Bayesian Networks

Prof. Giovanni Briganti
Unit of Computational Medicine and Neuropsychiatry
Chair of AI & Digital Medicine
Faculty of Medicine, University of Mons
Giovanni.BRIGANTI@umons.ac.be

LAKE COMO SCHOOL OF ADVANCED STUDIES

Computational Medicine and Neuropsychiatry

UMONS University of Mons

A small recap of what we did so far…

INTERNATIONAL JOURNAL OF METHODS IN PSYCHIATRIC RESEARCH Open Access

LONG DIDACTIC PAPER | Open Access

Network analysis: An overview for mental health research

This article relates to:

Giovanni Briganti, Marco Scutari, Sacha Epskamp, Denny Borsboom, Ria H. A. Hoekstra, Hudson Fernandes Golino, Alexander P. Christensen, Yannick Morvan, Omid V. Ebrahimi, Giulio Costantini, Alexandre Heeren, Jill de Ron, Laura F. Bringmann, Karoline Huth, Jonas M. B. Haslbeck, Adela-Maria Isvoranu, Maarten Marsman, Tessa Blanken, Allison Gilbert, Teague Rhine Henry, Eiko I. Fried, Richard J. McNally ... See fewer authors

First published: 14 November 2024 | https://doi.org/10.1002/mpr.2034 | Cit

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Methods

ISSN: 1082-989X

https://doi.org/10.1037/met0000479

A Tutorial on Bayesian Networks for Psychopathology Researchers

Giovanni Briganti[1, 2], Marco Scutari[3], and Richard J. McNally[1]

[1] Department of Psychology, Harvard University

[2] Laboratoire de Psychologie Médicale et Addictologie, Université libre de Bruxelles

[3] Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA), Lugano, Switzerland

Abstract

Bayesian Networks are probabilistic graphical models that represent conditional independence relationships among variables as a directed acyclic graph (DAG), where edges can be interpreted as causal effects connecting one causal symptom to an effect symptom. These models can help overcome one of the key limitations of partial correlation networks whose edges are undirected. This tutorial aims to introduce Bayesian Networks to identify admissible causal relationships in cross-sectional data, as well as how to estimate these models in R through three algorithm families with an empirical example data set of depressive symptoms. In addition, we discuss common problems and questions related to Bayesian networks. We recommend Bayesian networks be investigated to gain causal insight in psychological data.

Why think of mental disorders as networks?

- Step away from models that poorly fit clinical reality
- Solve the problem of heterogeneity
- Intervene on symptoms ?

Where do networks come from?

- Graph theory < AI
- Graph: relationships among entities
- Study of entities AND their connections
  - Between train stations
  - Between people (social networks)
  - Between brain regions (neuroanatomical networks)
  - **Between symptoms (psych networks)**

Example input

- Psychometric scales
- Individual Symptoms
- DSM Criteria
- Biological and physical observations

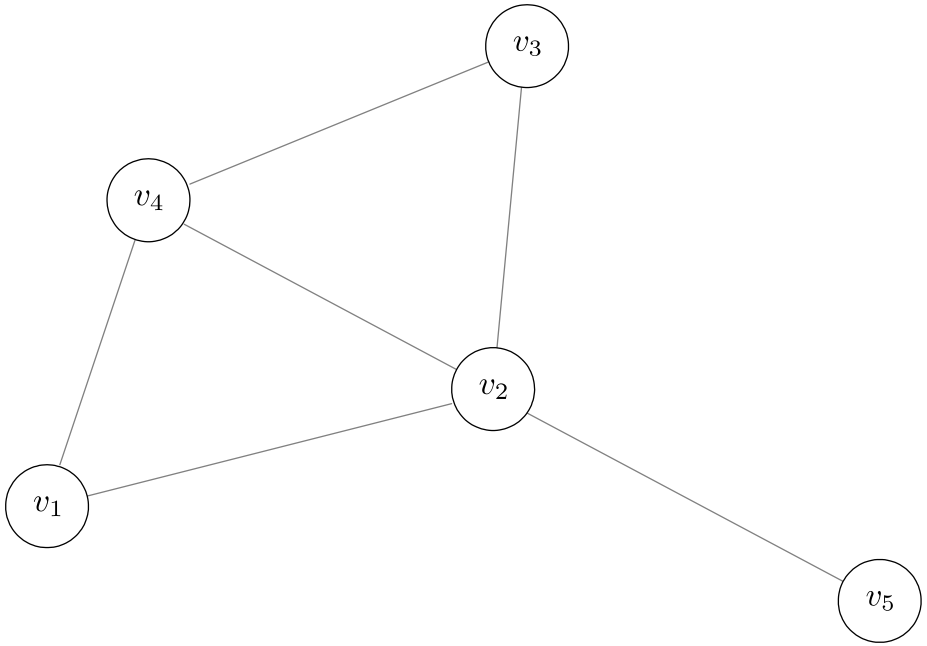
- … if what you are doing makes sense !

Uncertain Reasoning and Bayesian AI

- Reason in probabilistical terms when dealing with uncertainty
- The disease is unknown or partially known
- Tool: Bayesian AI
- Variables: real entities instead of a reasoning process
- In the case of psychiatry, for instance symptoms, signs, brain regions, biological items

Graph Theory

- Graph composed of
  - Nodes: symptoms
  - Edges: connections between 2 nodes
    - Weighted or Unweighted
    - Directed or Undirected
- Network
  - Graph
  - Probability distribution

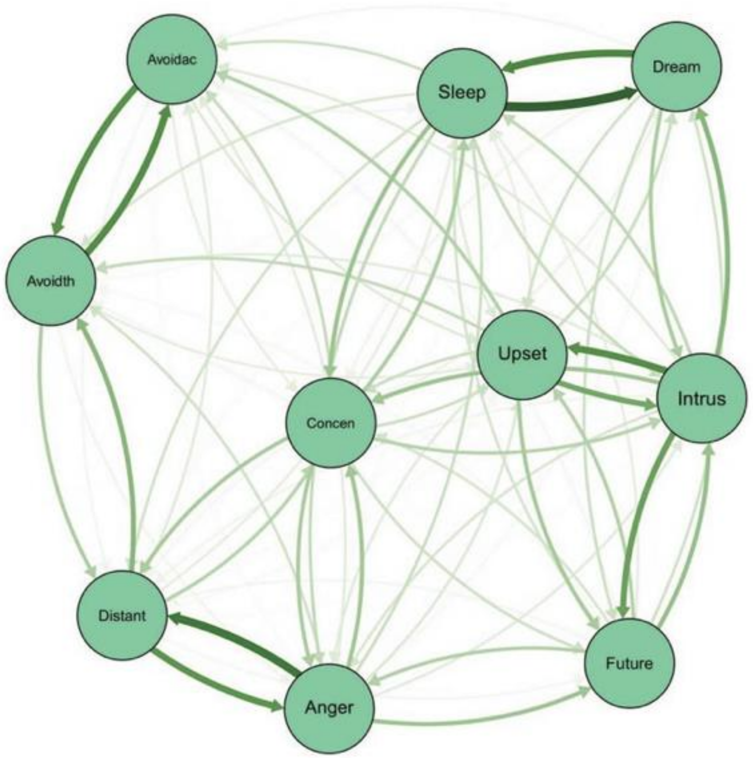


Briganti, G., Scutari, M., & McNally, R. J. (2022). A tutorial on bayesian networks for psychopathology researchers. *Psychological methods*.

Pairwise Markov Random Fields

- Components mutually influence each other
- Cycles and loops are allowed
- Example sleep problems → agitation → sleep problems → mania → sleep problems
- Estimating unobserved edges
  - Continuous. → GGM (partial correlation)
  - Ordinal → several methods exist
  - Binary → Ising model
  - Mixed → mgm
- Different frameworks !
  - “Frequentist” approach
  - ”Bayesian” approach

Relative importance networks beware the interpretation



Network inference

- Estimating a network structure is the first step
- But how to interpret a network?
- Which symptoms are the most interconnected?
    - Centrality / R^2
- Are networks of two groups different?
    - NCT etc.
    - Do symptoms aggregate in communities?
    - Community detection
- Are the results stable and accurate?
    - Bootstrapping (to a certain extent)

The need for Bayesian AI

- What to do with a central symptom
- What to do with a community
- Cause? Effect?
- We need to move things one step further
- Goal: provide an analysis of which are the admissible or plausible causal effects in our cross-sectional data set
- Additional steps (and sets of algorithms) are needed: these are provided by the Bayesian AI field

Bayesian Networks

Bayesian Networks

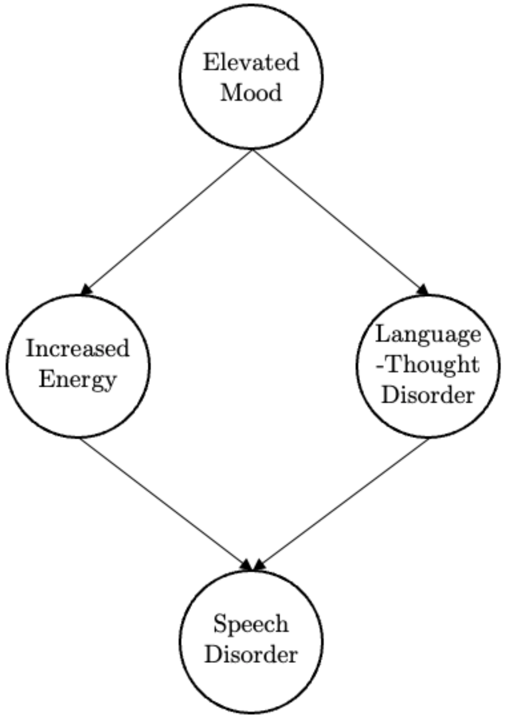
- Directed acyclic graph (DAG) + probability distribution
- Connections are directed
- No cycles
- No loops

- Tools for Causal Inference !

Conditions for rigorous causal inference using BNs

- Causal sufficiency: all the causes of a given variables are measured
  - No latent variables
  - No selection bias
- Causal Faithfulness: all the variables that are causally connected are probabilistically dependent
- Difficult to verify in psych data, but the same problem applies to other models in a similar way !

Directed Acyclic Graphs



Markov Blanket

- How to start?
- We look for nodes that are separated, given all combinations of all other nodes
- This is done algorithmically to determine a set S that separates vi and vj and for all possible node combination
- D-separation: two nodes X and Y are d-separated if conditioning on all members of S blocks all paths from X to Y
- The Markov blanket of a node X in a directed acyclic graph (DAG) is the minimal set of nodes that renders X conditionally independent from all other nodes in the graph.

Markov Blanket

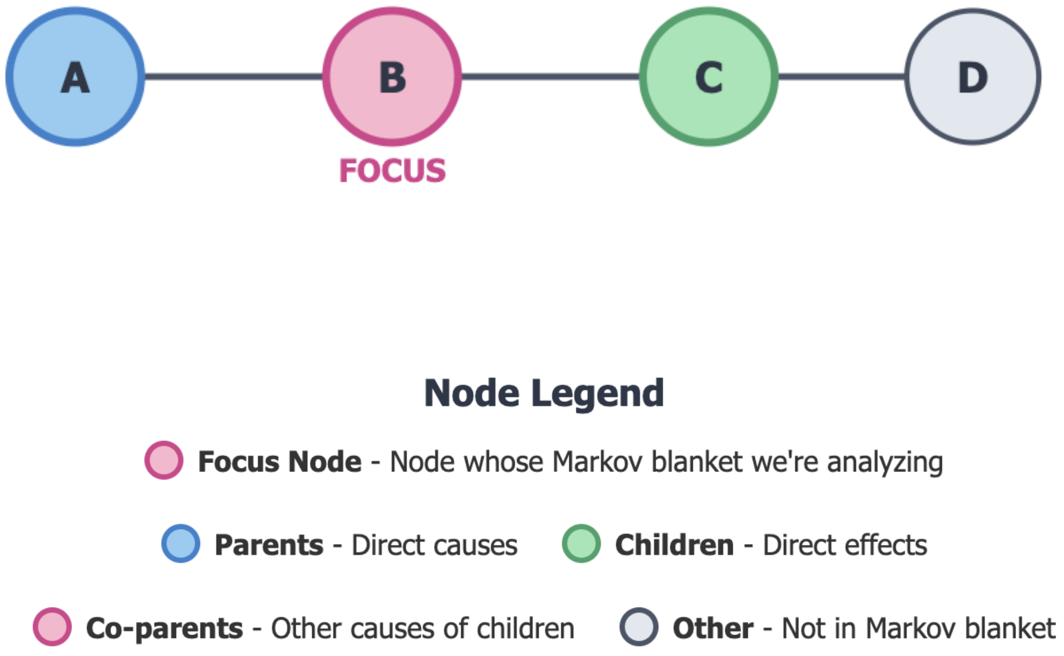
- For a node X in a DAG, the Markov blanket MB(X) consists of:

1. **Parents of X**: Pa(X) = {Y : Y → X}
2. **Children of X**: Ch(X) = {Y : X → Y}
3. **Co-parents (spouses) of X**: Sp(X) = {Y : Y → Z ← X for some Z ∈ Ch(X)}

- **Minimality :** the Markov blanket is the **minimal** set with the conditional independence property. No proper subset of MB(X) renders X independent from all other nodes.
- **Symmetry Property :** if Y ∈ MB(X), then X ∈ MB(Y). This means Markov blanket relationships are symmetric.
- **Local Markov Property :** a node X is conditionally independent of all non-descendants given its parents

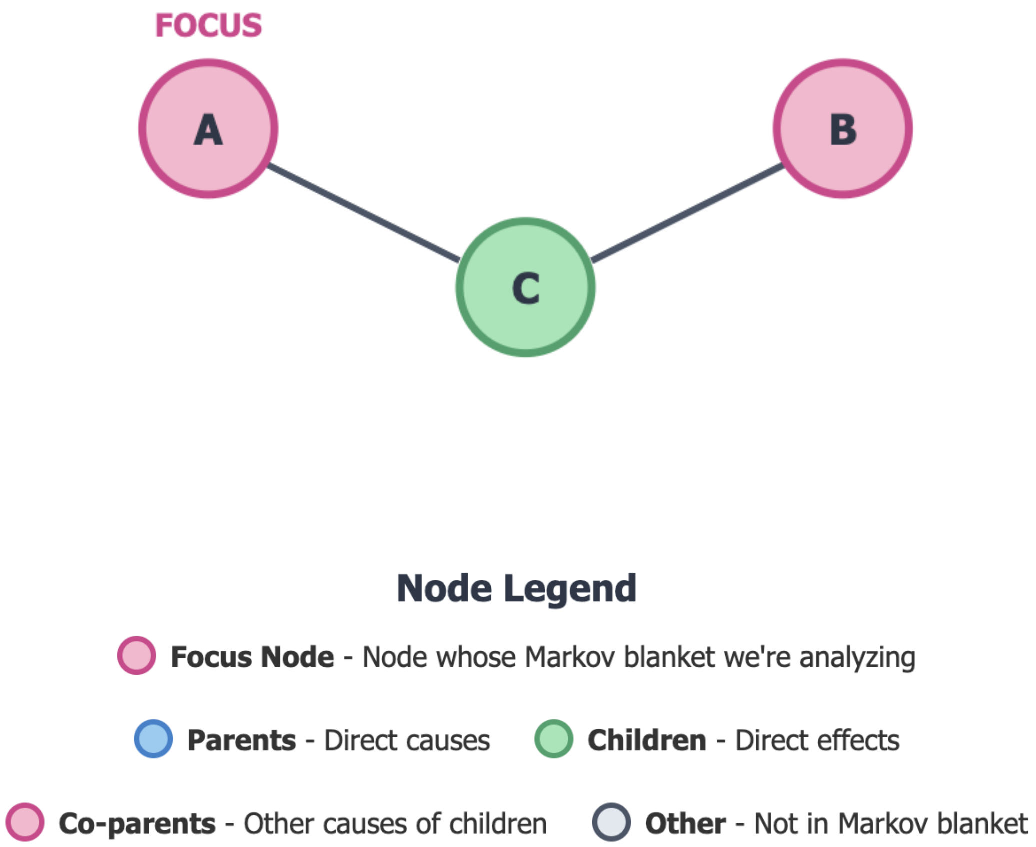
Example 1: Simple Chain

- **Graph**: $A \to B \to C$
- **MB(A)** = {B} (only child)
- **MB(B)** = {A, C} (parent and child)
- **MB(C)** = {B} (only parent)
- **Verification**: Given B, we have $A \perp\!\!\!\perp C \mid B$ ✓



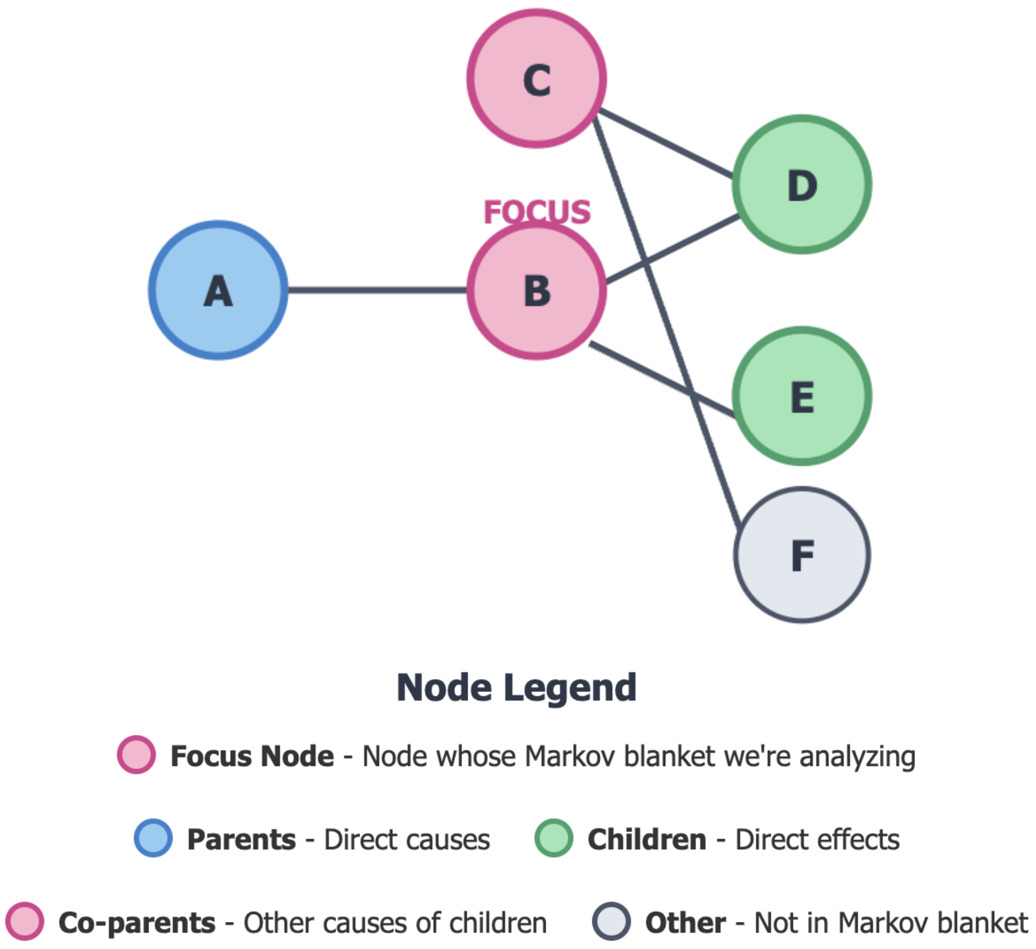
Example 2: V-Structure (Collider)

- **Graph**: $A \rightarrow C \leftarrow B$
- **MB(A)** = {B, C} (child C and co-parent B)
- **MB(B)** = {A, C} (child C and co-parent A)
- **MB(C)** = {A, B} (both parents)
- **Key insight**: A and B are co-parents, so each is in the other's Markov blanket.



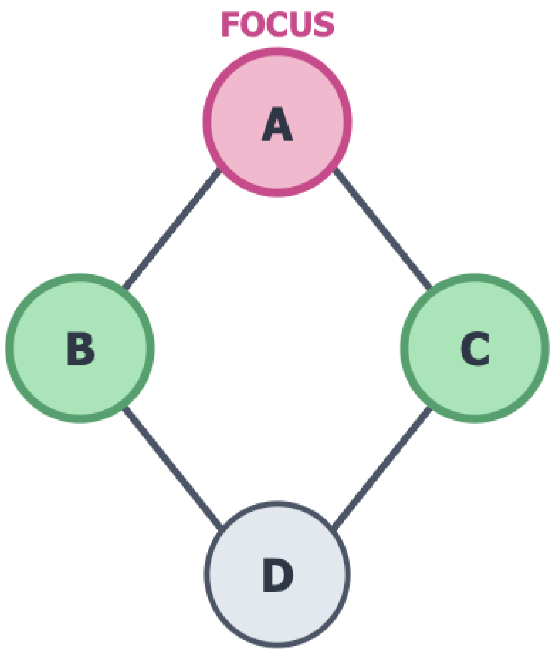
Example 3: complex structure

- **Graph**: A → B → D ← C, B → E, C → F
- **MB(A)** = {B}
- **MB(B)** = {A, C, D, E} (parent A, co-parent C, children D and E)
- **MB(C)** = {B, D, F} (co-parent B, children D and F)
- **MB(D)** = {B, C} (both parents)
- **MB(E)** = {B} (only parent)
- **MB(F)** = {C} (only parent)



Example 4: diamond structure

- MB(A) = {B,C}
- MB(B)= {A,C,D}
- MB(D)= {B,C}



Conditional Independence and the Need for D-Separation

- D-separation allows inference of conditional independencies from the DAG structure.
- Avoids relying on full probabilistic computations.
- Clarifies statistical vs. causal dependencies.
- Provides foundation for structure learning and causal inference.

Definition of D-Separation

- A path is blocked by a conditioning set Z if:
- There is a non-collider W on the path such that W ∈ Z.
- There is a collider W such that neither W nor its descendants are in Z.
- If all paths between X and Y are blocked by Z, then X and Y are d-separated given Z: $X \perp\!\!\!\perp_{\mathrm{n}} Y \mid Z$.

Implications of D-Separation

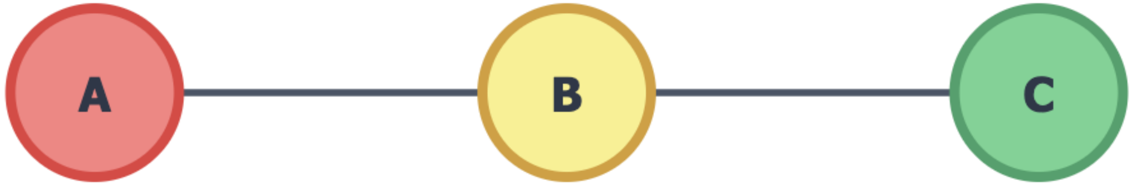
- D-Separation in the graph implies conditional independence in the distribution.
- Soundness: d-separation ⇒ statistical independence in any compatible P.
- Completeness (under faithfulness): independence in P ⇒ d-separation in G.
- Used in causal discovery (e.g., PC algorithm), do-calculus, and probabilistic inference.

Example 1: Chain Structure

✓ d-SEPARATED

A ⊥⊥ C | B

`A → B → C`



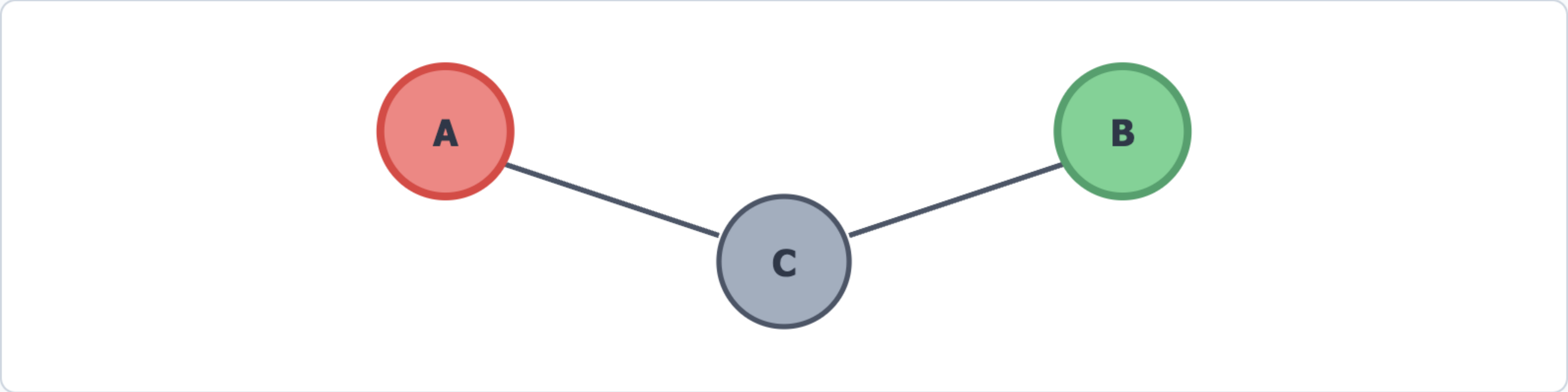
Explanation: *B is a non-collider on path A→B→C. Conditioning on B blocks this path, making A and C conditionally independent.*

Example 2: Collider Structure

✓ d-SEPARATED

$A \perp\!\!\!\perp B \mid \varnothing$

`A → C ← B`



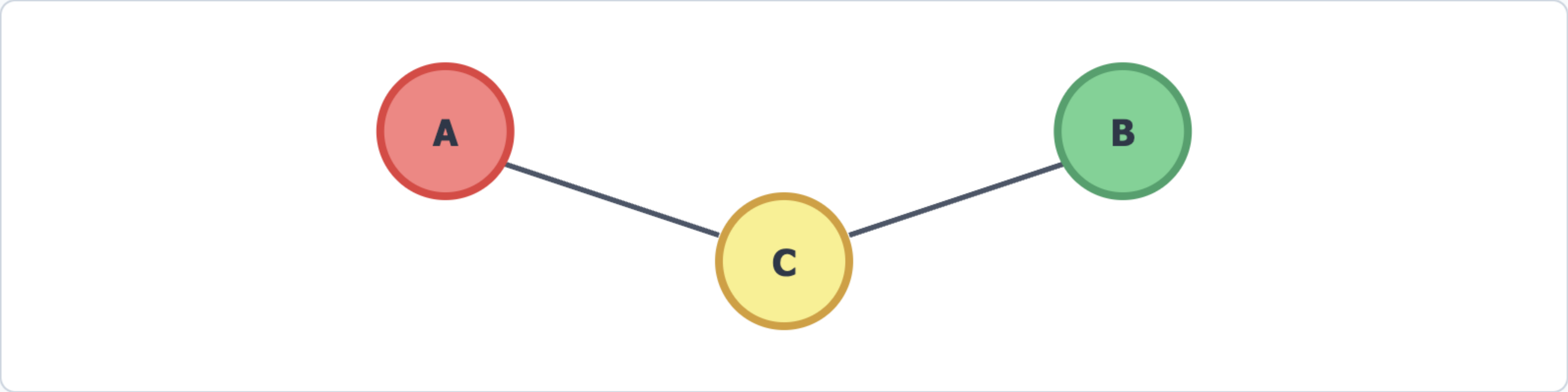
Explanation: *C is a collider on path A→C←B. Without conditioning on C (or its descendants), the path is naturally blocked, making A and B independent.*

Example 3: Conditioning on Collider

✗ NOT d-separated

A ⊥⊥ B | C

`A → C ← B`



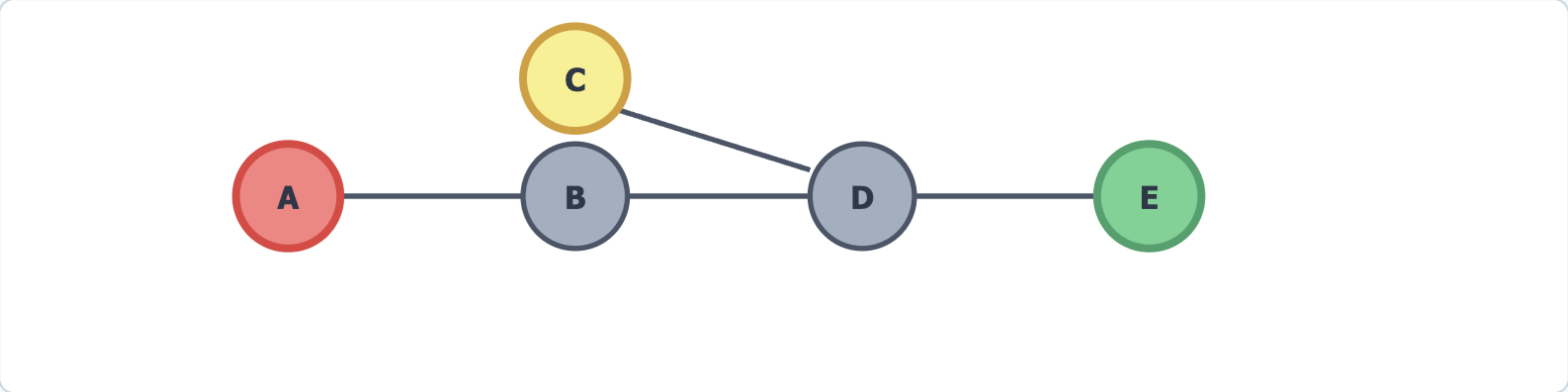
Explanation: *C is a collider on path A→C←B. Conditioning on the collider C "opens" the path, creating dependence between A and B. This demonstrates Berkson's paradox.*

Example 4: Complex Graph

✗ NOT d-separated

A ⊥⊥ E | C

`A → B → D ← C, D → E`



Explanation: *Path A→B→D→E exists. D is a collider on path A→B→D←C, but conditioning on C opens this collider. The path A→B→D→E is not blocked, so A and E are not d-separated given C.*

Markov Property

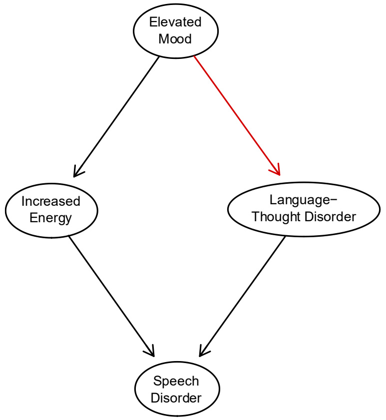
$$v_i \perp\!\!\!\perp_G v_j | v_k \Rightarrow v_i \perp\!\!\!\perp_P v_j | v_k$$

- If two nodes are not connected by an edge, then they are conditionally independent given some set S
- This is called the Markov Property
- Which makes it possible to write

$$\Pr(\mathbf{X}, \mathbf{\Theta}) = \prod_{i=1}^{N} \Pr\left(X_i \mid \Pi_{X_i}; \Theta_{X_i}\right),$$

- Where π are the parents of a node Xi

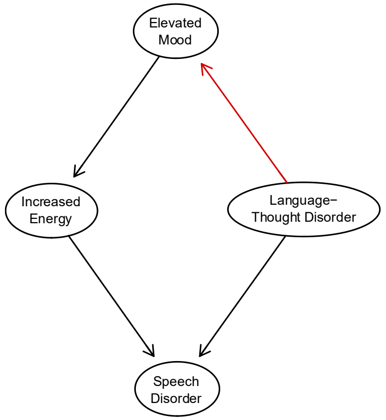
Colliders



$$\left.\begin{aligned}&\Pr(\text{Elevated Mood}, \text{Increased Energy},\\ &\quad\text{Language-Thought Disorder}, \text{Speech Disorder}) =\end{aligned}\right\}\ \text{Joint distribution}$$

$$\left.\begin{aligned}&\Pr(\text{Elevated Mood})\\ &\Pr(\text{Increased Energy} \mid \text{Elevated Mood})\\ &\Pr(\text{Language-Thought Disorder} \mid \text{Increased Energy},\\ &\quad\text{Elevated Mood})\\ &\Pr(\text{Speech Disorder} \mid \text{Language-Thought Disorder},\\ &\quad\text{Increased Energy}, \text{Elevated Mood}) =\end{aligned}\right\}\ \text{Chain rule}$$

$$\left.\begin{aligned}&\Pr(\text{Elevated Mood})\\ &\Pr(\text{Increased Energy} \mid \text{Elevated Mood})\\ &\Pr(\text{Language-Thought Disorder} \mid \text{Elevated Mood})\\ &\Pr(\text{Speech Disorder} \mid \text{Language-Thought Disorder},\\ &\quad\text{Increased Energy}) =\end{aligned}\right\}\ \text{Bayesian network, local distributions}$$



$$\left.\begin{aligned}&\cancel{\Pr(\text{Elevated Mood})}\\ &\Pr(\text{Increased Energy} \mid \text{Elevated Mood})\\ &\frac{\Pr(\text{Language-Thought Disorder}, \text{Elevated Mood})}{\cancel{\Pr(\text{Elevated Mood})}}\\ &\Pr(\text{Speech Disorder} \mid \text{Language-Thought Disorder},\\ &\quad\text{Increased Energy}) =\\ &\Pr(\text{Increased Energy} \mid \text{Elevated Mood})\\ &\Pr(\text{Language-Thought Disorder}, \text{Elevated Mood})\\ &\Pr(\text{Speech Disorder} \mid \text{Language-Thought Disorder},\\ &\quad\text{Increased Energy}) =\end{aligned}\right\}\ \text{Bayes theorem}$$

$$\left.\begin{aligned}&\Pr(\text{Increased Energy} \mid \text{Elevated Mood})\\ &\Pr(\text{Elevated Mood} \mid \text{Language-Thought Disorder})\\ &\Pr(\text{Language-Thought Disorder})\\ &\Pr(\text{Speech Disorder} \mid \text{Language-Thought Disorder},\\ &\quad\text{Increased Energy})\end{aligned}\right\}\ \text{Reversing a score-equivalent arc}$$

Briganti, G., Scutari, M., & McNally, R. J. (2022). A tutorial on bayesian networks for psychopathology researchers. *Psychological methods*.

Colliders

- In a collider the two causes are likely to be negatively correlated outside a DAG (for instance, a Markov Random Field).

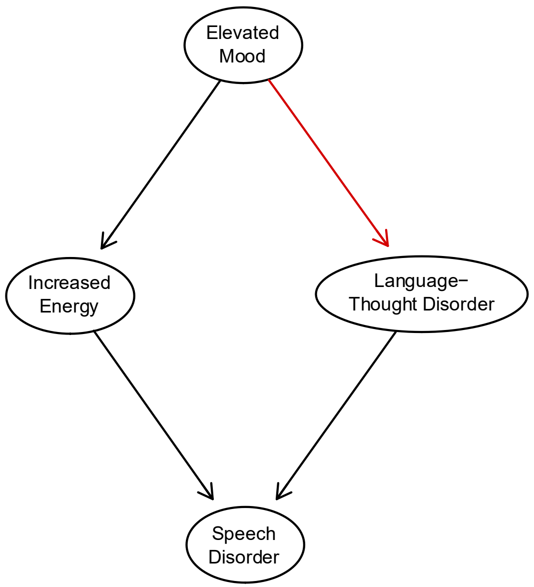
Equivalence classes

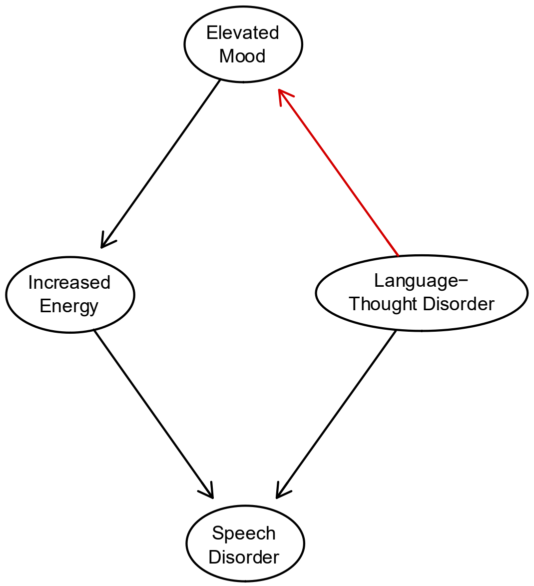
- Multiple configurations may be possible
- Except no new v structure

$$\underbrace{\Pr(v_j)\Pr(v_i \mid v_j)\Pr(v_k \mid v_i)}_{v_j \to v_i \to v_k} = \Pr(v_j)\,\frac{\Pr(v_i, v_j)}{\Pr(v_j)}\,\frac{\Pr(v_i, v_k)}{\Pr(v_i)} =$$

$$= \frac{\Pr(v_i, v_j)}{\Pr(v_i)}\Pr(v_i, v_k) = \underbrace{\Pr(v_i)\Pr(v_j \mid v_i)\Pr(v_k \mid v_i)}_{v_j \leftarrow v_i \to v_k} = \underbrace{\Pr(v_k)\Pr(v_j \mid v_i)\Pr(v_i \mid v_k)}_{v_j \leftarrow v_i \leftarrow v_k}$$

Briganti, G., Scutari, M., & McNally, R. J. (2022). A tutorial on bayesian networks for psychopathology researchers. *Psychological methods*.





$$\left.\begin{aligned}&\mathrm{Pr}(\text{Elevated Mood}, \text{Increased Energy},\\ &\quad \text{Language-Thought Disorder}, \text{Speech Disorder}) =\end{aligned}\right\}\ \text{Joint distribution}$$

$$\left.\begin{aligned}&\mathrm{Pr}(\text{Elevated Mood})\\ &\mathrm{Pr}(\text{Increased Energy} \mid \text{Elevated Mood})\\ &\mathrm{Pr}(\text{Language-Thought Disorder} \mid \text{Increased Energy},\\ &\quad \text{Elevated Mood})\\ &\mathrm{Pr}(\text{Speech Disorder} \mid \text{Language-Thought Disorder},\\ &\quad \text{Increased Energy}, \text{Elevated Mood}) =\end{aligned}\right\}\ \text{Chain rule}$$

$$\left.\begin{aligned}&\mathrm{Pr}(\text{Elevated Mood})\\ &\mathrm{Pr}(\text{Increased Energy} \mid \text{Elevated Mood})\\ &\mathrm{Pr}(\text{Language-Thought Disorder} \mid \text{Elevated Mood})\\ &\mathrm{Pr}(\text{Speech Disorder} \mid \text{Language-Thought Disorder},\\ &\quad \text{Increased Energy}) =\end{aligned}\right\}\ \text{Bayesian network, local distributions}$$

$$\left.\begin{aligned}&\cancel{\mathrm{Pr}(\text{Elevated Mood})}\\ &\mathrm{Pr}(\text{Increased Energy} \mid \text{Elevated Mood})\\ &\frac{\mathrm{Pr}(\text{Language-Thought Disorder}, \text{Elevated Mood})}{\cancel{\mathrm{Pr}(\text{Elevated Mood})}}\\ &\mathrm{Pr}(\text{Speech Disorder} \mid \text{Language-Thought Disorder},\\ &\quad \text{Increased Energy}) =\\ &\mathrm{Pr}(\text{Increased Energy} \mid \text{Elevated Mood})\\ &\mathrm{Pr}(\text{Language-Thought Disorder}, \text{Elevated Mood})\\ &\mathrm{Pr}(\text{Speech Disorder} \mid \text{Language-Thought Disorder},\\ &\quad \text{Increased Energy}) =\end{aligned}\right\}\ \text{Bayes theorem}$$

$$\left.\begin{aligned}&\mathrm{Pr}(\text{Increased Energy} \mid \text{Elevated Mood})\\ &\mathrm{Pr}(\text{Elevated Mood} \mid \text{Language-Thought Disorder})\\ &\mathrm{Pr}(\text{Language-Thought Disorder})\\ &\mathrm{Pr}(\text{Speech Disorder} \mid \text{Language-Thought Disorder},\\ &\quad \text{Increased Energy})\end{aligned}\right\}\ \text{Reversing a score-equivalent arc}$$

Probability distributions for BNs

- Gaussian
- Discrete
- Conditional Linear Gaussian

Structure Learning

- Learning the structure of BNs from data
- Possibly integrating expert knowledge
    - Ex. Whitelisting
    - Ex. Blacklisting
- Constraint based (d-separation)
- Score based (model fit criterion)
- Hybrid

Constraint based learning

- Conditional independences (d-separations) are the constraints
- Connect nodes that are not independent
- Constraint-based learning algorithms use a procedure that iteratively tests for (conditional) independence to build up the graph structure. Let us denote the skeleton of the graph as $G\_0=(V,E\_0)$, an undirected graph initially assumed to be fully connected.

Markov boundary

- Let us define the Markov boundary $\mathrm{MB}(\mathrm{X_i})$ as the minimal set of variables such that:

$$\mathrm{X_i} \perp\!\!\!\perp \mathrm{V} \setminus \{\mathrm{X_i} \cup \mathrm{MB}(\mathrm{X_i})\} \mid \mathrm{MB}(\mathrm{X_i}).$$

- Learning this boundary for each node provides a way to localize the search and is the basis of several local-to-global constraint-based algorithms (e.g., Grow-Shrink, HITON-PC).

- Moreover, under the assumption of faithfulness, the following property holds:

$(\mathrm{X_i} \perp\!\!\!\perp \mathrm{X_j} \mid S) \Rightarrow$ no edge between $\mathrm{X_i}$ and $\mathrm{X_j}$ in any DAG in the equivalence class,

- whereas

$(\mathrm{X_i} \not\!\perp\!\!\!\perp \mathrm{X_j} \mid S)$ for all $S \subset \mathrm{V} \setminus \{\mathrm{X_i}, \mathrm{X_j}\} \Rightarrow$ directed edge if part of a v−structure.

Edge Removal via Conditional Independence Tests

- For each pair of variables $\left(X_i, X_j\right)$, we attempt to find a subset $S \subseteq V \setminus \left\{X_i, X_j\right\}$ such that:

$$\left(X_i \perp\!\!\!\perp X_j \mid S\right)_P.$$

- If such a set exists, then the edge (X_i, X_j) is removed from the skeleton. The order of conditioning (i.e., the size of S) is typically increased iteratively.
- The procedure is made tractable by controlling the size of the conditioning sets, often under the assumption of sparsity (i.e., bounded in-degree or small Markov blankets).

Orienting Edges with Meek's Rules

- Once the skeleton is learned, the next step is to orient as many edges as possible into directed ones, while ensuring acyclicity and preservation of the conditional independencies. This is typically done using Meek's orientation rules and by identifying v-structures.

- A v-structure is a triple (X_i, X_k, X_j) such that:

$$X_i \to X_k \leftarrow X_j, \text{ but } X_i \nrightarrow X_j \text{ (nor vice versa).}$$

- This structure is identified when:

$$(X_i \not\perp X_j \mid S), \text{ for all } S \subseteq V \setminus \{X_i, X_j\} \text{ such that } X_k \in S.$$

Limitations of constraint-based learning

- Under appropriate assumptions (faithfulness, sufficiently large sample size), constraint-based methods are asymptotically consistent, i.e., they recover the correct Markov equivalence class of the DAG as $N \to \infty$.
- However, they suffer from several practical limitations:
- Multiple testing: procedure involves many statistical tests, leading to inflated Type I errors unless corrected.
- Statistical power: conditional independence tests lose power as the size of the conditioning set increases.
- Equivalence classes: constraint-based methods recover the CPDAG (completed partially directed acyclic graph), not the full DAG unless additional assumptions are made.

Testing for conditional independence relations

- Asymptotic $\chi 2$
- Hotelling's test (modified « t » test) for continuous
- Jonckheere-Terpstra test for ordinal variables (modified Wilcoxon test)

Asymptotic chi square for discrete data

- Under the null hypothesis $H_0: X \perp Y \mid Z$, one compares observed joint frequencies with expected frequencies computed under the assumption of conditional independence. The test statistic takes the form:

$$\chi^2 = \sum_{i,j,k} \frac{\left(O_{ijk} - E_{ijk}\right)^2}{E_{ijk}},$$

- where O_{ijk} denotes the observed frequency for each configuration of $X = i, Y = j, Z = k$, and $E_{ijk} = \frac{O_{i+k} \cdot O_{+jk}}{O_{++k}}$ is the expected frequency under independence. Under mild regularity conditions and large sample sizes, the statistic converges in distribution to a chi-square distribution with degrees of freedom equal to:

$$df = (|X| - 1)(|Y| - 1)|Z|$$

- where $|X|, |Y|$, and $|Z|$ are the cardinalities of the respective variable domains.

Hotelling’s modified t test

- Suppose $(X, Y, Z) \sim N$, and we wish to test whether $\rho_{XY \cdot Z} = 0$. Then the statistic:

$$t = \rho_{XY \cdot Z} \cdot \sqrt{\frac{n - |Z| - 2}{1 - \rho_{XY \cdot Z}^2}},$$

- is approximately distributed as a Student's t with $n - |Z| - 2$ degrees of freedom under the null. This is mathematically equivalent to a linear version of Hotelling's test when the conditioning set Z is projected out via regression residuals. Alternatively, for testing the joint mean vector μ across multiple variables, the Hotelling T^2 statistic is:

- $T^2 = n(\bar{x} - \mu_0)^T S^{-1} (\bar{x} - \mu_0)$, which under H_0 follows: $\frac{(n-p)}{p(n-1)} T^2 \sim F_{p,n-p}$

- where p is the dimension of x, n is the sample size, S is the sample covariance matrix.

Jonckheere-Terpstra test

- nonparametric rank-based test specifically designed for detecting ordered alternatives in independent samples (natural generalization of the Wilcoxon-Mann Whitney test) when the grouping variable has an ordinal structure.

- In the context of constraint-based learning, the JT test is used when one or more of the variables (e.g., X or Y) are ordinal, and the hypothesis of interest is: H_0 : The distribution of X is identical across ordered levels of Y (or Z), versus the alternative that there is a monotonic trend in the distribution across the ordered categories. The test statistic is constructed from the sum of Wilcoxon rank-sums across all pairs of ordered groups, and its null distribution is asymptotically normal:

$$\mathrm{Z} = \frac{\mathrm{U} - \mathrm{E}[\mathrm{U}]}{\sqrt{\mathrm{Var}(\mathrm{U})}},$$

- where U is the sum of concordant pairs across groups.

Inductive causation algorithm

Algorithm 1 Inductive Causation (IC) Algorithm

Require: A set of observed variables $\mathcal{V} = \{X_1, X_2, \ldots, X_n\}$
Require: An oracle (or statistical test) for conditional independence $X_i \perp\!\!\!\perp X_j \mid S$
Require: A significance level α
Ensure: A Completed Partially Directed Acyclic Graph (CPDAG) $G = (V, E)$

Initialize $G = (V, E)$ as the complete undirected graph over $\mathcal{V}$
for all pairs $(X_i, X_j) \in \mathcal{V} \times \mathcal{V},\ i \neq j$ **do**
    Set Sepset$[X_i][X_j] \leftarrow \emptyset$
end for

▷ Skeleton Construction

for all pairs $(X_i, X_j) \in \mathcal{V} \times \mathcal{V},\ i \neq j$ **do**
    for all subsets $S \subseteq \mathcal{V} \setminus \{X_i, X_j\}$ **do**
        if $X_i \perp\!\!\!\perp X_j \mid S$ at level α **then**
            Remove edge (X_i, X_j) from G
            Set Sepset$[X_i][X_j] \leftarrow S$, Sepset$[X_j][X_i] \leftarrow S$
            break
        end if
    end for
end for

Inductive causation algorithm

▷ V-Structure Orientation

for all triples $(X_i, X_j, X_k) \in \mathcal{V}^3$ **do**
    if X_i and X_k are non-adjacent, but both adjacent to X_j **then**
        if $X_j \notin \text{Sepset}[X_i][X_k]$ **then**
            Orient $X_i \rightarrow X_j \leftarrow X_k$
        end if
    end if
end for

Inductive causation algorithm

▷ Apply Meek's Orientation Rules

repeat
Apply the following rules:
Rule 1: If $X_i \to X_j$ and $X_j - X_k$, with $X_i \not\sim X_k$, then orient $X_j - X_k$ as $X_j \to X_k$
Rule 2: If a directed path $X_i \to \cdots \to X_j$ exists and $X_i - X_j$, orient $X_i - X_j$ as $X_i \to X_j$
Rule 3: If $X_i - X_j$, $X_j \to X_k$, and $X_i \to X_k$, then orient $X_i - X_j$ as $X_i \to X_j$
until no more edges can be oriented
return the resulting CPDAG G

Inductive causation algorithm

- First & simplest constraint based algorithm
- We start with a complete network (every node is connected)
- For each pair A and B, search for a set S(AB)
- If S exists, then remove edge A-B

Inductive Causation Algorithm

- For each unconnected A and B such as A-C and B-C, if C is not in S, then A→ C← B
- Determining v-structures is fundamental, it builds the essential structure of the BN
- If A-B in a completely directed path , then A→B
- If A→C and C-B, then A→ C→ B

Score-based algorithms

- Assign a score to each candidate BN
- Explore different DAG propositions by single edge addition, removal and reversal
- Goal: no new network has a score such as

Score(new network) > maxscore

Exemple : Hill-Climbing Algorithm

Algorithm 2 Hill-Climbing Algorithm

Require: Dataset $\mathcal{D}$ over variables $\mathcal{V} = \{X_1, X_2, \ldots, X_n\}$
Require: Scoring function $\text{Score}(G, \mathcal{D})$
Ensure: A DAG $G^* = (V, E^*)$ that (locally) maximizes the score

- Initialize $G \leftarrow G_0$ (typically the empty graph)
- Compute $S \leftarrow \text{Score}(G, \mathcal{D})$
- **repeat**
  - $\text{BestScore} \leftarrow S$
  - $G_{\text{best}} \leftarrow G$
  - **for all** legal operations $\mathcal{O} \in \{\text{Add}, \text{Delete}, \text{Reverse}\}$ **do**
    - **for all** node pairs $(X_i, X_j) \in \mathcal{V} \times \mathcal{V},\ i \neq j$ **do**
      - **if** $\mathcal{O}$ is admissible on (X_i, X_j) without creating a cycle **then**
        - $G' \leftarrow \mathcal{O}(G, X_i, X_j)$
        - $S' \leftarrow \text{Score}(G', \mathcal{D})$
        - **if** $S' > \text{BestScore}$ **then**
          - $\text{BestScore} \leftarrow S'$
          - $G_{\text{best}} \leftarrow G'$
        - **end if**
      - **end if**
    - **end for**
  - **end for**
  - $G \leftarrow G_{\text{best}}$
  - $S \leftarrow \text{BestScore}$
- **until** $G = G_{\text{best}}$ ▷ No operation improved the score
- **return** G

Hybrid algorithms

- Conditional independence constraints to reduce the number of candidate network
- Then compute a score for the remaining candidates

- Exemple: the rsmax2 algorithm

The RSmax2 algorithm

Algorithm 3 RSmax2 Algorithm

Require: Dataset $\mathcal{D}$ over variables $\mathcal{V} = \{X_1, X_2, \ldots, X_n\}$
Require: Conditional independence test procedure CI-Test$(\cdot)$
Require: Scoring function Score$(G, \mathcal{D})$
Require: Significance level α
Ensure: A DAG $G^* = (V, E^*)$ approximating the underlying causal structure

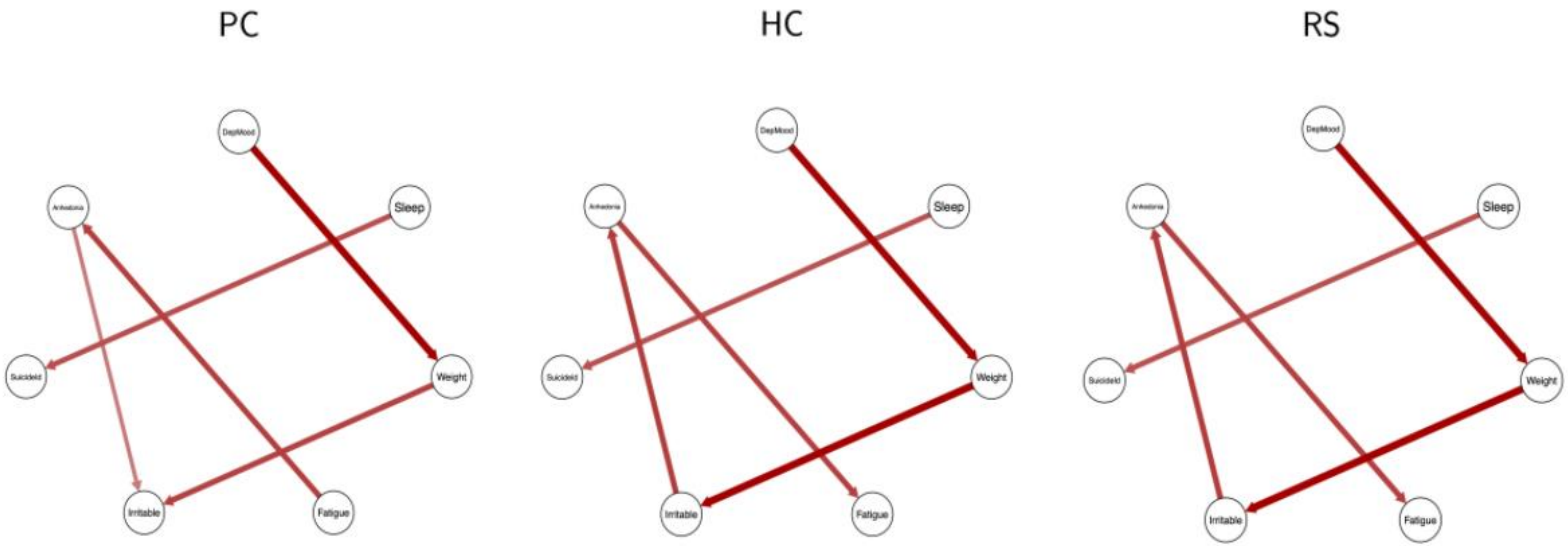
$\triangleright$ Phase I: Constraint-Based Local Structure Discovery

for all variables $X_i \in \mathcal{V}$ **do**
  Initialize candidate parent set $\text{Cand}(X_i) \leftarrow \emptyset$
  for all $X_j \in \mathcal{V} \setminus \{X_i\}$ **do**
    if $X_i \not\perp X_j \mid S$ for all $S \subseteq \mathcal{V} \setminus \{X_i, X_j\}$ **then**
      Add X_j to $\text{Cand}(X_i)$
    end if
  end for
end for
Define a constrained search space $\mathcal{S}$ where arcs $X_j \to X_i$ are allowed only if $X_j \in \text{Cand}(X_i)$

$\triangleright$ Phase II: Score-Based Search over the Restricted Space

Initialize graph $G \leftarrow G_0$ (typically empty)
Compute $S \leftarrow \text{Score}(G, \mathcal{D})$
repeat
  $\text{BestScore} \leftarrow S$
  $G_{\text{best}} \leftarrow G$
  for all legal operations $\mathcal{O} \in \{\text{Add}, \text{Delete}, \text{Reverse}\}$ **do**
    for all node pairs $(X_i, X_j) \in \mathcal{V} \times \mathcal{V},\ i \neq j$ **do**
      if $\mathcal{O}(X_i, X_j) \in \mathcal{S}$ and acyclicity is preserved **then**
        $G' \leftarrow \mathcal{O}(G, X_i, X_j)$
        $S' \leftarrow \text{Score}(G', \mathcal{D})$
        if $S' > \text{BestScore}$ **then**
          $\text{BestScore} \leftarrow S'$
          $G_{\text{best}} \leftarrow G'$
        end if
      end if
    end for
  end for
  $G \leftarrow G_{\text{best}}$
  $S \leftarrow \text{BestScore}$
until $G = G_{\text{best}}$
return G

Choice of algorithm



Briganti, G., Scutari, M., & McNally, R. J. (2022). A tutorial on bayesian networks for psychopathology researchers. *Psychological methods*.

BNs or PMRFs? It depends

- PMRFs cannot distinguish whether a symptom X is more likely to cause or be caused by other variables
- The collider problem

Stability of BNs

- Bootstrapping the structure learning
- Example
- Include edges that exist in 85+% of bootstrapped networks
- Direction in more than 50%

What can Bayesian Networks do for us?

Retrospective data can inform every step of knowledge generation

- **Identify a research question**
- Design a study protocol
- Data collection (how to)
- Analyze the data (how to)
- **Interpret the analysis to get an answer**

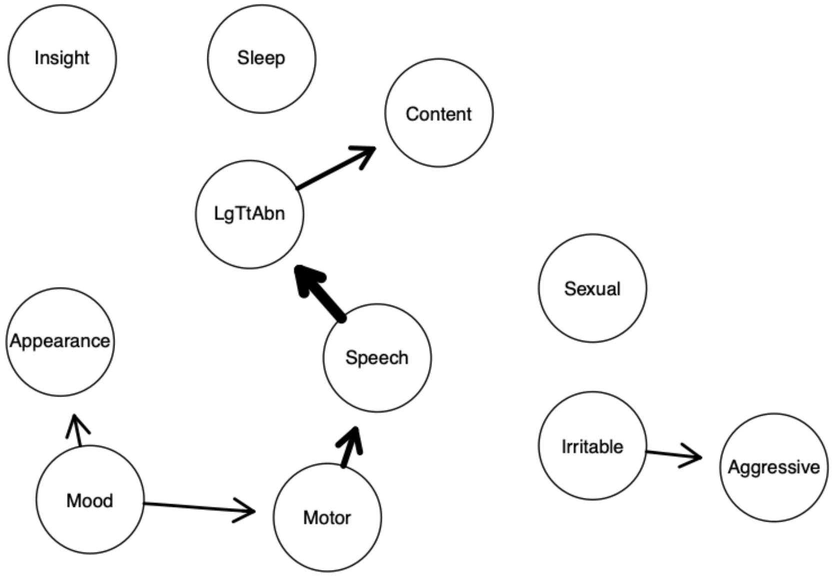
Limitations of statistical language

- Failing to address the pressing causal questions
- Limiting the pace for hypothesis generation

Bayesian Networks

- We can use them to perform causal inference (if assumptions are met)
- We can use them to generate many research hypotheses fast (if assumptions not met)
- Address the uncertainty in a more quantitative way in all cases

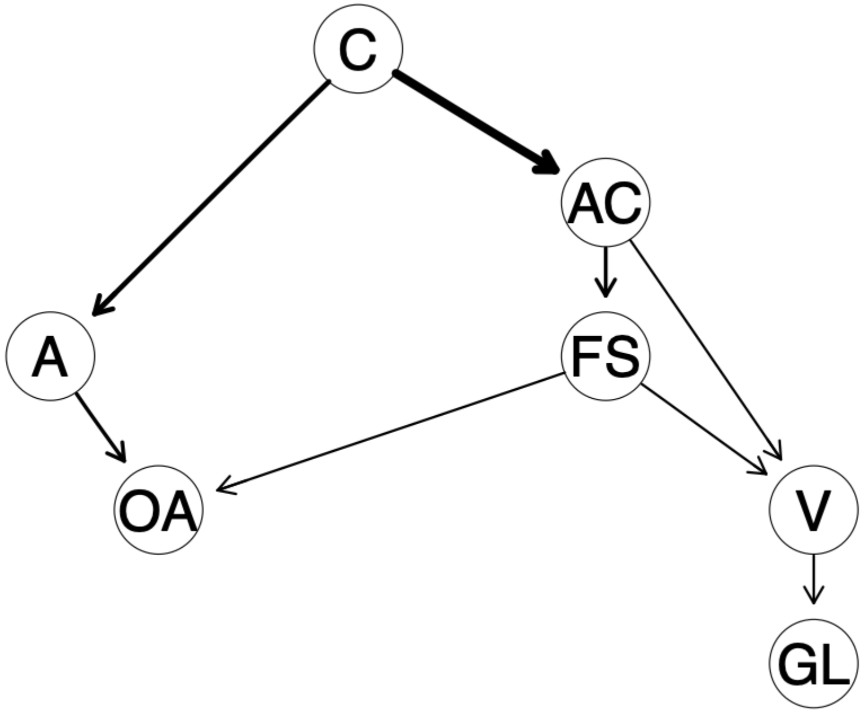
Example 1 – Bipolar disorder
Mood and Motor activity/energy



Item	Symptom
Mood	Elevated Mood
Motor	Increased Motor Activity-Energy
Sexual	Sexual Interest
Sleep	Sleep
Irritable	Irritability
Speech	Speech (Rate and Amount)
LgTtAbn	Language-Thought Disorder
Content	Content
Aggressive	Disruptive-Aggressive Behavior
Appearance	Appearance
Insight	Insight

Briganti, G. (2022). On the use of Bayesian artificial intelligence for hypothesis generation in psychiatry. *Psychiatria Danubina*, *34*.

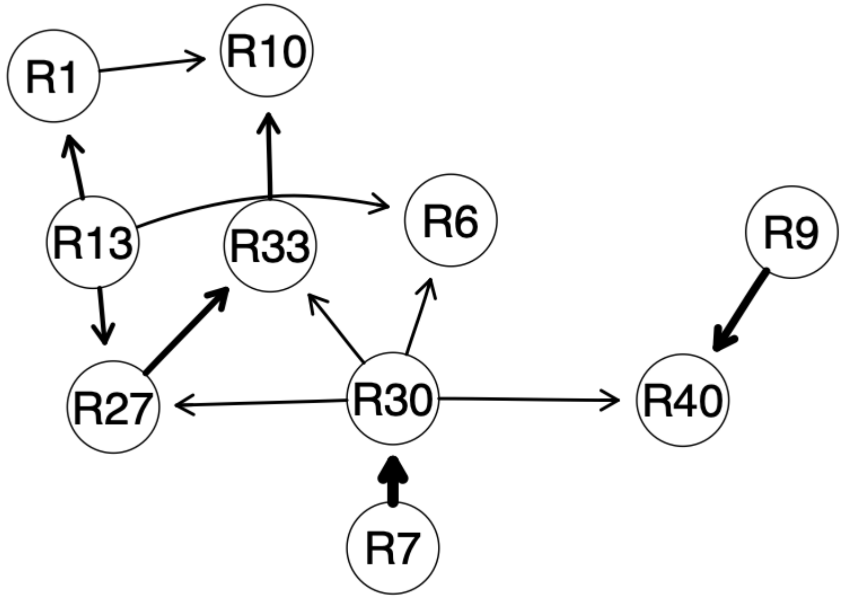
Example 2 – Self-Worth
The isolation of God's love



Item	Domain
FS	Family Support
C	Competition
A	Appearance
GL	God's Love
AC	Academic Competence
V	Virtue
OA	Other's Approval

Briganti, G. (2022). On the use of Bayesian artificial intelligence for hypothesis generation in psychiatry. *Psychiatria Danubina*, *34*.

Example 3 – Narcissistic personality
Exhibitionism, authority and superiority as parent nodes



N°	0	1	Domain
R1	I am not good at influencing people	I have a natural talent for influencing people	Authority
R6	I try to accept the consequences of my behavior	I can usually talk my way out of anything	Exploitativeness
R7	I prefer to blend in with the crowd	I like to be the center of attention	Exhibitionism
R9	I am no better or worse than most people	I think I am a special person	Superiority
R10	I am not sure if I would make a good leader	I see myself as a good leader	Authority
R13	I don't like it when I find myself manipulating other people	I find it easy to manipulate people	Exploitativeness
R27	Power for its own sake doesn't interest me	I have a strong will to power	Entitlement
R30	It makes me uncomfortable to be the center of attention	I really like to be the center of attention	Exhibitionism
R33	It makes little difference to me whether I am a leader or not	I would prefer to be a leader	Authority
R40	I am much like everybody else	I am an extraordinary person	Superiority

Briganti, G. (2022). On the use of Bayesian artificial intelligence for hypothesis generation in psychiatry. *Psychiatria Danubina*, *34*.

What can BNs do for you? -1

- Identify a great number of sets of independent and dependent variables
- Identify a great number of potential clinical prediction models
- For each symptom in the network, identify local predictors
- Improve the quality of clinical prediction models

What can BNs do for you? -2

- Identify the potential v-structures in a multivariate data set
- Interpreting negative edges in PMRFs

What can BNs do for you? -3

- A “white-box” kind of AI
- Nodes are observed
- Local models can be easily identified for future investigation
- Address the inherent complexity of such models

Briganti, G., Decety, J., Scutari, M., McNally, R. J., & Linkowski, P. (2022). Using Bayesian networks to investigate psychological constructs: The case of empathy. *Psychological reports*, 00332941221146711.

Conclusion

- Plenitude of methods to study networks in human behavior
- Application should be methodologically rigorous
- Advancing methodology needs expert knowledge
- Next step: multilevel network modeling
- Towards Precision Psychiatry!

Thank you!