Partial Correlation Networks and DAGs: The causal skeleton in the closet

Oisín Ryan[1], Laura Bringmann[2], Noémi Schuurman[1]

[1] *Department of Methodology and Statistics, Utrecht University*
[2] *Psychometrics and Statistics Department, University of Groningen*

The Network approach as causal inference

- Network approaches to psychopathology are theories of causal systems
- Alternative theoretical framework than observed indicators all caused by a latent disorder
- Disorders are themselves networks made up of symptoms and the causal relations between them

Pairwise Markov Random Fields

Researchers with cross-sectional data often estimate *Pairwise Markov Random Fields* (PMRFs)

- Partial Correlation Networks, Gaussian Graphical Models (GGMs), Mixed Graphical Models (MGMs), the Ising model
- Undirected (weighted) networks
- An edge is present between a pair of nodes if there is dependence between these two nodes **conditional on the set of all other nodes in the network**
- 23 applied papers in psychology according to Haslbeck & Fried (2017)

Why PMRFs?

PMRFs are often suggested for two reasons

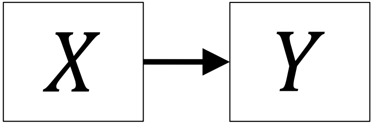
1. Sparse way of representing multivariate densities
2. More "causally meaningful" than full correlation networks
   - Argument usually that correlation networks show spurious associations, so partial correlations are more insightful for causal structure

A formal framework for causal inference

Intervenionist approaches to causal inference define causal effects in terms of outcomes of hypothetical experiments (e.g. Rubin, 1974; Pearl 1988)

Directed Acyclical Graphs (DAGs) can be used to represent causal structure graphically (Pearl, 1995, Spirtes, Glymour, Scheines, 1993)

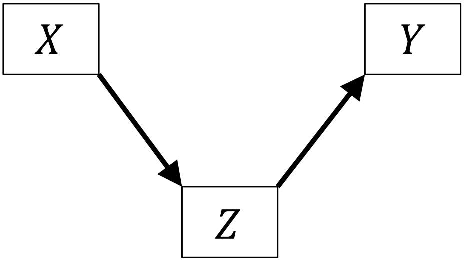
- Parents (X) cause children (Y)
- Children of Y are *descendants* of X
- Cannot be both a descendant and a parent of the same variable (no loops)
- Describes multivariate conditional dependency: X is conditionally independent of non-descendants, given the parents of X



DAGs and marginal/conditional dependencies

Mediation

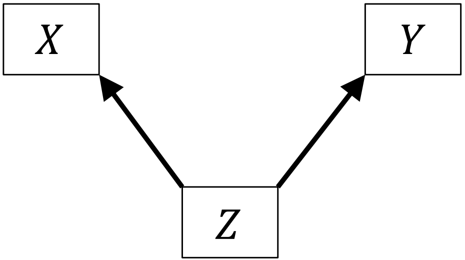
- Full correlation between X and Y is non-zero
  - X and Y are marginally dependent ($X \not\perp\!\!\!\perp Y$)
- Partial correlation between X and Y conditioning on Z is zero
  - X and Y are independent when we condition on Z, ($X \perp\!\!\!\perp Y \mid Z$)



DAGs and marginal/conditional dependencies

Common Cause

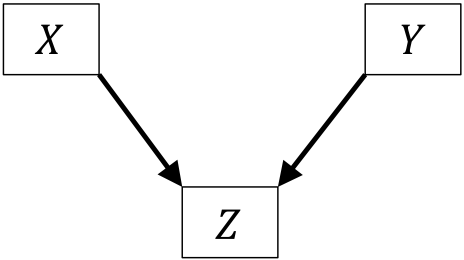
- Full correlation between X and Y is non-zero
  - X and Y are marginally dependent ($X \not\perp\!\!\!\perp Y$)
- Partial correlation between X and Y conditioning on Z is zero
  - X and Y are independent when we condition on Z, ($X \perp\!\!\!\perp Y \mid Z$)



DAGs and marginal/conditional dependencies

Collider / Common Effect

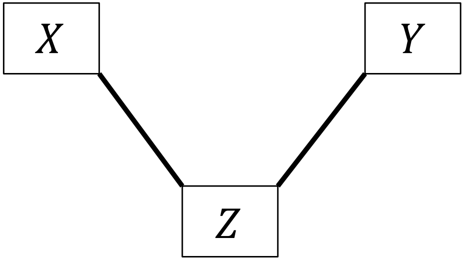
- Full correlation between X and Y is zero
  - X and Y are marginally independent ($X \perp\!\!\!\perp Y$)
- Partial correlation between X and Y conditioning on Z is non-zero
  - X and Y are dependent when we condition on Z, ($X \not\perp\!\!\!\perp Y \mid Z$)



DAGs and PMRFs

Example (Epskamp, Waldorp, Mottus & Borsboom, 2016)

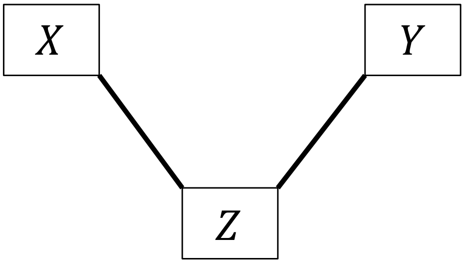
Partial Correlation Network



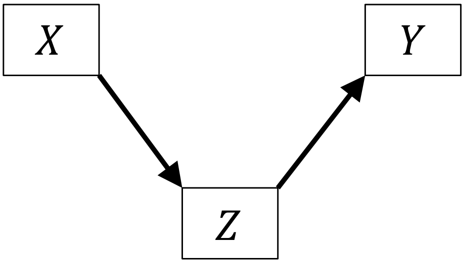
DAGs and PMRFs

Example (Epskamp, Waldorp, Mottus & Borsboom, 2016)

Partial Correlation Network



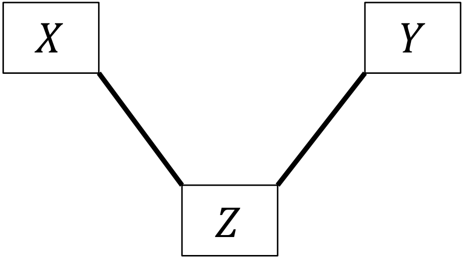
Possible True DAG 1



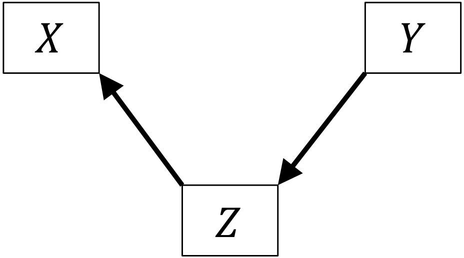
DAGs and PMRFs

Example (Epskamp, Waldorp, Mottus & Borsboom, 2016)

Partial Correlation Network



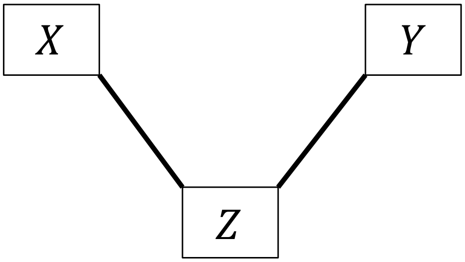
Possible True DAG 2



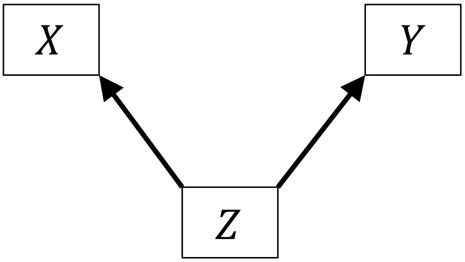
DAGs and PMRFs

Example (Epskamp, Waldorp, Mottus & Borsboom, 2016)

Partial Correlation Network



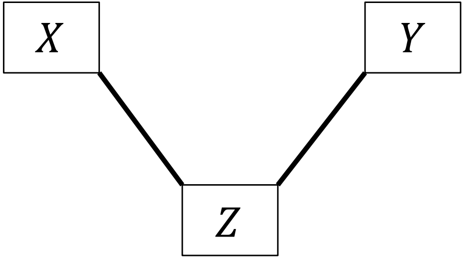
Possible True DAG 3



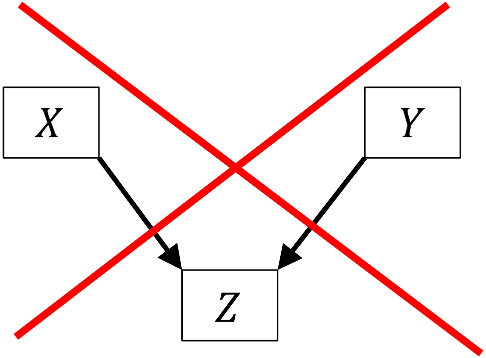
DAGs and PMRFs

Example (Epskamp, Waldorp, Mottus & Borsboom, 2016)

Partial Correlation Network



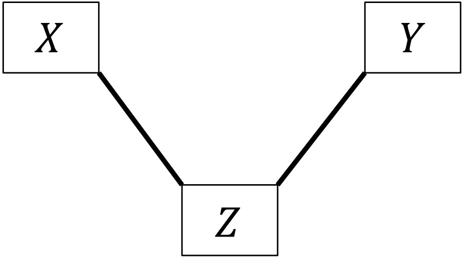
Impossible DAG 1



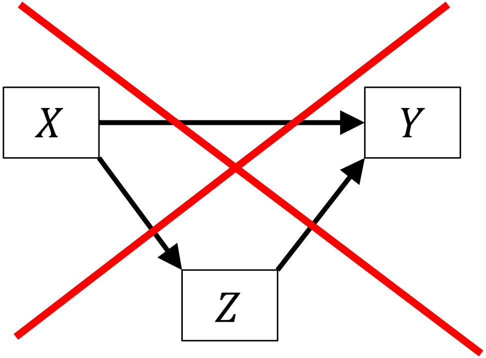
DAGs and PMRFs

Example (Epskamp, Waldorp, Mottus & Borsboom, 2016)

Partial Correlation Network



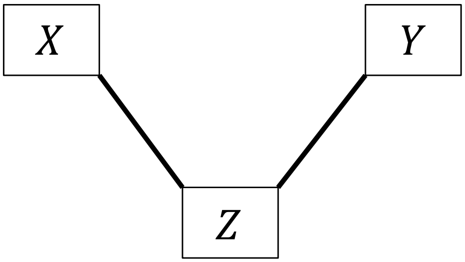
Impossible DAG 2



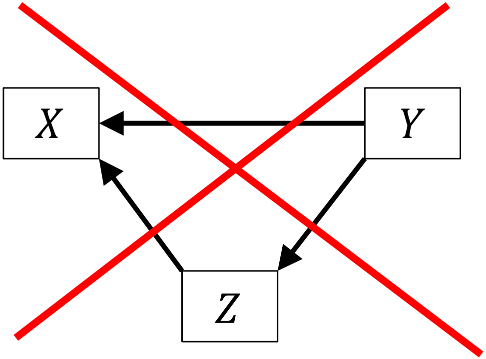
DAGs and PMRFs

Example (Epskamp, Waldorp, Mottus & Borsboom, 2016)

Partial Correlation Network



Impossible DAG 3



DAGs and PMRFs

In this specific instance the PMRF can be considered a good estimate of the *skeleton* of the underlying DAG

- The *skeleton* is the undirected graph found by ignoring the directions of the edges in the DAG (Glymour and Scheines, 2000)

DAGs and PMRFs

In this specific instance the PMRF can be considered a good estimate of the *skeleton* of the underlying DAG

- The *skeleton* is the undirected graph found by ignoring the directions of the edges in the DAG (Glymour and Scheines, 2000)

Unfortunately this property does not scale up well

- In larger networks PMRFs will induce edges due to conditioning on colliders

DAGs and PMRFs

In this specific instance the PMRF can be considered a good estimate of the *skeleton* of the underlying DAG

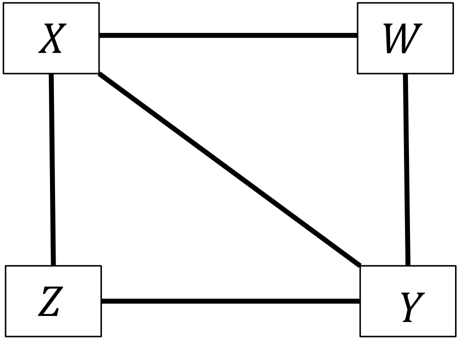
- The *skeleton* is the undirected graph found by ignoring the directions of the edges in the DAG (Glymour and Scheines, 2000)

Unfortunately this property does not scale up well

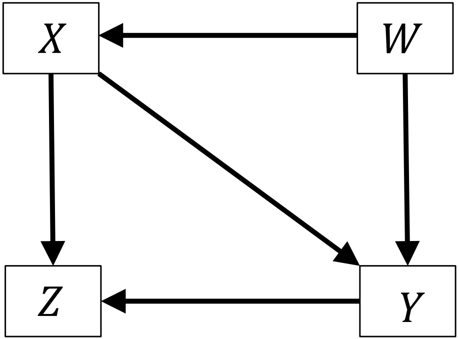
- In larger networks PMRFs will induce edges due to conditioning on colliders
- PMRFs estimate the *moral graph* - the undirected graph found by marrying all unmarried parents in the DAG, and then ignoring all directions
- In the general case this will be denser than and imply more possible true DAGs (a larger *equivalence class*) than the skeleton

DAGs and PMRFs

Partial Correlation Network

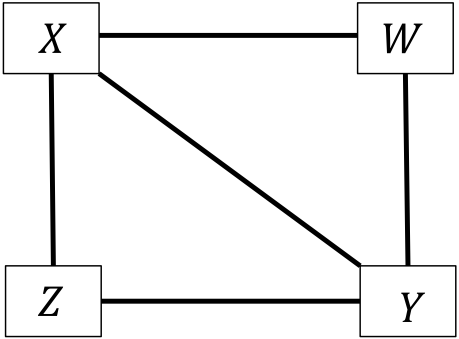


Possible True DAG 1

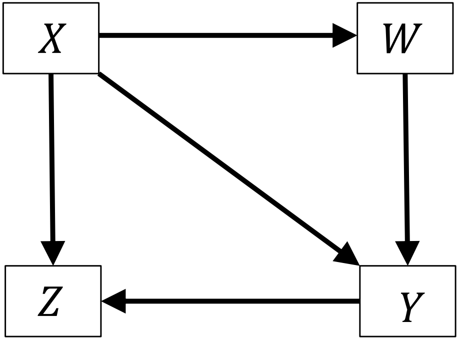


DAGs and PMRFs

Partial Correlation Network

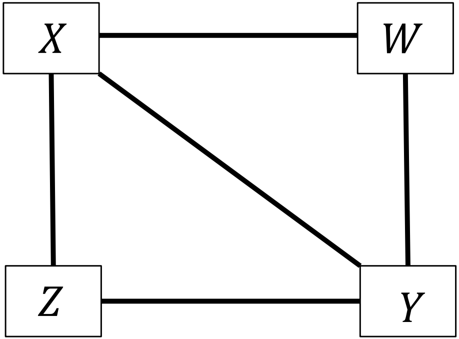


Possible True DAG 2

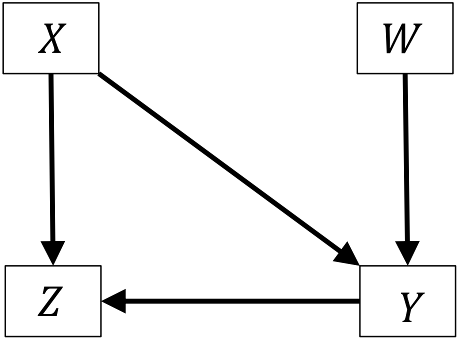


DAGs and PMRFs

Partial Correlation Network

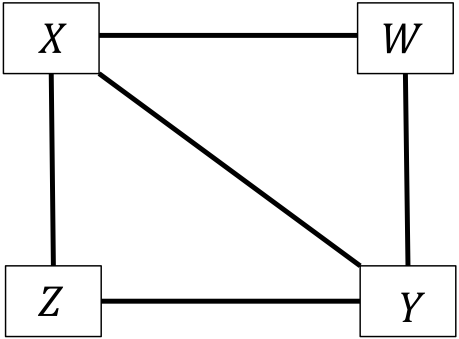


Possible True DAG 3

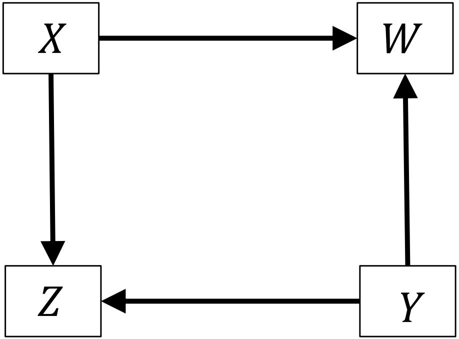


DAGs and PMRFs

Partial Correlation Network



Possible True DAG 4



PMRFs as DAG Skeletons

Problem: Researchers interpret PMRFs as though they are DAG skeletons

A quick review of 28 cross-sectional PMRF papers showed

- 4 papers in which the PMRF is referred to as the “causal skeleton”
- 10 papers which contain an interpretation of the PMRF as an estimate of the DAG skeleton

PMRFs as DAG Skeletons

Problem: Researchers interpret PMRFs as though they are DAG skeletons

A quick review of 28 cross-sectional PMRF papers showed

- 4 papers in which the PMRF is referred to as the "causal skeleton"
- 10 papers which contain an interpretation of the PMRF as an estimate of the DAG skeleton
    - Interpretation of subgraphs A-B-C as reflecting mediation
    - Statements that the graph only lacks direction of causal effects
    - If there is an edge in the PMRF, its likely theres an edge in the corresponding location in the DAG

Quantifying the problem

Given a multivariate gaussian density with an underlying DAG structure, how well can the skeleton of the DAG be retrieved by

- PMRF1: GGM with non-significant parameters set to zero
  - Partial Correlation matrix with a certain alpha level
- PMRF2: GGM using regularization
  - EBICglasso from qgraph (Epskamp et al 2012)
- Full correlation matrix, non-significant parameters set to zero
- The PC algorithm
  - DAG search algorithm, using pcalg (KalisCh et al, 2012)

Causal search: the PC algorithm

A method for finding DAGs from (observational) data

0. Start with a fully connected undirected network
1. Delete edges between nodes which are marginally independent.
2. In the graph created by the previous step, test each pair of neighbours for conditional dependence given each q-node subset of their adjacency set, with q=1. If conditionally independent, delete the edge between them.
3. Repeat step 2, increasing the value of q (each 2 node subset, new graph, then each 3 node subset, new graph). Repeat until the adjacency set is smaller than q.

The result of step 3 is an estimate of the DAG skeleton.

Simulation: DAG generation

For erdos-renyi:

1. Create an empty $n_{node} \times n_{node}$ weight matrix matrix
2. For each element in the lower triangle, draw from a bernoulli distribution with $p = P_{edge}$
3. If a 1 is drawn, assign that element in the weight matrix a value by drawing from a uniform(lb, ub).
4. Starting from the first row, draw N samples of each variable based on the weight matrix and a standard normal residual

For other graph-types:
Use randDAG and rmvDAG from pcalg, with topological.sort from igraph

Simulation conditions

Parameter			
Edge p	.1	.25	.5
Nodes	10	20	
N (sample size)	1000	5000	
(LB, UB)	(-1,1)	(0.1,1)	

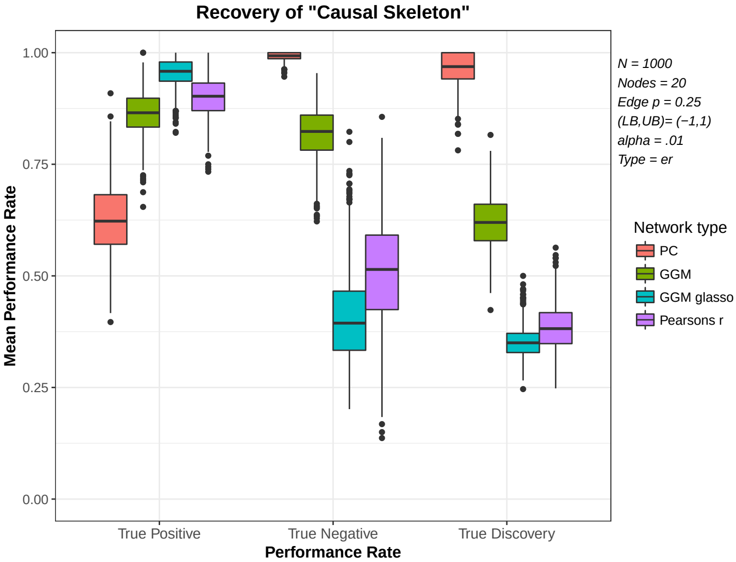
$\alpha = .01$
Graph type = erdos-renyi

- For limited conditions, results avialable for different graph types
  - power, bipartite, regular, watts
- Results also available for lower sample sizes (100, 500) and larger networks (Nodes=40,60)

Simulation: Measures

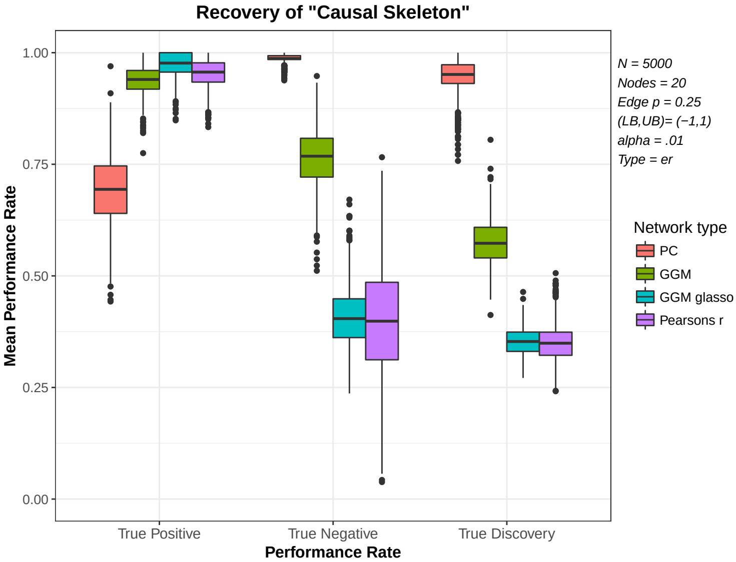
1. True Positive Rate
   - Sensitivity. Ratio of correctly estimated edges to number of true edges
2. True Negative Rate
   - Specificity. Ratio of correctly estimated gaps to number of true gaps
3. True Discovery Rate
   - Precision. Ratio of correctly estimated edges to total number of estimated edges

Results 1

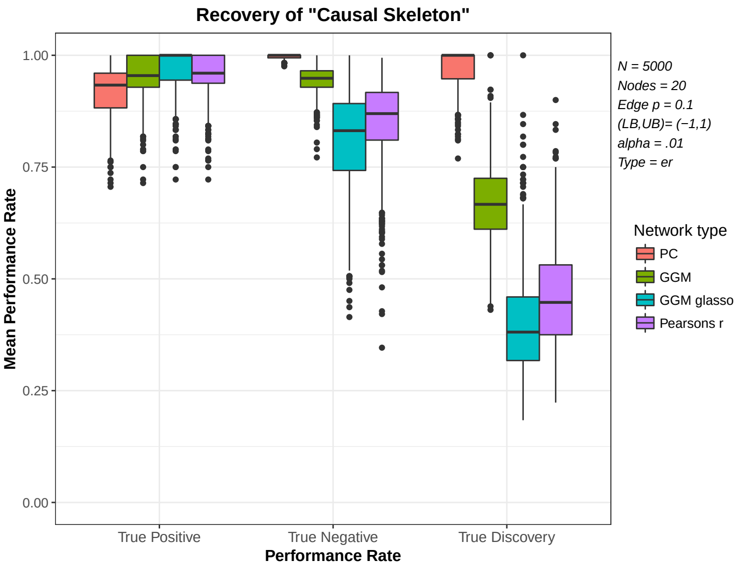


Recovery of "Causal Skeleton"
Mean Performance Rate
1.00
0.75
0.50
0.25
0.00
True Positive
True Negative
True Discovery
Performance Rate
N = 1000
Nodes = 20
Edge p = 0.25
(LB,UB)= (−1,1)
alpha = .01
Type = er
Network type
PC
GGM
GGM glasso
Pearsons r

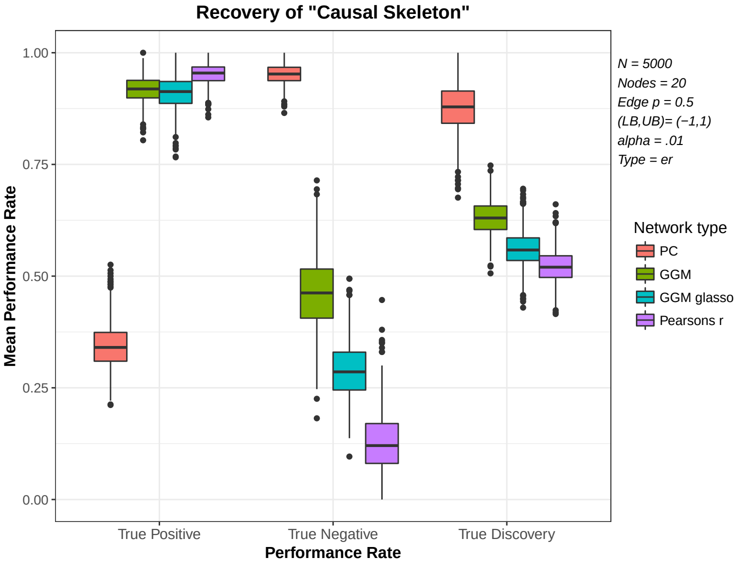
Results 1: Increase Sample Size



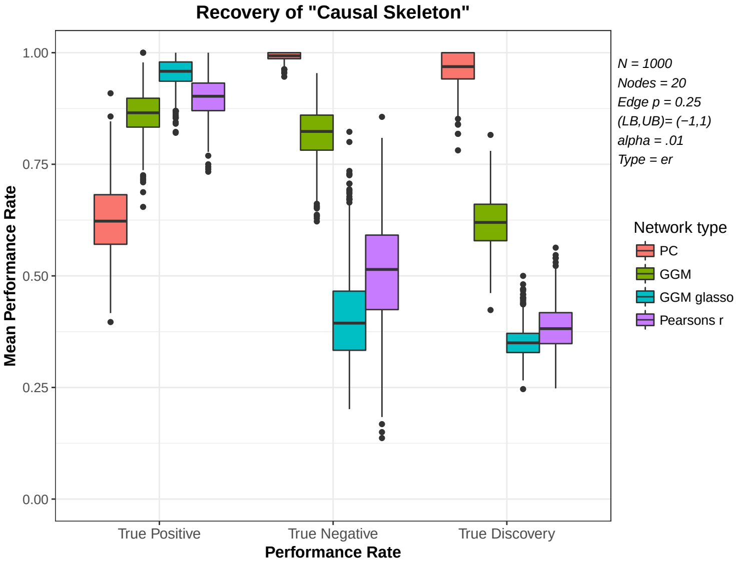
Results 1: Decrease edge probability



Results 1: Increase Edge Probability

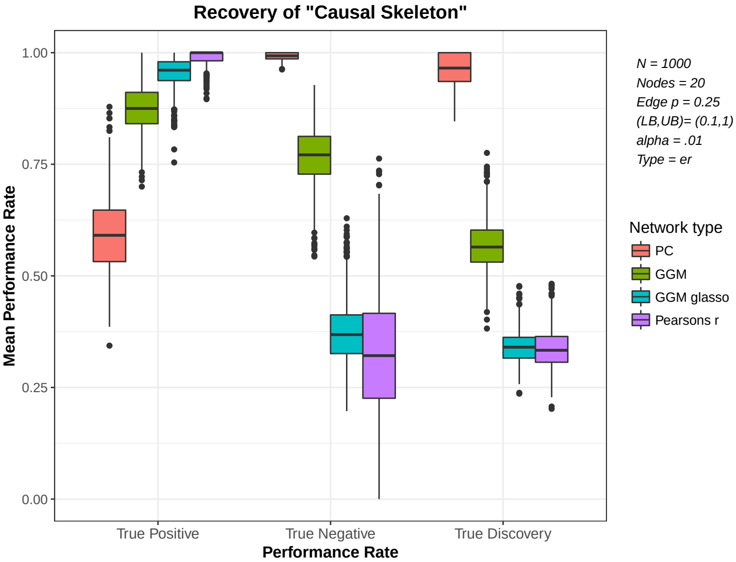


Results 1

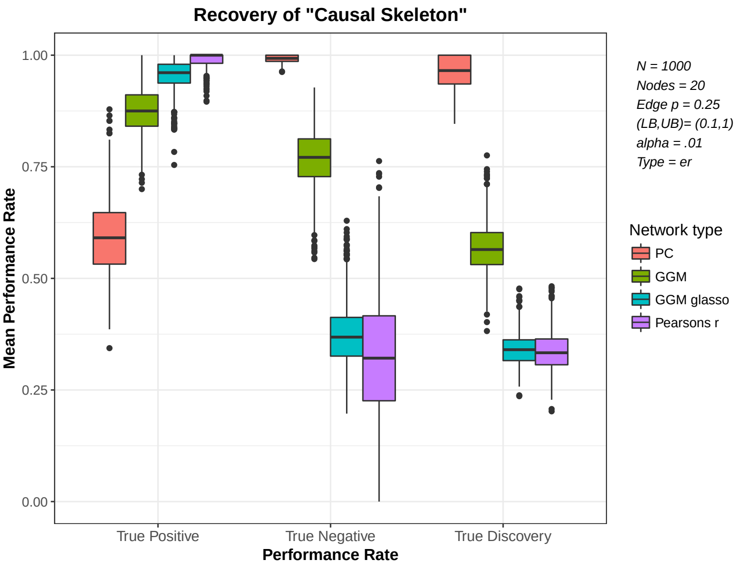


Recovery of "Causal Skeleton"
Mean Performance Rate
1.00
0.75
0.50
0.25
0.00
True Positive
True Negative
True Discovery
Performance Rate
N = 1000
Nodes = 20
Edge p = 0.25
(LB,UB)= (−1,1)
alpha = .01
Type = er
Network type
PC
GGM
GGM glasso
Pearsons r

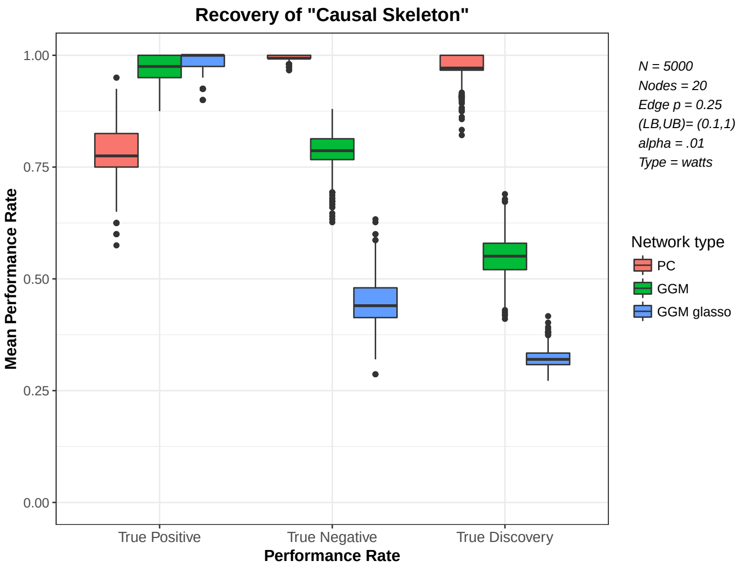
Results 1: Positive Manifold



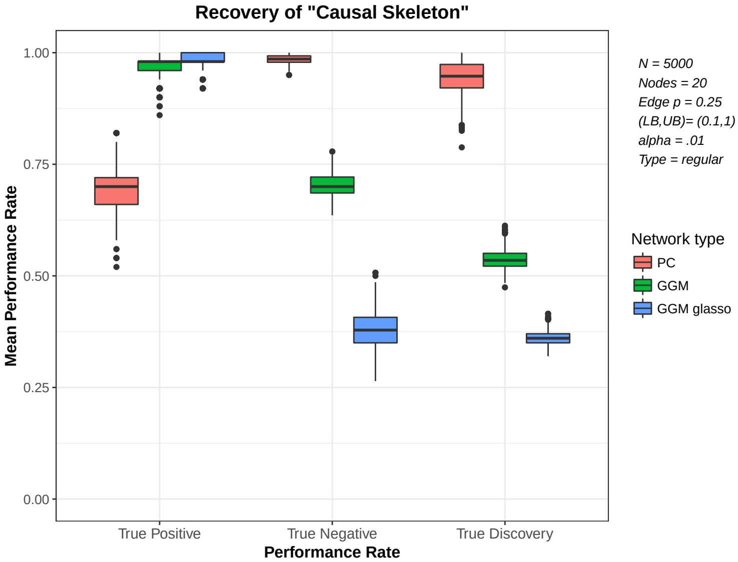
Results 1: Positive Manifold



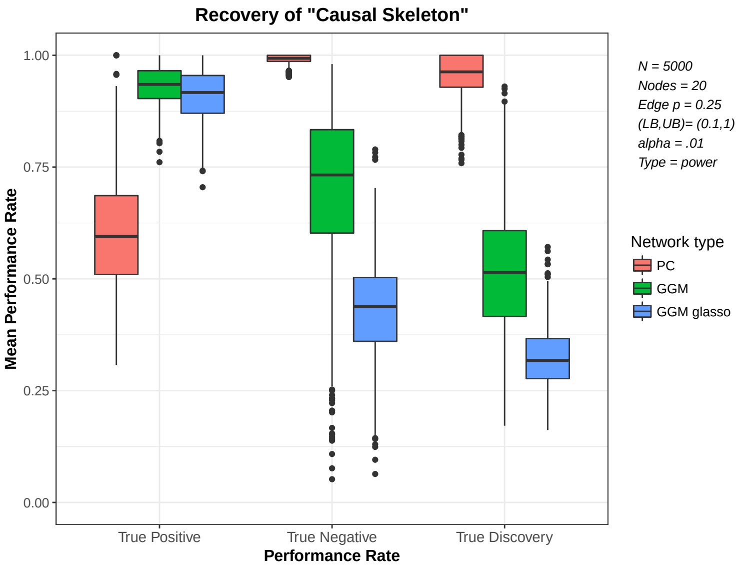
Results 1: Positive Manifold Graph types



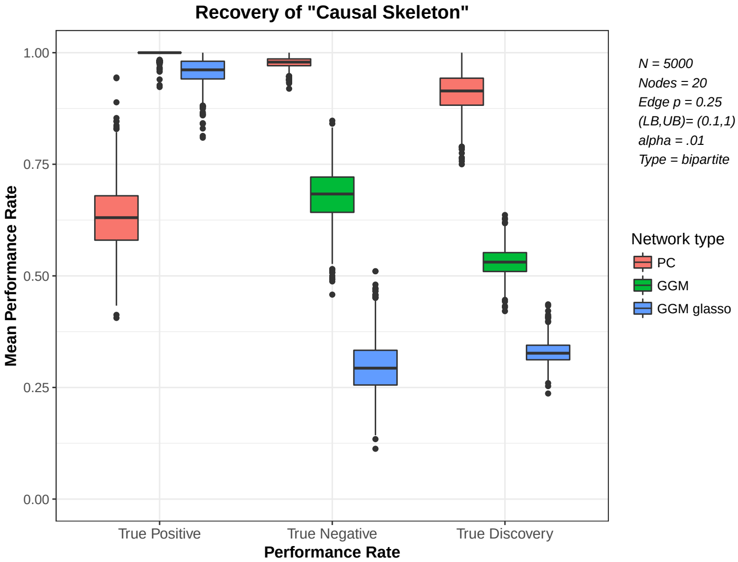
Results 1: Positive Manifold Graph types



Results 1: Positive Manifold Graph types



Results 1: Positive Manifold Graph types



Summary

- An edge being present in a PMRF with regularizations can mean there is only a 32.5 percent chance there is an actual edge in the underlying DAG
- Regularization techniques do not help with this kind of “false positive” - comparable to full correlation networks

Discussion

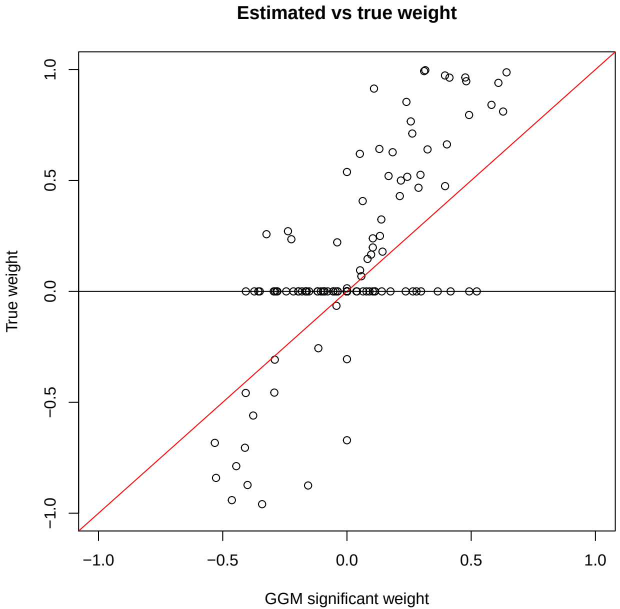
1. If want to investigate causal structures using DAGs and associated theory:
   - Perhaps we shouldn't be using PMRFs
   - Increasingly many DAG search algorithms (beyond the PC algorithm) exist for this
   - Directions rarely fully identified, but should imply less possible DAGs (smaller equivalence class) than PMRFs

Discussion

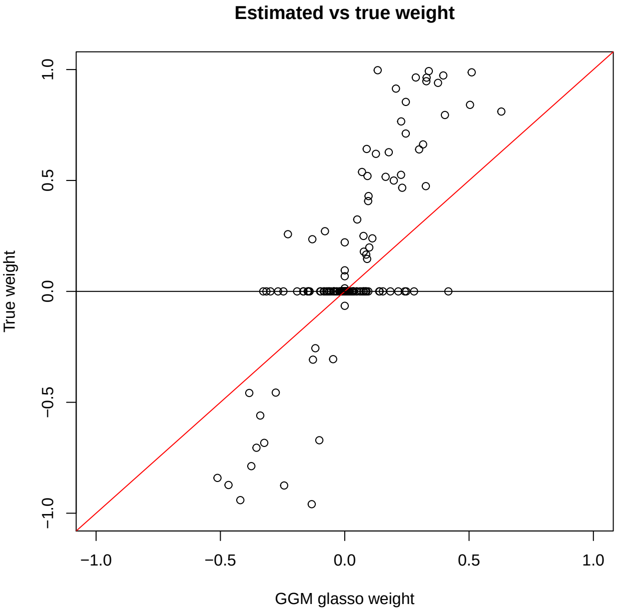
1. If want to investigate causal structures using DAGs and associated theory:
   - Perhaps we shouldn't be using PMRFs
   - Increasingly many DAG search algorithms (beyond the PC algorithm) exist for this
   - Directions rarely fully identified, but should imply less possible DAGs (smaller equivalence class) than PMRFs

2. If we don't think DAGs are appropriate:
   - Why?
   - What frameworks for causal inference do we use?
   - Are PMRFs the most appropriate tool for that framework? (e.g. Richardson 1996)

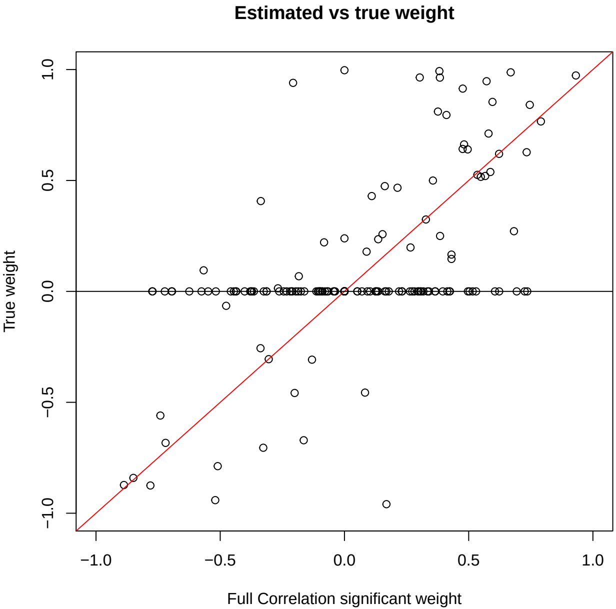
GGM edge weight example



GGM glasso edge weight example



Full correlation edge weight example



Centrality: Snapshot

Positive Manifold	Nodes	Edge P	Degree GGM glasso	Degree GGM	Degree Full correlation	Betweenness GGM glasso	Betweenness GGM	Betweenness Full correlation
Yes	10	.25	.616	.590	.493	.669	.727	.578
		.5	.367	.363	.243	.400	.397	.339
	20	.25	.295	.287	.151	.421	.383	.201
		.5	.052	.103	.078	.116	.148	.116
No	10	.25	.677	.650	.592	.718	.743	.638
		.5	.480	.461	.353	.566	.540	.426
	20	.25	.423	.393	.245	.453	.410	.294
		.5	.210	.245	.143	.239	.270	.140

Table: Proportion of estimates in which (one of) the node(s) with greatest estimated centrality corresponds the most central node in the weighted DAG skeleton. Alpha=.01, N_participants =5000

Quotes regarding partial correlations as causal

- "if one is interested in knowing which of these symptoms are truly related[...] then correlations may not provide optimal information [...] the matrix of partial correlations [...] may be considered to provide clues about the causal skeleton of a network (an undirected pattern of direct relations between variables)". (Borsboom & Cramer, 2013)
- "correlation networks generally show a high rate of spurious associations between nodes [...]. To gain more insights regarding the causal structure of the system, we, therefore, computed concentration networks in which edges represent partial correlations" Deserno et al. (2016)

Quotes regarding PMRF as DAG skeleton

- "Our empirical network... showed that 'sleep disturbances'.. was not directly connected to any of the depression-related problems... but was connected via worry... This may imply that sleep disturbances do not directly cause depression-like symptoms, but can lead to depression via the **mediating** role of excessive worry" Boschloo et al (2016b)
- "Whenever the partial correlation is exactly zero, no connection is drawn between two nodes, indicating that two variables are independent after controlling for all other variables in the network. ... such a missing connection indicates one of the two variables could not have caused the other (Pearl, 2000) As such, **whenever there is a connection present, it highlights a potential causal pathway** between two variables" Epskamp & Fried (2016)

Quotes regarding PMRF as DAG skeleton

- “number of social contacts does not **directly** relate to feeling happy, but **influences happiness through** social satisfaction” Deserbi et al (2016)
- “ due to the cross-sectional nature of the data, the estimated networks are undirected, and centrality estimates do not provide information whether a symptom mostly actively **triggers** other symptoms ... **or is triggered** by other nodes .” Fried et al (2016c)
- “the Markov random field network can be viewed as a causal skeleton that encodes the **existence but not the direction of putative causal relations** in the population.” van Borkulo et al (2015)