Causal Inference in Psychopathology: an overview

Giovanni Briganti, MD, PhD
Postdoctoral Researcher
Richard J. McNally Laboratory
Harvard University
gbriganti@fas.harvard.edu

Laboratory Meeting March 22, 2021

HARVARD UNIVERSITY

The paper

A tutorial on Bayesian Networks for psychopathology researchers (submitted)

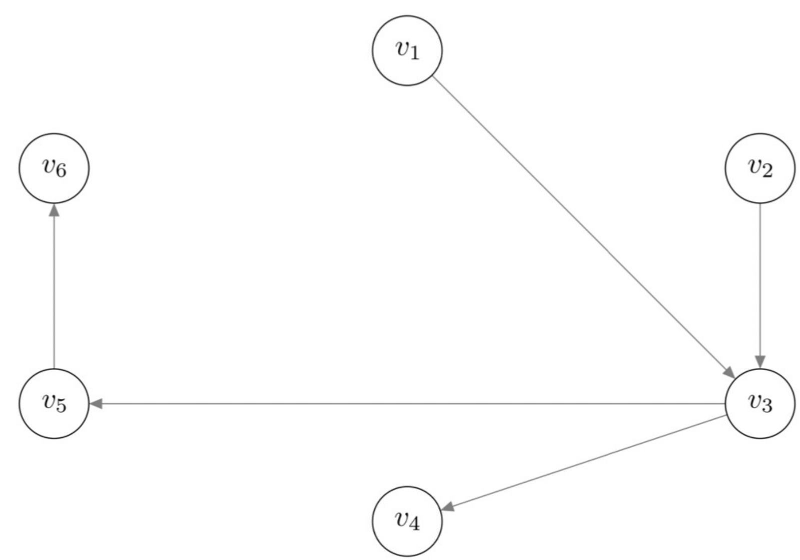
Briganti, Scutari and McNally

The network framework

- Network theory: mental disorder arises from interactions among its constituent symptoms
- Network analysis: partial correlation networks
- Undirected networks
- Key limitation: A → B ? Or B → A ?
- Bayesian networks → rigorous causal inference

Bayesian Networks (BNs)

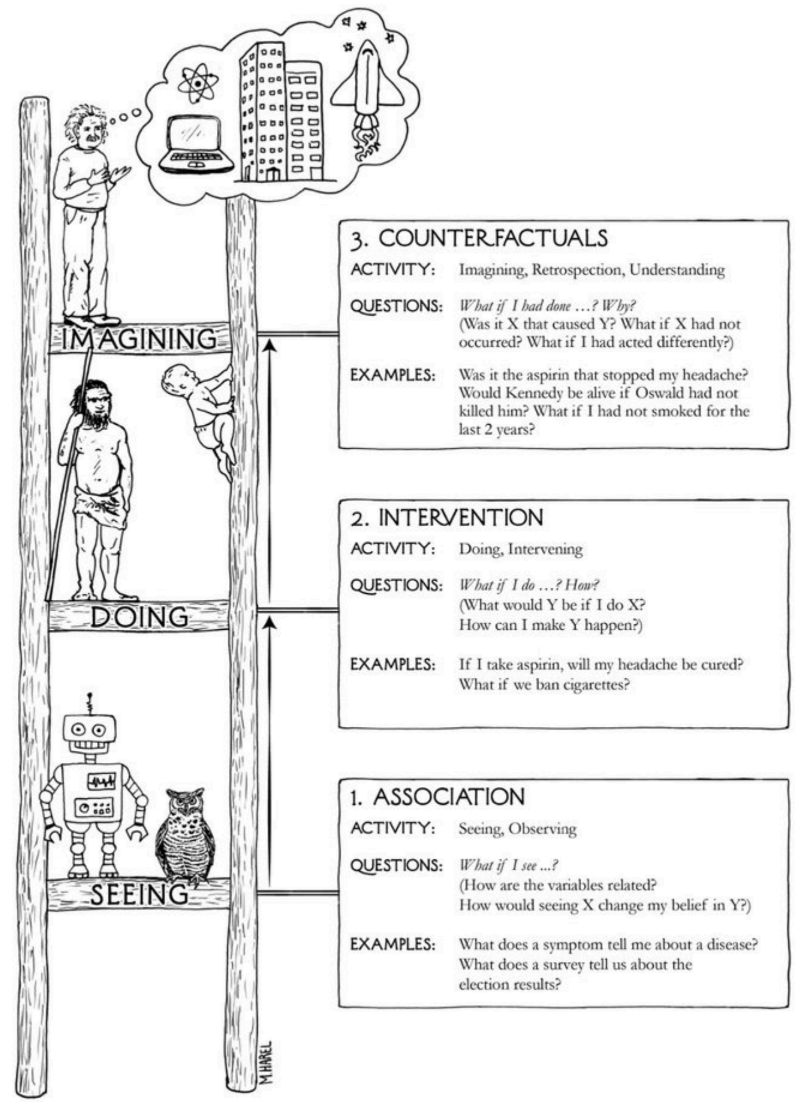
- Defined by
  - Directed Acyclic Graph (DAG)
  - Joint probability distribution of variables
- Goal of DAGs: expressing *conditional independence relations* through graphical separation
- The rest of variables: expression of *causal effects*
- Admissible causal relations in observational data



Let us take a step back...

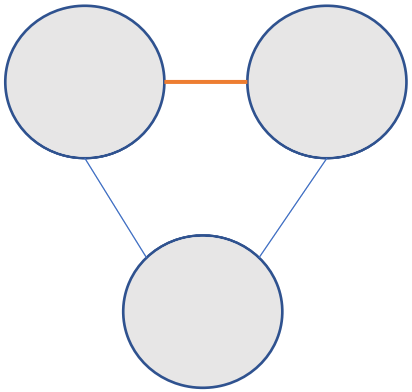
The Ladder of Causation (Judea Pearl)

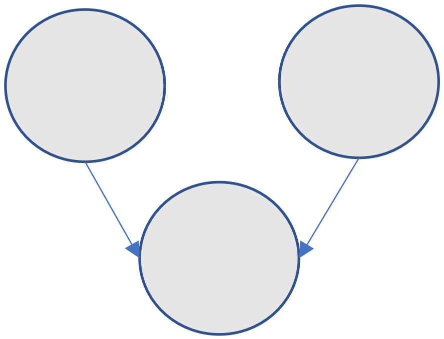
- The case of psychopathology
- The data generating process is unknown
- We try to uncover it with cross-sectional studies
- The number/type of variables and interrelations



What can BNs add to the network field?

- Addressing centrality
  - Centrality = measure of interconnectedness
  - Based on edge weight
  - Undirected networks
  - Central node A: a cause or effect?
- The issue of negative edges
  - Berkson's Bias?
  - True negative connection?
- Causal insight is needed





Directed Acyclic Graphs (1)

- Math object G(V, A) of a set of nodes and edges
- Directed edges: (vi,vj) =/= (vj,vi)
- Acyclicity: no loops (vi→vi) or cycles
- Markov property
  - graphical separation → probabilistic independence
  - Decomposition of larger model into smaller models

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

- Markov blanket → independence map (d-separations)

Directed Acyclic Graphs (2)

- We are not sure whether the causal structure proposed in the *only* possible, based on the data

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

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

- A number of edges can be directed either way
- Those edges are rended *undirected*
- Completed partially directed graph (cpDAG)

Probability distributions

- Discrete BNs: $Xi|\prod Xi$ are multinomial
  - Conditional probability tables
- Gaussian BNs → partial correlation matrix (GGM)
  - Local distributions are linear regressions
  - $\prod Xi$ are regressors
- Conditional Linear Gaussian BN
  - Both discrete and continuous nodes
  - Continuous nodes have a set of linear regression, one for each configuration of discrete parents, with continuous parents as regressors

Structure Learning (1) - Assumptions

- Relationships are *conditional independencies*
- Each node is one variable
- Cross-sectional settings (if not, then dynamic BNs)
- Global probability distribution is strictly positive
  - Every event is observable

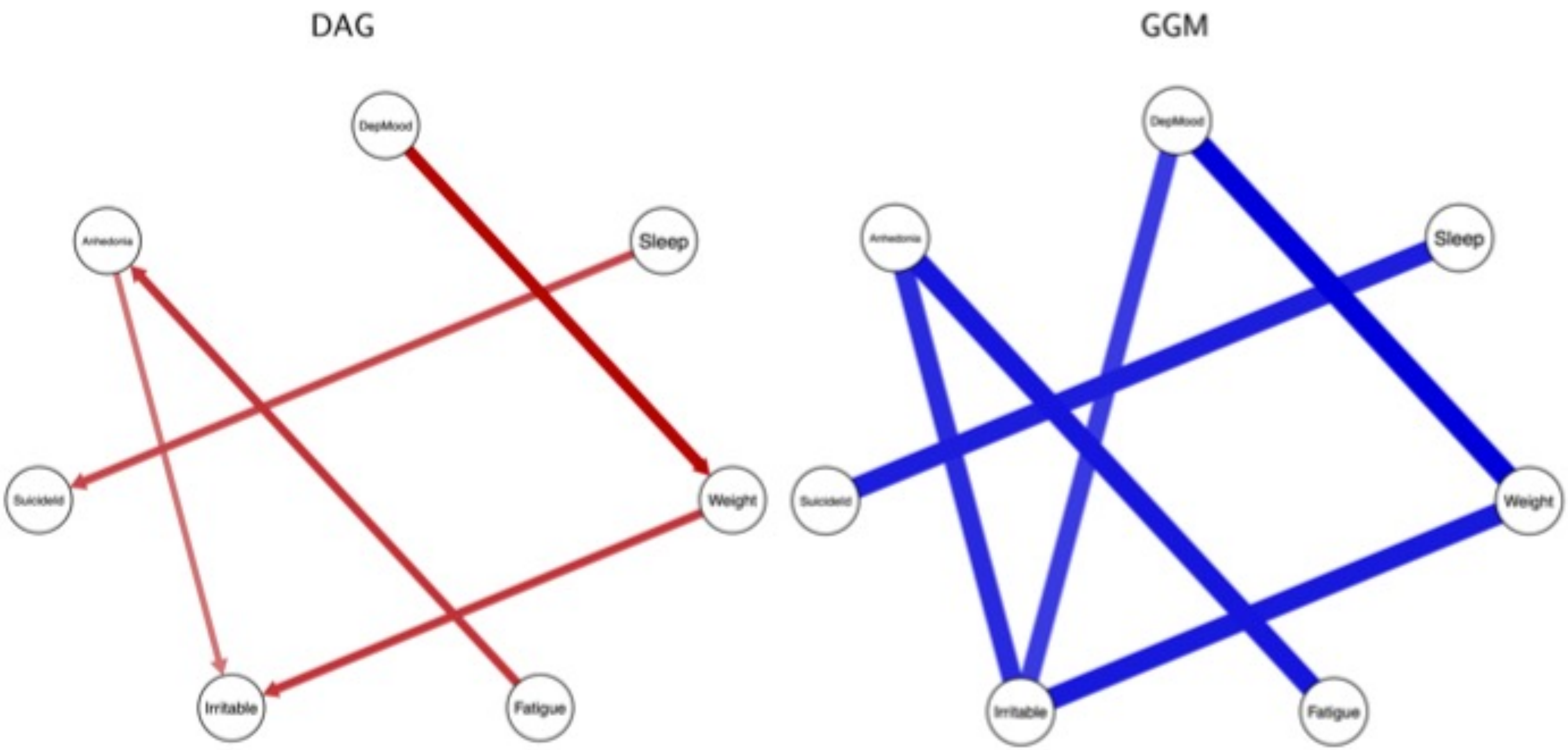
- All these four assumptions are usually met in psychopathological data

Structure learning (2) - Stability

- Bootstrapping
- Strength: Edge appears in $> 85\%$ of samples
- Minimum direction: direction appears in $> 50\%$ of samples
- Note : contrary to the GGM literature, the BN methodology directly shows the bootstrapped network

Structure Learning (3) Constraint-based algorithms

- Test for conditional independence → constraint
- Inductive causation algorithm
  - For each A and B in X
  - S(AB) of nodes such that A and B are conditionally independent given S(AB). Deletes edge between A and B if S(AB) exists
  - If A and B are both connected to C, and C ∈/ S(A,B), then A→C ←B
  - All other edges are directed accordingly
    - A and B are adjacent in a path where all other edges are directed then A→B
    - If A-C and C-B, then A→C→B


DAG
DepMood
Anhedonia
Sleep
Suicideld
Weight
Irritable
Fatigue
GGM
DepMood
Anhedonia
Sleep
Suicideld
Weight
Irritable
Fatigue

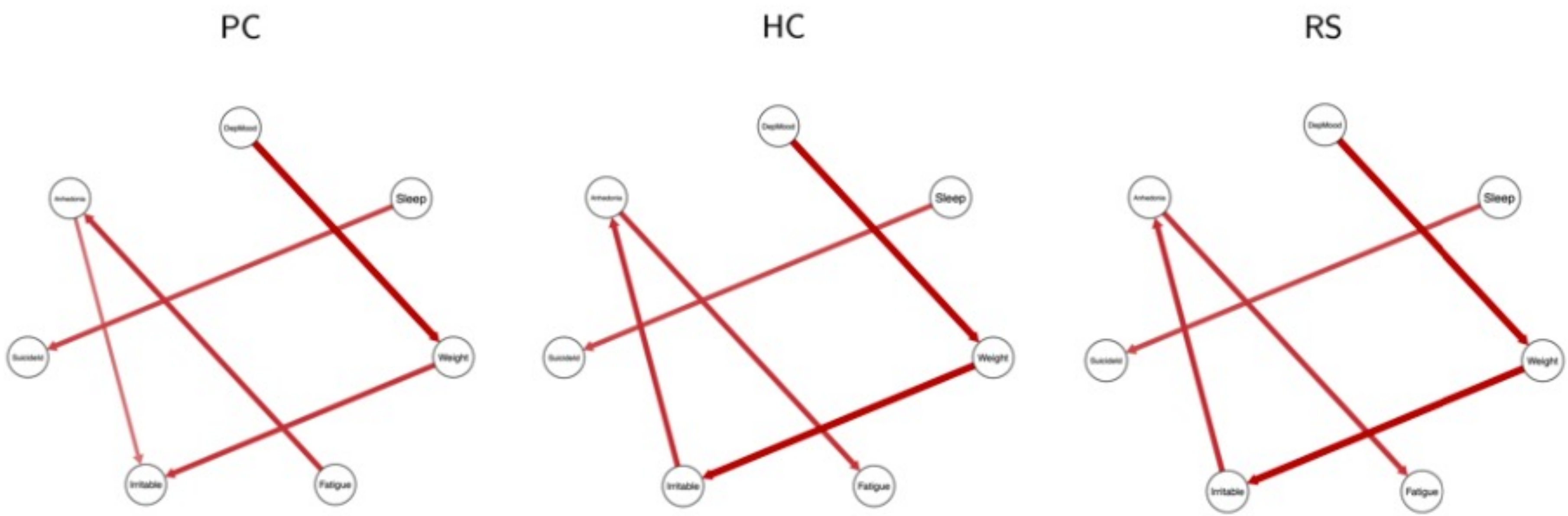
Structure Learning (4) – Score-based algorithms

- Score assigned to each candidate BN, such as BIC
- Hill-climbing algorithm
  - Greedy Search (edge by edge re-evaluation)
  - First network Network G ← Score (G)
  - Maxscore ← Score (G)
  - Algorithm searches for a modified network G* such as Score (G*) > Score (G) → G* becomes the new candidate network G
  - _ _ _ until no G* has Score (G*) > Score (G).

Structure Learning (5) – Hybrid algorithms

- Restricts parents of a node to small subset of candidates
- Sparse Candidate Algorithm
  - Parents of a node Xi in a set Ci
  - Maximize Score(G*) among networks where parents of Xi are included in Ci
  - Sets new network G*=G

Structure learning (6) – Comparing algorithms



Common issues

- No systematic differences as to the choice of algorithm
- Impossible to distinguish equivalent DAGs
- Assumptions for causal inference
  - Causal Markov assumptions (conditional independence of a node with its non-effects)
  - D-separations in the DAG are the only ones in the distribution
  - No latent variable
    - Redundancies due to factors (goldbricker, UVA)
- Way forward: RCTs
- Blacklisting clinically implausible edges (McNally 2016)
  - Comparing machine and human expert knowledge

Thank you!