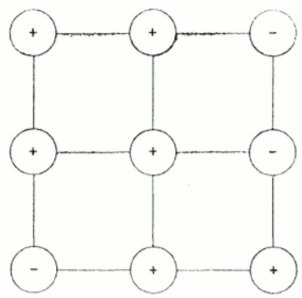


FIG. 1. A possible configuration of a finite square lattice. The energy of this configuration is $E=-2U+3\mu H$.

Energy $E(\sigma) = \sum_{i\sim j} U\sigma_i\sigma_j + \mu H \sum_i \sigma_i$.

Lenz (1920); Ising (1925)

884 Reviews of Modern Physics · October 1967

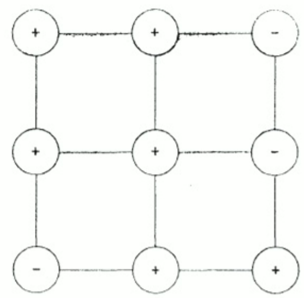


Fig. 1. A possible configuration of a finite square lattice. The energy of this configuration is $E=-2U+3\mu H$.

Energy $E(\sigma) = \sum_{i\sim j} U\sigma_i\sigma_j + \mu H \sum_i \sigma_i$.

Probability of state $p(\sigma) \propto \exp\{-E(\sigma)/kT\}$.

Wright (1921), Genetics

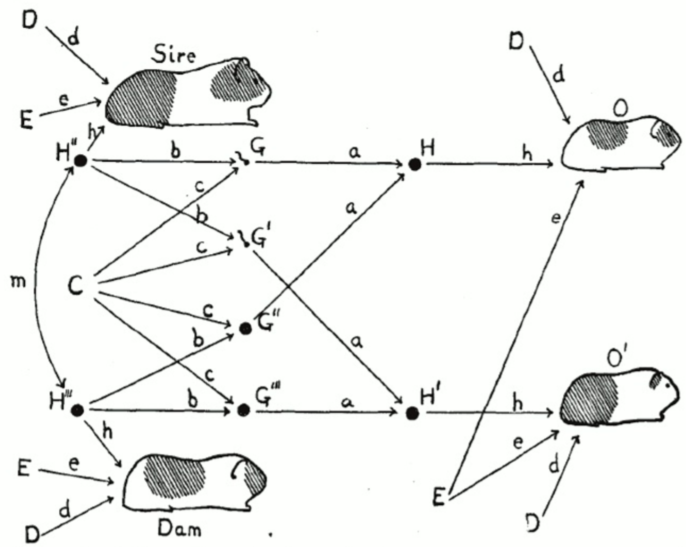


FIGURE 2.—A diagram illustrating the relations between two mated individuals and their progeny. H, H', H'' and H''' are the genetic constitutions of the four individuals. G, G', G'' and G''' are four germ-cells. E and D represent tangible external conditions and chance irregularities as factors in development. C represents chance at segregation as a factor in determining the composition of the germ-cells. Path coefficients are represented by small letters.

Wright (1934), Formal theory.

THE METHOD OF PATH COEFFICIENTS

By

SEWALL WRIGHT
Department of Zoology, The University of Chicago.

Introduction

The method of path coefficients was suggested a number of years ago (Wright 1918, more fully 1920, 1921), as a flexible means of relating the correlation coefficients between variables in a multiple system to the functional relations among them. The method has been applied in quite a variety of cases. It seems desirable now to make a restatement of the theory and to review the types of application, especially as there has been a certain amount of misunderstanding both of purpose and of procedure.

Basic Formulae

The object of investigation is a system of variable quantities, arranged in a typically branching sequential order representative of some chosen point of view toward the functional relations. Such a system is conveniently represented in a diagram such as Fig. 1. Those variables which are treated as dependent are connected with those of which they are considered functions by arrows. The system of factors back of each variable may be made formally complete by the introduction of symbols representative of total residual determination (as V_u in Fig. 1). A residual correlation between variables is represented by a double-headed arrow. It will be assumed that all relations are linear.[1] Thus each variable is related to those from which uni-

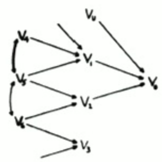


FIG. 1

[1] Relations which are far from linear with respect to the absolute values of the variables may be approximately linear with respect to variations, if the coefficients of variability are small. Thus if $V_0 = f(V_1, V_2, \cdots, V_n)$,

Goodman (1970), Log-linear models

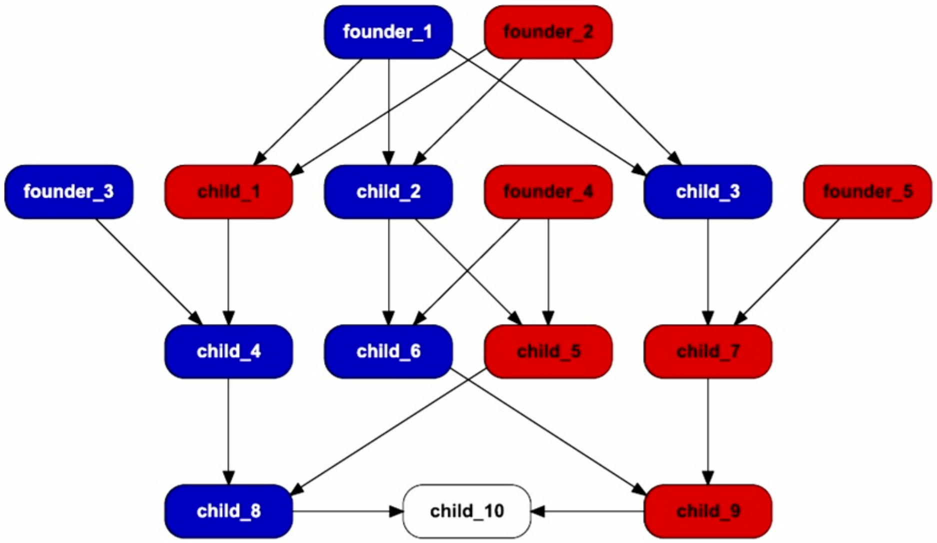
$$\log p_{ijkl} = \alpha^{\emptyset} + \alpha_i^{A} + \cdots + \alpha_{ij}^{AB} + \cdots + \alpha_{ijkl}^{ABCD}$$

Table 4. EACH HIERARCHICAL HYPOTHESIS PERTAINING TO THE 4-WAY CONTINGENCY TABLE

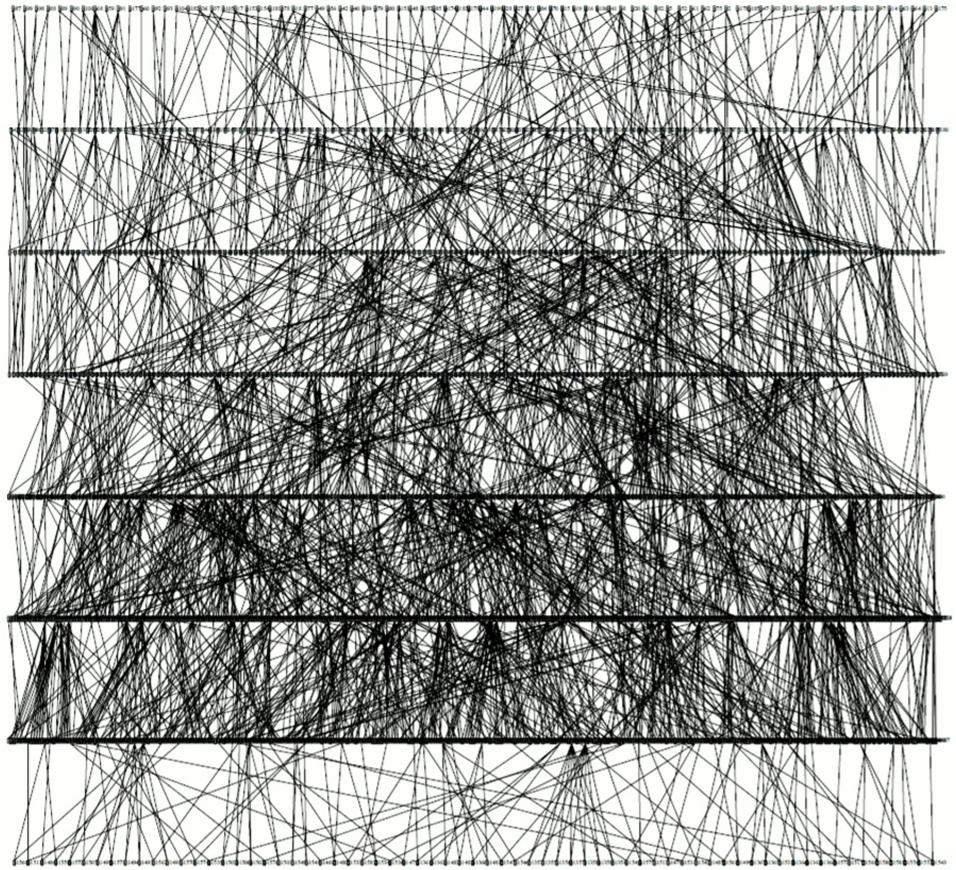
Kind of hypothesis	*Fitted marginals and description in conventional terms*		*Degrees of freedom*[a]	*Number of hypotheses of this kind*
	I. Elementary hypotheses			
1	$\{ABC\}, \{ABD\}$	$[\overline{C} \otimes \overline{D} \mid \overline{A}\overline{B}]$	4	6
2	$\{ABC\}, \{AD\}$	$[\overline{B}\overline{C} \otimes \overline{D} \mid \overline{A}]$	6	12
3	$\{ABC\}, \{D\}$	$[\overline{A}\overline{B}\overline{C} \otimes \overline{D}]$	7	4
4	$\{ABC\}$	$[\overline{D} = \Phi \mid \overline{A}\overline{B}\overline{C}]$	8	4
5	$\{AB\}, \{AC\}, \{AD\}$	$[\overline{B} \otimes \overline{C} \otimes \overline{D} \mid \overline{A}]$	8	4
6	$\{AB\}, \{AC\}, \{BD\}$	$[\overline{B} \otimes \overline{C} \mid \overline{A}] \cap [\overline{A}\overline{C} \otimes \overline{D} \mid \overline{B}]$	8	12
7	$\{AB\}, \{AC\}, \{D\}$	$[\overline{B} \otimes \overline{C} \mid \overline{A}] \cap [\overline{A}\overline{B}\overline{C} \otimes \overline{D}]$	9	12
8	$\{AB\}, \{CD\}$	$[\overline{A}\overline{B} \otimes \overline{C}\overline{D}]$	9	3
9	$\{AB\}, \{AC\}$	$[\overline{B} \otimes \overline{C} \mid \overline{A}] \cap [\overline{D} = \Phi \mid \overline{A}\overline{B}\overline{C}]$	10	12
10	$\{AB\}, \{C\}, \{D\}$	$[\overline{A}\overline{B} \otimes \overline{C} \otimes \overline{D}]$	10	6
11	$\{AB\}, \{C\}$	$[\overline{A}\overline{B} \otimes \overline{C}] \cap [\overline{D} = \Phi \mid \overline{A}\overline{B}\overline{C}]$	11	12
12	$\{A\}, \{B\}, \{C\}, \{D\}$	$[\overline{A} \otimes \overline{B} \otimes \overline{C} \otimes \overline{D}]$	11	1
13	$\{AB\}$	$[\overline{C}\overline{D} = \Phi \mid \overline{A}\overline{B}]$	12	6
14	$\{A\}, \{B\}, \{C\}$	$[\overline{A} \otimes \overline{B} \otimes \overline{C}] \cap [\overline{D} = \Phi \mid \overline{A}\overline{B}\overline{C}]$	12	4
15	$\{A\}, \{B\}$	$[\overline{A} \otimes \overline{B}] \cap [\overline{C}\overline{D} = \Phi \mid \overline{A}\overline{B}]$	13	6
16	$\{A\}$	$[\overline{B}\overline{C}\overline{D} = \Phi \mid \overline{A}]$	14	4
17	n	$[\overline{A}\overline{B}\overline{C}\overline{D} = \Phi]$	15	1
	II. Non-elementary hypotheses			
18	$\{ABC\}, \{ABD\}, \{ACD\}, \{BCD\}$		1	1
19	$\{ABC\}, \{ABD\}, \{ACD\}$		2	4
20	$\{ABC\}, \{ABD\}, \{CD\}$		3	6
21	$\{ABC\}, \{AD\}, \{BD\}, \{CD\}$		4	4
22	$\{ABC\}, \{AD\}, \{BD\}$		5	12
23	$\{AB\}, \{AC\}, \{AD\}, \{BC\}, \{BD\}, \{CD\}$		5	1
24	$\{AB\}, \{AC\}, \{AD\}, \{BC\}, \{BD\}$		6	6
25	$\{AB\}, \{AC\}, \{AD\}, \{BC\}$		7	15
26	$\{AB\}, \{AC\}, \{BC\}, \{D\}$		8	4
27	$\{AB\}, \{AC\}, \{BC\}$		9	4

[a] Degrees of freedom for the 2^4 table.

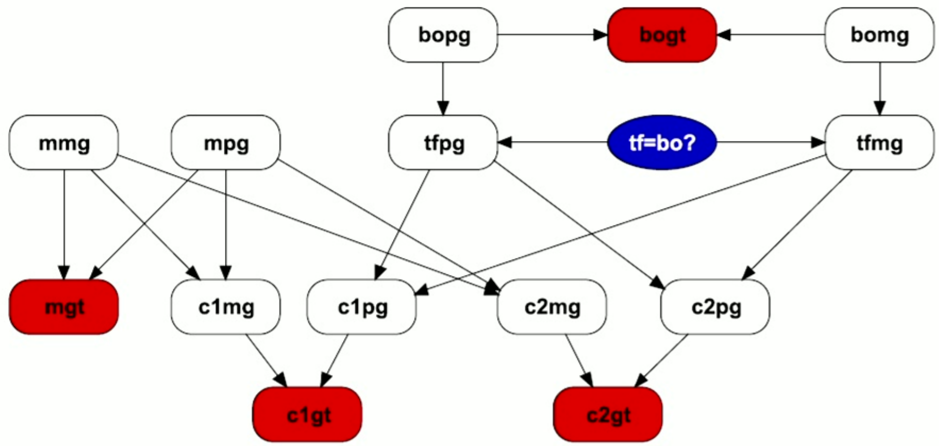
Pedigrees, Werner's syndrome



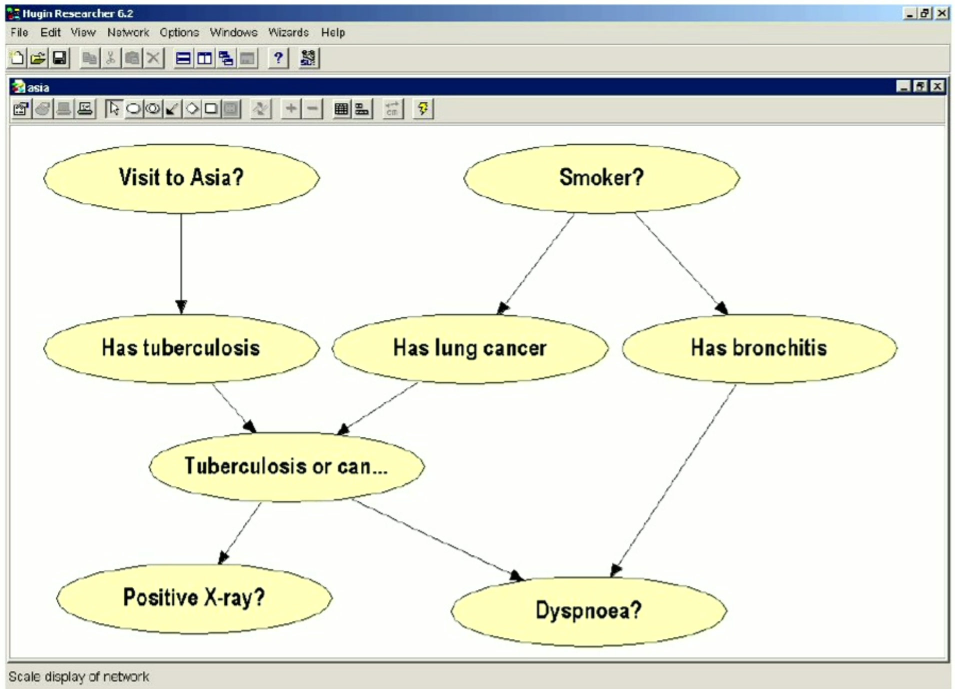
Pedigrees, Eskimo relationships



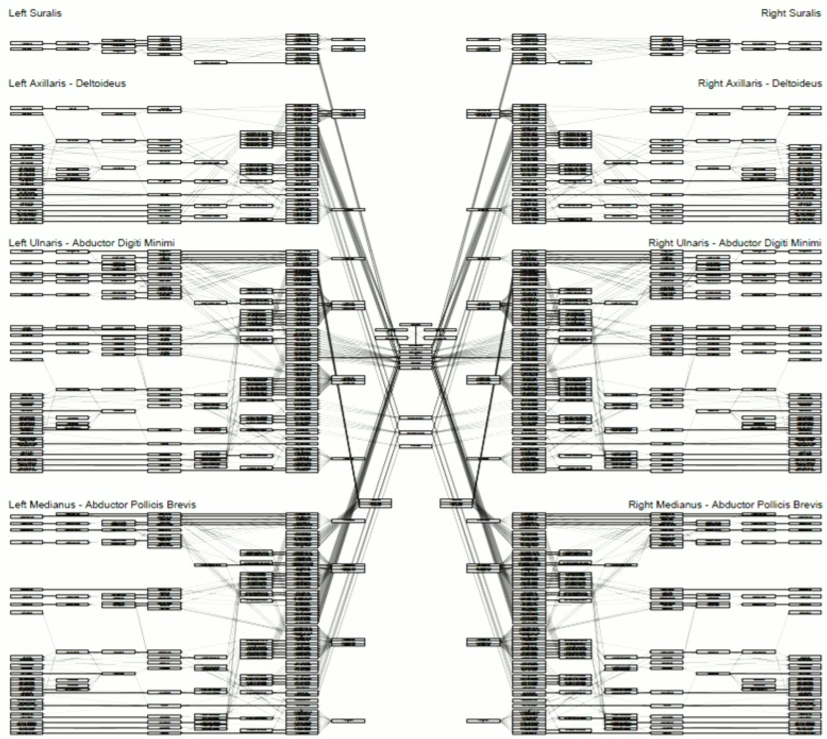
Forensic genetics, Body identification



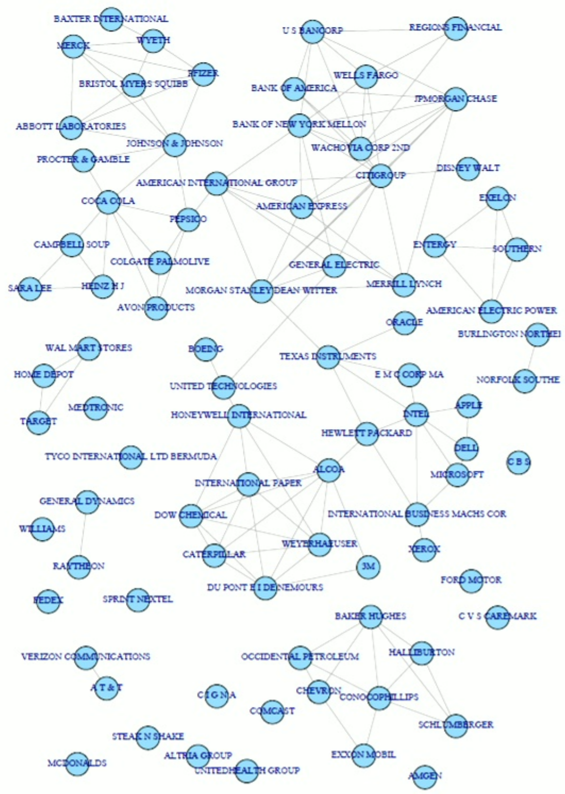
Medical diagnosis, expert systems



Electromyography, MUNIN



Financial analysis, stock prices



Other modern applications of graphical models

- digital communication (coding and decoding);

Other modern applications of graphical models

- digital communication (coding and decoding);
- insurance fraud;
- credit evaluation;
- intelligent cars;
- space technology;
- games (both game theory and actual games);
- image analysis;
- quantum systems?
- etc. etc.

Important milestones: Phase I

- Visit to Terry Speed, Perth, Australia 1976, *basic paper* Darroch et al. (1980);
- Swedish summerschool, Särö 1979: Lecture notes, met Nanny Wermuth, *extension to mixed discrete continuous variables and chain graphs* (Lauritzen and Wermuth, 1989);
- Artificial intelligence and expert systems; solves *computational problem for sparse directed acyclic graphs* (Lauritzen and Spiegelhalter, 1988; Pearl, 1988);
- *Markov chain Monte Carlo methods* and complex Bayesian graphical models (Thomas et al., 1992),exploiting methods used in physics.

Important milestones: Phase II

- *Abstract theory* of conditional independence: semigraphoids etc. (Pearl, 1988; Studený, 1992; Studený, 2005);
- *Causal interpretation* of directed graphical models and search methods (Spirtes et al., 1993; Pearl, 1995);
- *Plethora of graph types* for describing conditional independence (Cox and Wermuth, 1993; Andersson et al., 2001; Richardson and Spirtes, 2002).
- *Variational inference* in graphical models (mean field method) (Jordan et al., 1999); again transferred from physics;
- General methods for *structure identification* (Edwards and Havránek, 1985; Dawid and Lauritzen, 1993; Friedman et al., 2008).

Conditional Independence

Random variables X and Y are *conditionally independent* given the random variable Z if

$$\mathcal{L}(X \mid Y, Z) = \mathcal{L}(X \mid Z).$$

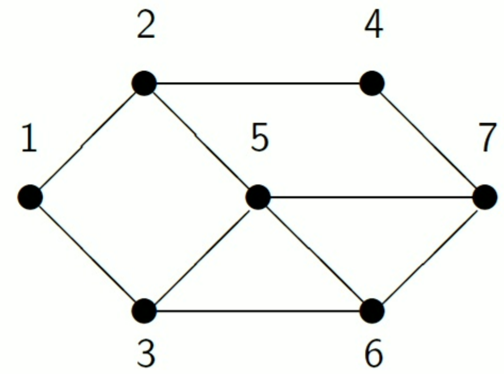
We then write $X \perp\!\!\!\perp Y \mid Z$ (or $X \perp\!\!\!\perp_P Y \mid Z$)

Intuitively: Knowing Z renders Y *irrelevant* for predicting X.

Factorisation of densities:

$$\begin{aligned} X \perp\!\!\!\perp Y \mid Z \quad & \iff \quad f(x,y,z)f(z) = f(x,z)f(y,z) \\ & \iff \quad \exists a, b : f(x,y,z) = a(x,z)b(y,z). \end{aligned}$$

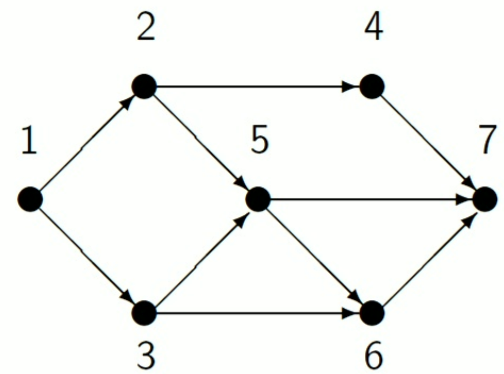
Undirected graphical models



Complex systems of conditional independence can for example be described by undirected graphs.

Directed graphical models

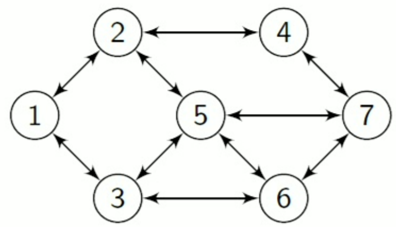
Directed graphs are also natural models for conditional indpendence:



Any node is *conditional independent of its non-descendants, given its immediate parents.* So, for example, in the above picture we have

$$5 \perp\!\!\!\perp \{1,4\} \mid \{2,3\}, \quad 6 \perp\!\!\!\perp \{1,2,4\} \mid \{3,5\}.$$

Markov theory: bidirected graphs



Here the global Markov property entails that $1 \perp\!\!\!\perp 4, 7 \mid 3$ because all paths from 1 to $\{4, 7\}$ pass through *the complement* of these nodes.

This would be true for a Gaussian distribution if

$$i \leftrightarrow j \iff \sigma_{ij} \neq 0,$$

where σ_{ij} is the covariance between X_i and X_j.

For random variables X, Y, Z, and W it holds

(C1) If $X \perp\!\!\!\perp Y \mid Z$ then $Y \perp\!\!\!\perp X \mid Z$;

(C2) If $X \perp\!\!\!\perp Y \mid Z$ and $U = g(Y)$, then $X \perp\!\!\!\perp U \mid Z$;

(C3) If $X \perp\!\!\!\perp Y \mid Z$ and $U = g(Y)$, then $X \perp\!\!\!\perp Y \mid (Z, U)$;

(C4) If $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp W \mid (Y, Z)$, then $X \perp\!\!\!\perp (Y, W) \mid Z$;

If density w.r.t. product measure $f(x, y, z, w) > 0$ also

(C5) If $X \perp\!\!\!\perp Y \mid (Z, W)$ and $X \perp\!\!\!\perp Z \mid (Y, W)$ then $X \perp\!\!\!\perp (Y, Z) \mid W$.

An *independence model* $\perp_\sigma$ is a ternary relation over subsets of a finite set V. *Semi-graphoid* if for all subsets A, B, C, D:

(S1) if $A \perp_\sigma B \mid C$ then $B \perp_\sigma A \mid C$ (symmetry);

(S2) if $A \perp_\sigma (B \cup D) \mid C$ then $A \perp_\sigma B \mid C$ and $A \perp_\sigma D \mid C$ (decomposition);

(S3) if $A \perp_\sigma (B \cup D) \mid C$ then $A \perp_\sigma B \mid (C \cup D)$ (weak union);

(S4) if $A \perp_\sigma B \mid C$ and $A \perp_\sigma D \mid (B \cup C)$, then $A \perp_\sigma (B \cup D) \mid C$ (contraction).

It is a *graphoid* if (S1)–(S4) holds and

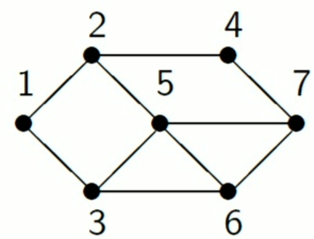
(S5) if $A \perp_\sigma B \mid (C \cup D)$ and $A \perp_\sigma C \mid (B \cup D)$ then $A \perp_\sigma (B \cup C) \mid D$ (intersection).

It is *compositional* if also

(S6) if $A \perp_\sigma B \mid C$ and $A \perp_\sigma D \mid C$ then $A \perp_\sigma (B \cup D) \mid C$ (composition).

Separation in undirected graphs

Let $\mathcal{G} = (V, E)$ be finite and simple undirected graph (no self-loops, no multiple edges).



For $A, B, S \subseteq V$, let $A \perp_{\mathcal{G}} B \mid S$ denote that *S separates A from B in $\mathcal{G}$*, i.e. all paths from A to B intersect S.

Fact: *The relation $\perp_{\mathcal{G}}$ on subsets of V is a compositional graphoid.*

Systems of random variables

For a system V of *labeled random variables* $X_v, v \in V$, we use the shorthand

$$A \perp\!\!\!\perp B \mid C \iff X_A \perp\!\!\!\perp X_B \mid X_C,$$

where $X_A = (X_v, v \in A)$.

Systems of random variables

For a system V of *labeled random variables* $X_v, v \in V$, we use the shorthand

$$A \perp\!\!\!\perp B \mid C \iff X_A \perp\!\!\!\perp X_B \mid X_C,$$

where $X_A = (X_v, v \in A)$.

The properties (C1)–(C4) imply that $\perp\!\!\!\perp$ *satisfies the semi-graphoid axioms* for such a system, and the *graphoid axioms* if the joint density of the variables is strictly positive.

An independence model of this kind is said to be *probabilistic*.

Systems of random variables

For a system V of *labeled random variables* $X_v, v \in V$, we use the shorthand

$$A \perp\!\!\!\perp B \mid C \iff X_A \perp\!\!\!\perp X_B \mid X_C,$$

where $X_A = (X_v, v \in A)$.

The properties (C1)–(C4) imply that $\perp\!\!\!\perp$ *satisfies the semi-graphoid axioms* for such a system, and the *graphoid axioms* if the joint density of the variables is strictly positive.

An independence model of this kind is said to be *probabilistic*.

Markov theory is concerned with relations between a probabilistic independence model and independence models given by graphs.

General Markov theory

A node u is a *collider* on a walk if arrowheads meet at u

$$(\ldots \to u \leftarrow \ldots), (\ldots \leftrightarrow u \leftarrow \ldots), (\ldots \to u \leftrightarrow \ldots), (\ldots \leftrightarrow u \leftrightarrow \ldots)$$

A walk is *connecting relative to a set S* if all non-colliders are outside S and all colliders are inside S.

For a graph $\mathcal{G}$ (with three types of edges) we say that *A is $\mathcal{G}$-separated from B by S* if there is no walk from A to B that connects relative to S and we write $A \perp_{\mathcal{G}} B \mid S$.

A probability distribution P is *Markov* w.r.t. $\mathcal{G}$ if

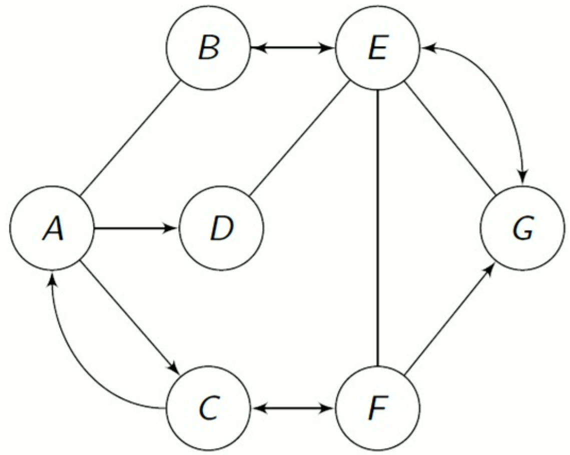
$$A \perp_{\mathcal{G}} B \mid S \implies A \perp\!\!\!\perp_P B \mid S$$

A probability distribution P is *faithful* to $\mathcal{G}$ if

$$A \perp_{\mathcal{G}} B \mid S \iff A \perp\!\!\!\perp_P B \mid S$$

Markov theory: general graphs

Challenge: which general graphs can represent conditional independence properties of probability distributions?



In other words, *does there exist a P that is faithful to $\mathcal{G}$?*

Max-linear Bayesian networks for extremes

Structural equation models of the form

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

where $\mathrm{pa}(i)$ denotes parents of i in a directed acyclic graph; studied by Gissibl and Klüppelberg (2018), Gissibl et al. (2021),

Max-linear Bayesian networks for extremes

Structural equation models of the form

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

where pa(i) denotes parents of i in a directed acyclic graph; studied by Gissibl and Klüppelberg (2018), Gissibl et al. (2021),

Is a *max-linear model* (Wang and Stoev, 2011):

$$X = B \odot Z$$

where the matrix product $\odot$ is given as

$$(M \odot N)_{ij} = \bigvee_k m_{ik} n_{kj}$$

and $B = C^{\odot d} = \lim_{n \to \infty} C^{\odot n} = B \odot B$.

Max-linear Bayesian networks for extremes

Structural equation models of the form

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

where pa(i) denotes parents of i in a directed acyclic graph; studied by Gissibl and Klüppelberg (2018), Gissibl et al. (2021),

Is a *max-linear model* (Wang and Stoev, 2011):

$$X = B \odot Z$$

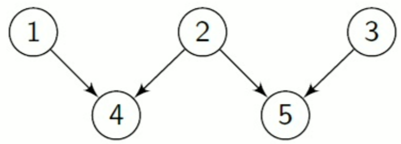
where the matrix product $\odot$ is given as

$$(M \odot N)_{ij} = \bigvee_k m_{ik} n_{kj}$$

and $B = C^{\odot d} = \lim_{n\to\infty} C^{\odot n} = B \odot B$.

To study Markov properties, use *tropical linear algebra* (Amendola et al., 2022).

Special Markov properties appear

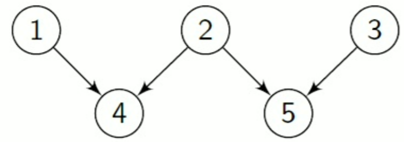


Here it holds for any MLBN that $1 \perp\!\!\!\perp 3 \mid \{4,5\}$ whereas $\neg(1 \perp_{\mathcal{D}} 3 \mid \{4,5\})$!

Assume all coefficients equal to 1. Then the conditional distribution of (X_1, X_2, X_3) given $(X_4, X_5) = (x_4, x_5)$ is determined by imposing the following inequalities on (X_1, X_2, X_3):

$$\max(X_1, X_2) \leq x_4, \quad \max(X_2, X_3) \leq x_5.$$

Example continued



$$\max(X_1, X_2) \leq x_4, \quad \max(X_2, X_3) \leq x_5.$$

Now distinguish three cases:

1. If $x_4 < x_5$, the condition is equivalent to

$$\max(X_1, X_2) \leq x_4, \quad X_3 \leq x_5.$$

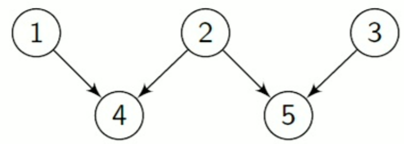
2. If $x_4 > x_5$, the condition is equivalent to

$$\max(X_1) \leq x_4, \quad \max(X_2, X_3) \leq x_5.$$

3. If $x_4 = x_5$ we must have $X_2 = x_4 = x_5$ and hence equiv

$$X_1 \leq x_4, \quad X_3 \leq x_5, \quad X_2 = x_4 = x_5.$$

Example continued



In all three (contexts) we have independence of X_1 and X_3

1. If $x_4 < x_5$, *x_5 cannot be caused by X_2 but only by X_3 or Z_5.*
2. If $x_4 > x_5$, *x_4 cannot be caused by X_2 but only by X_1 or Z_4.*
3. If $x_4 = x_5$ *both must be caused by X_2.*

Spirtes, P., Glymour, C., and Scheines, R. (1993). *Causation, Prediction and Search*. Springer-Verlag, New York. Lecture Notes in Statistics 81.

Studený, M. (1992). Conditional independence relations have no finite complete characterization. In *Transactions of the 11th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*, pages 377–396, Prague. Academia.

Studený, M. (2005). *Probabilistic Conditional Independence Structures*. Information Science and Statistics. Springer-Verlag, London.

Thomas, A., Spiegelhalter, D. J., and Gilks, W. R. (1992). BUGS: A program to perform Bayesian inference using Gibbs Sampling. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 4*, pages 837–842, Oxford, United Kingdom. Clarendon Press.

Wang, Y. and Stoev, S. A. (2011). Conditional sampling for spectrally discrete max-stable random fields. *Advances in Applied Probability*, 43(2):461–483.

Wright, S. (1921). Systems of mating. I. The biometric relations between parent and offspring. *Genetics*, 6:111–123.

Wright, S. (1934). The method of path coefficients. *The Annals of Mathematical Statistics*, 5:161–215.