

Predicting putative intervention effects

after causal structure learning from survey data

Giusi Moffa

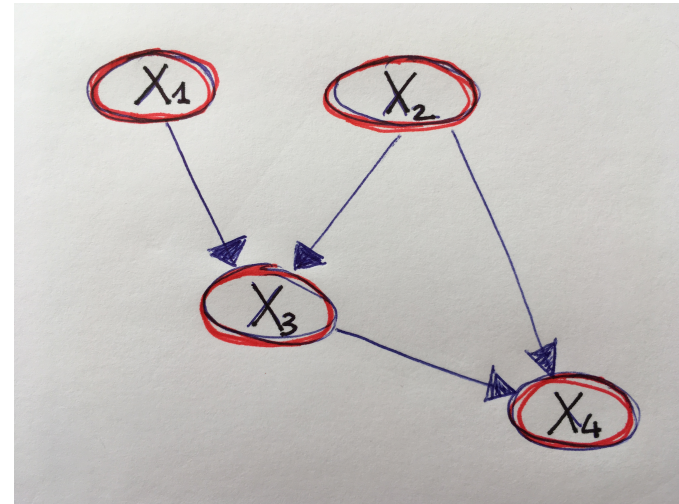
BBS Seminar, 1 November 2019, Basel

Probabilistic DAGs - Bayesian networks

Compact representation of multivariate probability distributions

Probabilistic graphical models for a set of variables $\{X_1, X_2, \dots, X_n\}$ characterized by

- a graphical structure, directed and acyclic, whose nodes are the variables
- a probability model for each node describing the relationship with its parents
- edges encode conditional independencies (any variable is conditionally independent of its non-descendant given its parents)



$$P(X_1)P(X_2)P(X_3|X_1, X_2)P(X_4|X_2, X_3)$$

e.g. $X_4 \perp\!\!\!\perp X_1 | (X_2, X_3)$

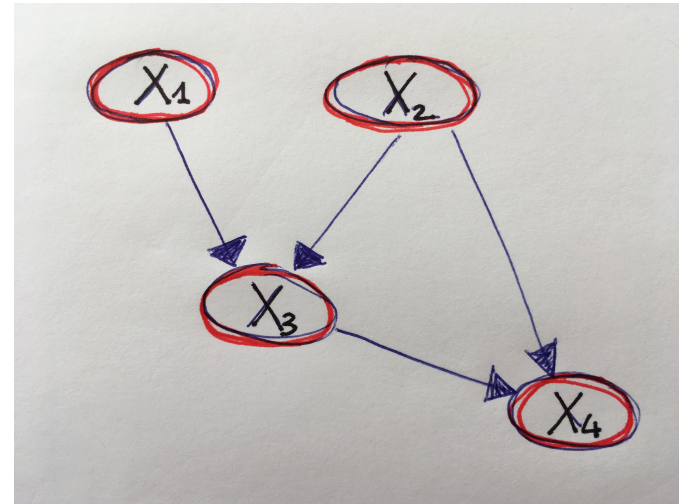
$$\{X_1, X_2, \dots, X_n\} \sim P(X_1, X_2, \dots, X_n) = \prod_{i=1}^n P(X_i | \mathbf{Pa}_i)$$

Causal DAGs and Markov condition

Causal interpretation

Graphical representation of structural causal models

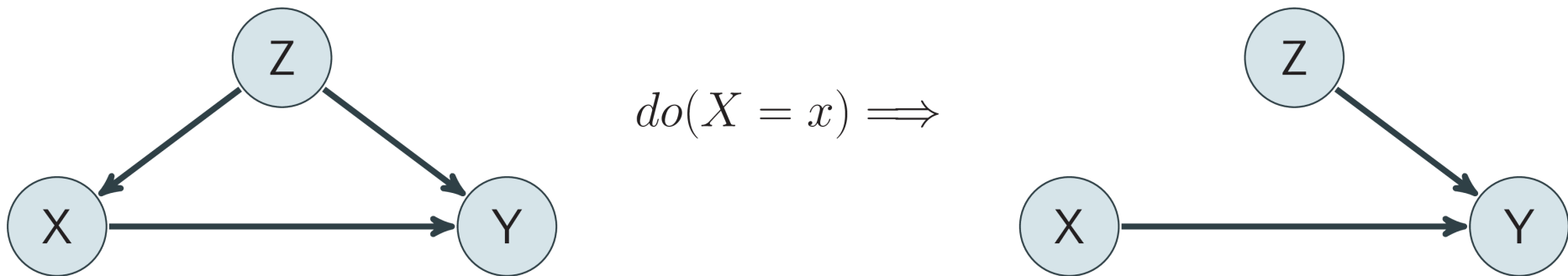
- qualitative
- directed edges imply direct causes (e.g. X_2 is a direct cause of X_4)
- directed paths imply potential causes (e.g. X_1 is a potential cause of X_4)



All common causes, even if unmeasured, of any pair of variables on the graph are themselves on the graph

Causal Markov condition: given a causal DAG representation of a system, it also represents its conditional independence (CI) properties

Intervention effects from (a known) causal DAGs



Pearl do calculus

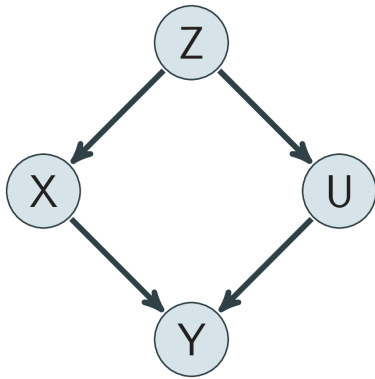
Average causal effect: $P(Y = 1|do(X = 1)) - P(Y = 1|do(X = 0))$ [causal estimand]

Causal effect: $P(Y = y|do(X = x)) = P_m(Y = y|X = x)$ [conditional probability in manipulated model]

Adjustment formula: $P(Y = y|do(X = x)) = \sum_z P(Y = y|X = x, Z = z)P(Z = z)$
[only in terms of preintervention probabilities]

Intervention effects in terms of propensity scores

Given a causal DAG



More generally:

$$P(Y = y | do(X = x)) = \sum_z P(Y = y | X = x, \mathbf{Pa}(X) = z) P(\mathbf{Pa}(X) = z)$$

$\mathbf{Pa}(X)$: parents of X in the DAG

$$P(Y = y | do(X = x)) = \sum_z \frac{P(Y = y, X = x, \mathbf{Pa}(X) = z)}{\underbrace{P(X = x | \mathbf{Pa}(X) = z)}_{\text{Propensity score}}}$$

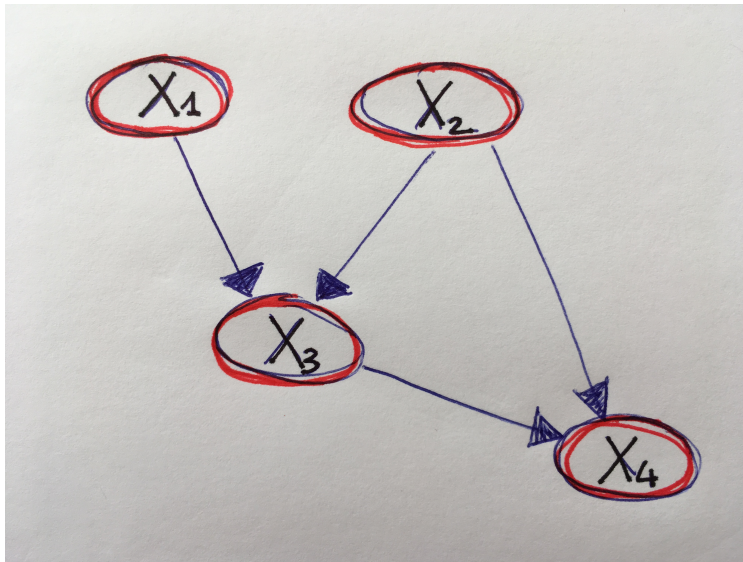
Reweighting samples \Rightarrow fictitious population from **post-intervention** distribution

Given a DAG, graphical criteria (e.g. **back and front door**) inform identifiability of causal effects

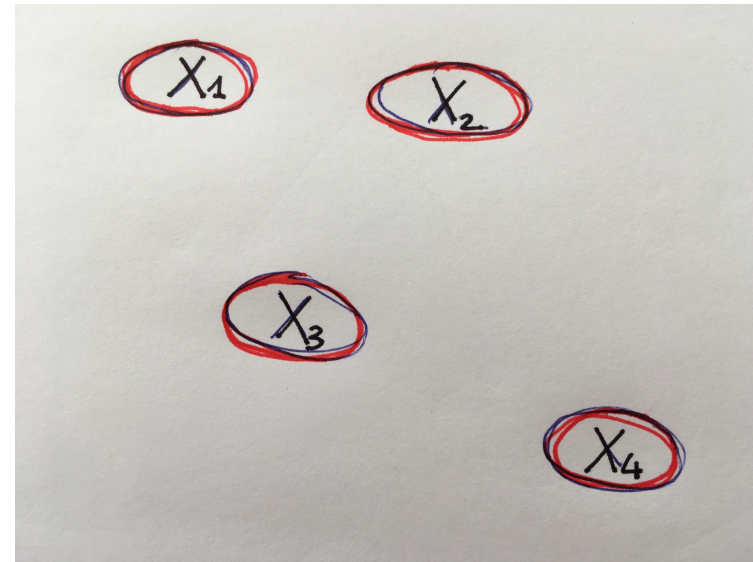
Inference in probabilistic graphical models

Two main tasks

- Parameter estimation (for a given probabilistic model of a node conditional on its parents)
- Structure learning (identify the connections, the more challenging task)



$$P(X_i | \mathbf{Pa}_i) = ?$$



$$X_i ? X_j$$

Structure learning approaches

Constraint-based methods

- PC (Peter and Clark) algorithm: reverse-engineering of the CIs of the distribution

Score and search algorithms

- Scoring function typically derived from a Bayesian approach

$$P(G|D) \propto P(D|G)P(G) \quad \text{Likelihood} \times \text{Prior}$$

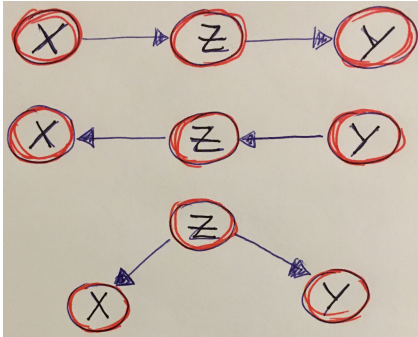
- e.g. Greedy search, hill climbing, dynamic programming, ILP

MCMC: posterior sampling, with recent developments with partition MCMC (Kuipers and Moffa, JASA 2017)

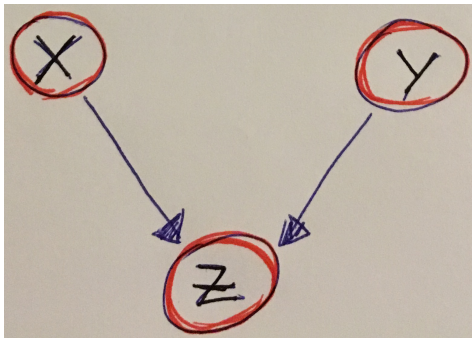
Hybrid methods

- First prune the search space
- Then score and search
 - Recent efficient procedure in Kuipers, Suter and Moffa; [arXiv:1803.07859 \(2018\)](https://arxiv.org/abs/1803.07859) and [BiDAG R package on CRAN](https://cran.r-project.org/web/packages/BiDAG/index.html)

Markov equivalence

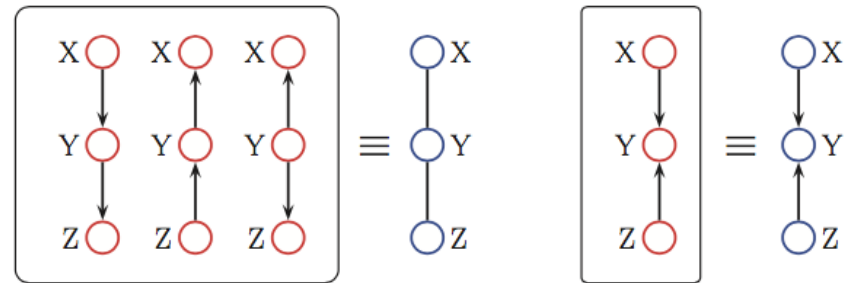


$$Y \perp\!\!\!\perp X|Z \equiv X \perp\!\!\!\perp Y|Z$$



$$Y \perp\!\!\!\perp X$$

Even from perfect data \Rightarrow learning up to an equivalence class



CPDAG (Completed Partially DAG)

Causal discovery of DAGs - Some assumptions

Causal representation: There exists some DAG G that is a causal DAG representation of the system.

Causal Markov condition: The identical DAG G also represents (by means of the Markov condition) the probabilistic conditional independence properties of the system.

Causal faithfulness: The causal DAG G is a probabilistically faithful representation of the system

- in plain English: all and only the independencies of the probability distribution are encoded in the graph
- beware of poligamy: the same set of conditional independence relationships can be described by different DAGs, so the same distribution may be faithful to many DAGs

Causal sufficiency: No unmeasured confounders

A. Philip Dawid, 2009, Beware of the DAG!

A case study in Psychosis - Medical background

Psychosis: medical equivalent of the lay idea of madness

Schizophrenia: best known psychotic disorder, 0.5% prevalence in the general population

- defined in terms of particular symptoms (delusions, hallucinations)
- but many more: worry, anxiety, depression
- search for physical causes of little success, e.g. 200 genes with small effects and unclear interactions, and neurophysiological abnormalities not consistently identified

Alternative explanations in the aetiology of Schizophrenia

- Social causes, e.g. stressful experiences, traumas like sexual abuse and bullying
- Interactions between symptoms

Bullying - a damaging experience

- Characterised by
 - abuse, intrusiveness, threat and the actuality of harm
 - exaggeration and distortion of power relationships
 - short and long term consequences
- Effects likely to operate through **cognitive-emotional biases** (with lowered mood):
 - increased self-focus,
 - often catastrophic reduction in self-regard,
 - anticipation of further episodes,
 - negative interpretation of ambiguous events
- Commonly leads to
 - mood disorders and suicidal ideation
 - psychotic symptoms and disorders, particularly persecutory ideation

Bullying - research question

Is it possible that the cognitive and emotional consequences of bullying are responsible for the psychotic manifestations that are associated with it?

Focus of the study:

- evaluate the link between
 - a history of being bullied
 - mood symptoms
 - psychotic symptoms (persecutory ideation, hallucinations)
- quantify potential intervention effects on **persecutory ideation**

Underlying assumption

- Interactional model of symptoms

A case study in Psychosis - the data

Data from the English National Survey of Psychiatric Morbidity, 2007 and 2000

Psychological questionnaire

- symptomatic and experiential variables
- cross sectional
- 8580 subjects in 2000 survey
- 9 selected variables

- Social variable: a history of bullying
- Psychological/behavioural variables
 - Persecutory thinking
 - Auditory hallucinations
 - Mood instability
 - Depression
 - Anxiety
 - Worry
 - Sleep problems
 - Cannabis use (physical effect on emergence of psychotic symptoms?)

Binary data - BDe score

Bayesian Dirichlet equivalent (BDe) score
Heckerman and Geiger, UAI 1995

Score equivalence and **score modularity**

Binary case for DAG G

- node X with m parents Y
- each state of Y has parameter θ_Y

$$P(X = 1|Y) = \theta_Y$$

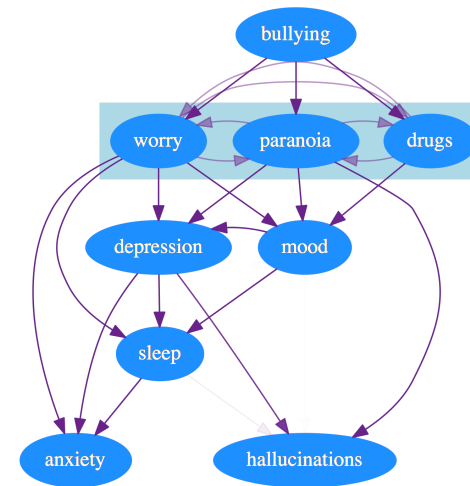
- beta prior on θ with hyperparameter

$$\alpha = \beta = \frac{\chi}{2^m}$$

BDe metric is marginal likelihood $P(D|G)$

BDe score is posterior $P(G|D)$

Partition MCMC to sample $\sim P(G|D)$

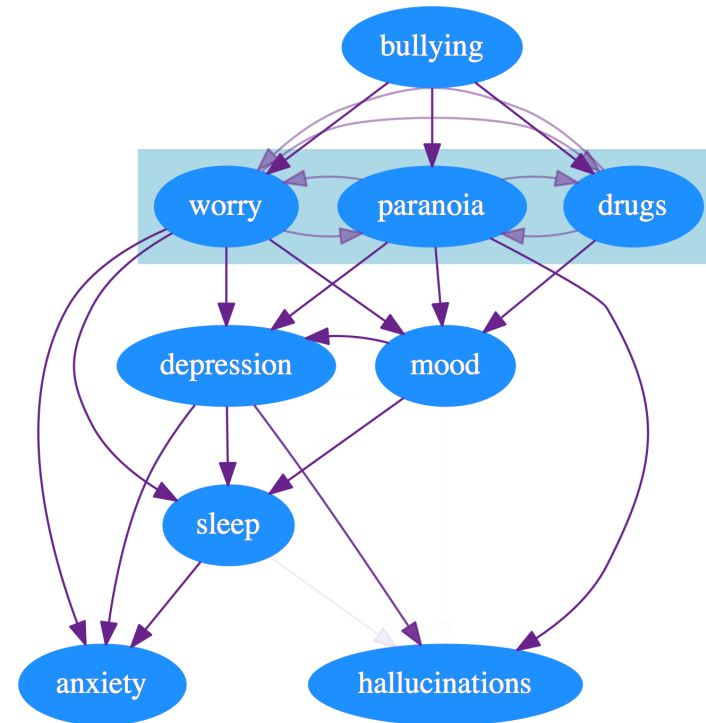


sample of 50,000 DAGs

Bullying - a case study with Bayesian networks

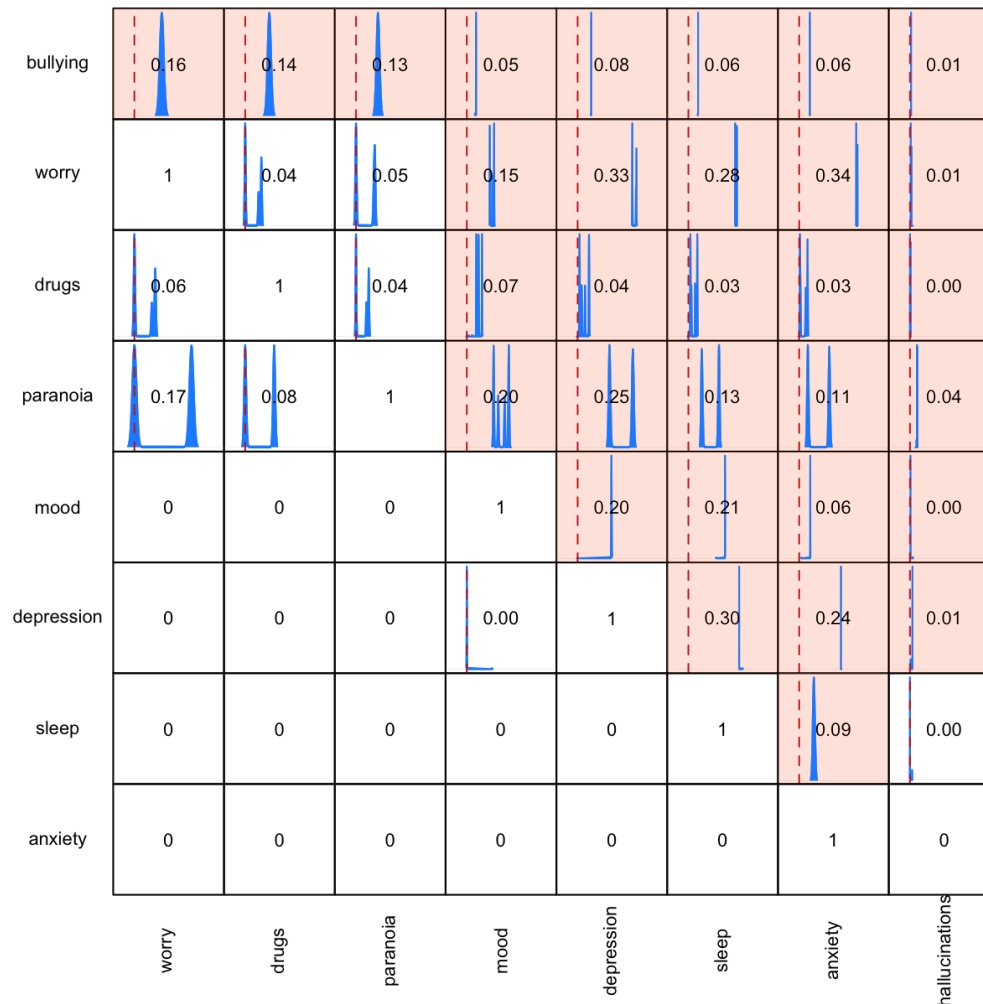
Quantify potential intervention effects on **persecutory ideation**

- Social variable **history** of **bullying** assumed antecedent
- **Interactional model** of symptoms explored by means of Bayesian networks (represented by DAGs)
- double arrows imply equivalence classes
- colour intensity reflects the strength of the links
- For each graph and each variable derive potential intervention effect on downstream nodes ($do(1) - do(0)$)



sample of 50,000 DAGs

Intervention effects: from a DAG ensemble from the posterior



- posterior distribution of causal effects of row label on column label (downstream only)
- 0 indicates no effect
- truncated to (-.1, .5) for clarity
- red line → zero causal effect
- box coloured if 95% credible interval does not straddle the zero line
- numerical values → posterior mean of the causal effect

Moffa et al, Schiz Bull 2017

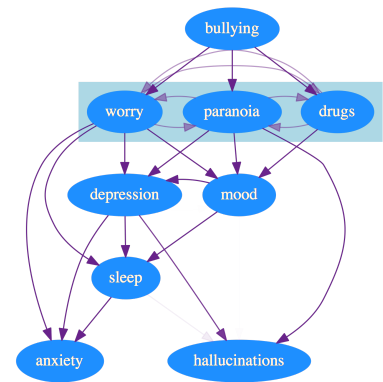
Psychological significance of findings

- Many hypothesised mediators did not meet the criteria for mediation: depression, anxiety, sleep disturbance, and hallucinations
- Links between worry, mood instability and persecutory ideation could not be disambiguated, cannot be resolved from these data, hence no evidence that bullying leads to paranoia by disturbing the mood
- In addition to highlight plausible causal links, the method also allows us to estimate the distributions of potential intervention effects
- There were studies underway involving attempts to alleviate persecutory ideation by reducing sleep disturbance and modifying depressive cognitions. Based on the present analysis they may prove unsuccessful (still not aware about results)

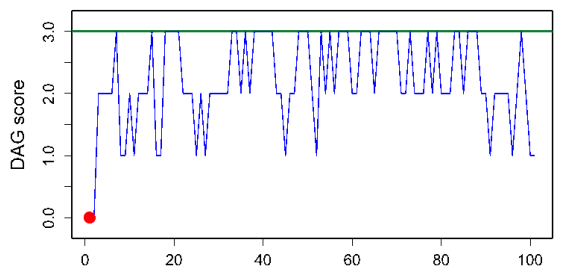
Significant limitation for psychology data

Inability to model feedback loops, partly addressed by Dynamic Bayesian networks

Thank you



bullying	0.16	0.14	0.13	0.05	0.08	0.06	0.06	0.01
worry	1	0.04	0.05	0.15	0.33	0.28	0.34	0.01
drugs	0.06	1	0.04	0.07	0.04	0.03	0.03	0.00
paranoia	0.17	0.08	1	0.20	0.25	0.13	0.11	0.04
mood	0	0	0	1	0.20	0.21	0.06	0.00
depression	0	0	0	0.00	1	0.30	0.24	0.01
sleep	0	0	0	0	0	1	0.09	0.00
anxiety	0	0	0	0	0	0	1	0
hallucinations								



Essential references

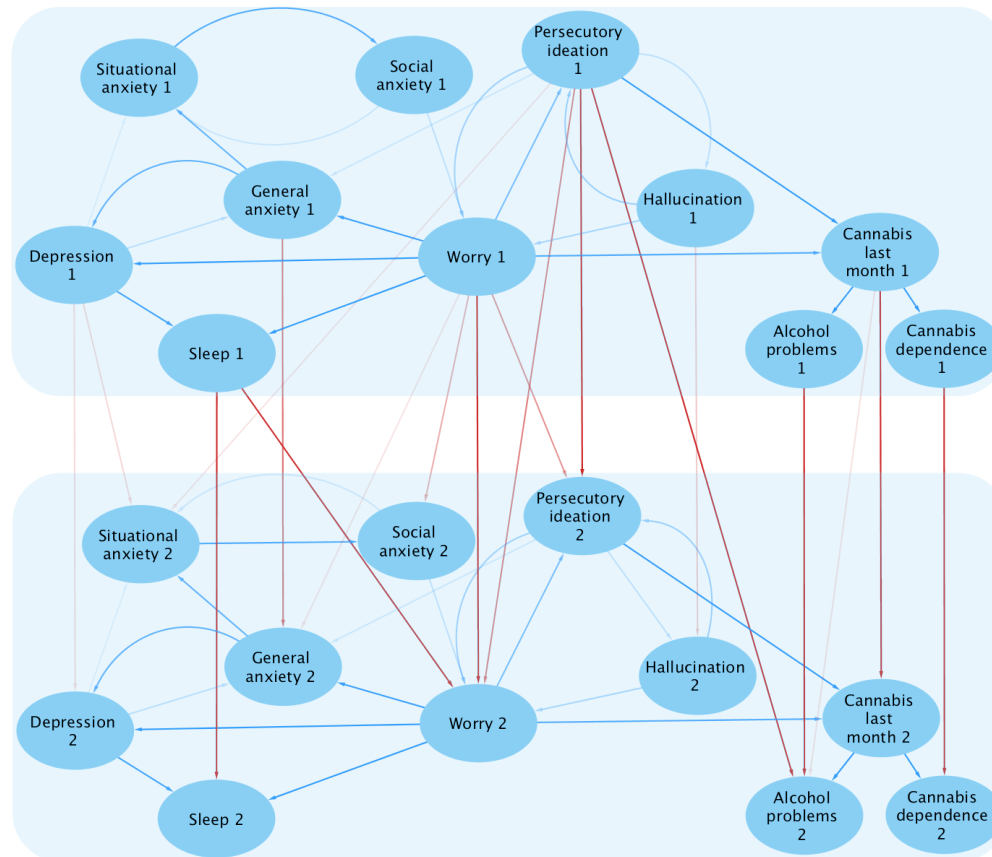
- Moffa, Giusi, et al. "Using directed acyclic graphs in epidemiological research in psychosis: an analysis of the role of bullying in psychosis." *Schizophrenia Bulletin* 43.6 (2017): 1273-1279.
- Jack Kuipers*, Giusi Moffa*, Elizabeth Kuipers, Daniel Freeman and Paul Bebbington. "Links between psychotic and neurotic symptoms in the general population: an analysis of longitudinal British National Survey data using Directed Acyclic Graphs." *Psychological Medicine* (2018): 1-8.
- Kuipers, Jack, and Giusi Moffa. "Partition MCMC for inference on acyclic digraphs." *Journal of the American Statistical Association* 112.517 (2017): 282-299.
- Kuipers, Jack, Polina Suter, and Giusi Moffa. "Efficient Structure Learning and Sampling of Bayesian Networks." *arXiv preprint arXiv:1803.07859* (2018).
- Kuipers, Jack, Thomas Thurnherr, Giusi Moffa, et al. "Mutational interactions define novel cancer subgroups." *Nature Communications* 9, (2018).
- Dawid, A. Philip. "Beware of the DAG!." *Causality: Objectives and Assessment*. 2010.
- Pearl, Judea. *Causality*. Cambridge university press, 2009.

Companion slides

Dynamic Bayesian network - graph

2000 British National Psychiatric Morbidity survey and its 18-month follow-up data (N=2406)

- one node for each variable at each time slice
- assume stationarity over time
- edges only displayed if they appear in at least 10% of the sampled DAGs



sample of 10,000 DAGs

Considerations about the psychological significance

- Worry appears to have a central role in the links between symptoms;
 - with plausible direct effects on insomnia, depressed mood and generalised anxiety.
- The relationship between persecutory ideation and worry is indeterminate, consistent with cross-sectional analysis
- Not all variables appear self-predicting of their state at the second time point
 - interestingly these are made up of affective variables (depression, social anxiety, and situational anxiety): a possibility is that they fluctuate significantly over the 18 months of follow up
 - general anxiety, worry, sleep problems, and persecutory ideation are strongly selfpredicting, suggesting they tend to persist over the follow-up period
- The relationship over the 18-month follow-up period between persecutory ideation and worry is suggestive of a putative feedback loop