Node Centrality Measures are a poor substitute for Causal Inference

Fabian Dablander & Max Hinne

Department of Psychological Methods, University of Amsterdam

Background

Theoretical network models:

- Constitute a shift in how we view psychological constructs.
- Example: Depression does not cause symptoms such as worry, sadness, and lack of energy but rather arises out of their direct causal interactions [1].

Statistical network models:

- Majority of research uses Gaussian graphical model and cross-sectional data [2].
- Node centrality measures used to gauge importance of nodes; many researchers suggest that highly central nodes may constitute good intervention targets [2].

Observations:

- Mismatch between theoretical and statistical models.
- Intervention is a *causal* notion. Here [3] we ask: how well do node centrality measures map onto measures of causal influence?

Node-specific Measures

Node Centrality Measures

There exist a number of node centrality measures for weighted graphs [5]. We focus on *degree*, *strength*, *closeness*, *betweenness*, and *eigenvector-centrality*.

Causal Influence Measures

We use the following two causal influence measures [4]:

$$ACE(X_j \rightarrow X_i) := \mathbb{E}[X_i | do(X_j = x_j)] - \mathbb{E}[X_i | do(X_j = x_j + 1)] \tag{1}$$

$$CE_{KL}(X_j \rightarrow X_i) := KL(P||P_S) = \frac{1}{2}\left(\text{tr}\left[\Sigma_S^{-1}\Sigma\right] - \log\frac{\det\Sigma}{\det\Sigma_S} - n\right) \quad , \tag{2}$$

where we compute the total causal influence of node X_j by summing its causal influence on all of its children.

Simulation setup

For each network type (Figure 1), we randomly generate a DAG with number of nodes $v = \{10, 20, 30, 40, 50\}$ and network density $d = \{0.1, 0.2, \ldots, 0.9\}$. We sample $n = 5000$ observations, estimate a Gaussian graphical model, and compute (a) the rank-correlation between node centrality and causal influence and (b) the probability of correctly classifying the node with the highest causal influence based on centrality.

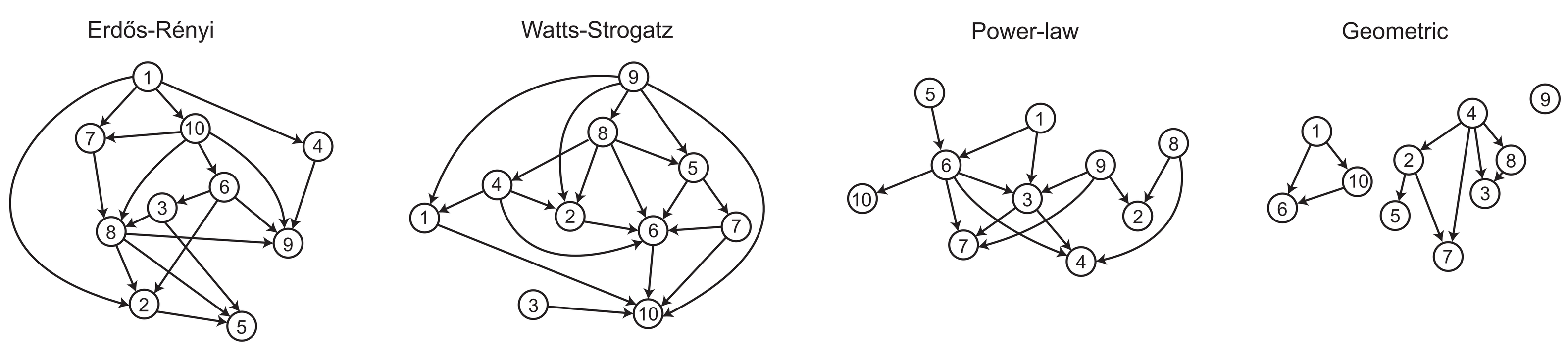

Figure 1: Shows examples of the generated directed acyclic graphs that are used as ground truth in the simulations.

Example

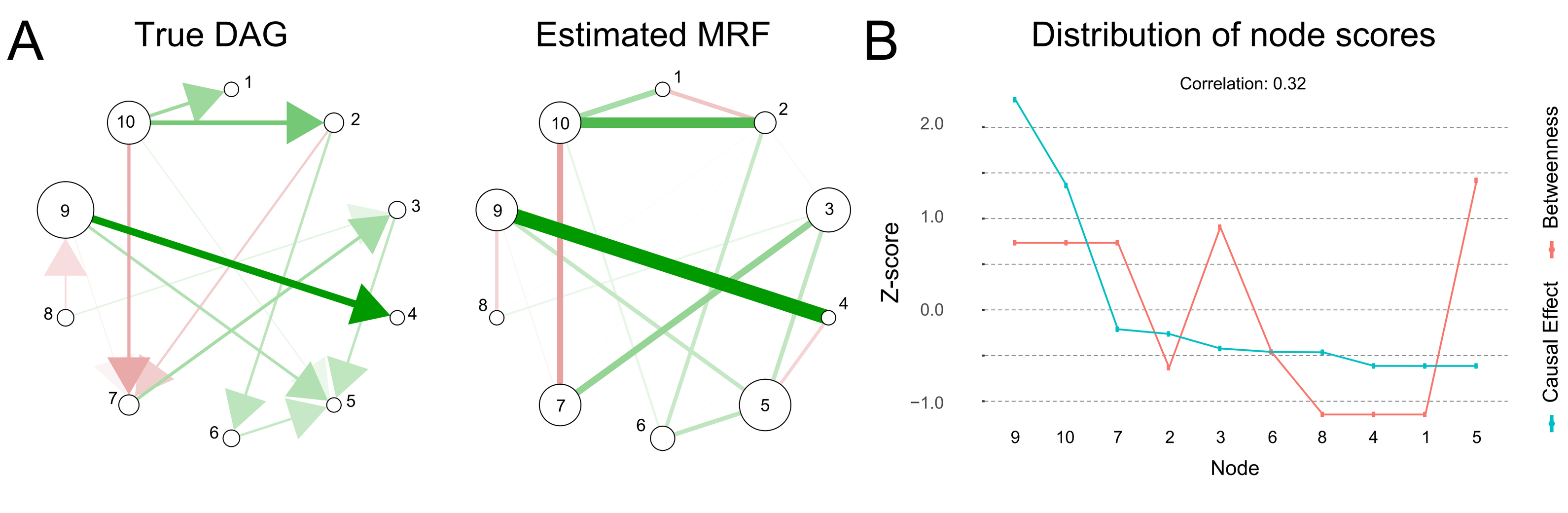

Figure 2: **A**. Shows the true DAG and the estimated undirected network for $v = 10$ and $d = 0.3$. **B**. Shows the distributions of CE_{KL} and betweenness centrality ($r_s = 0.32$). Interactive plot: https://fdabl.shinyapps.io/causality-centrality-app/.

Results

Figure 3 shows a homogeneous picture of correlations across graphs. The average rank-correlation is between 0.30 and 0.50, with 10% and 90% quantiles ranging from -0.10 to 0.70. Eigenvector-centrality is the exception, because in contrast to the other measures, its distribution over nodes does not become left-tailed.

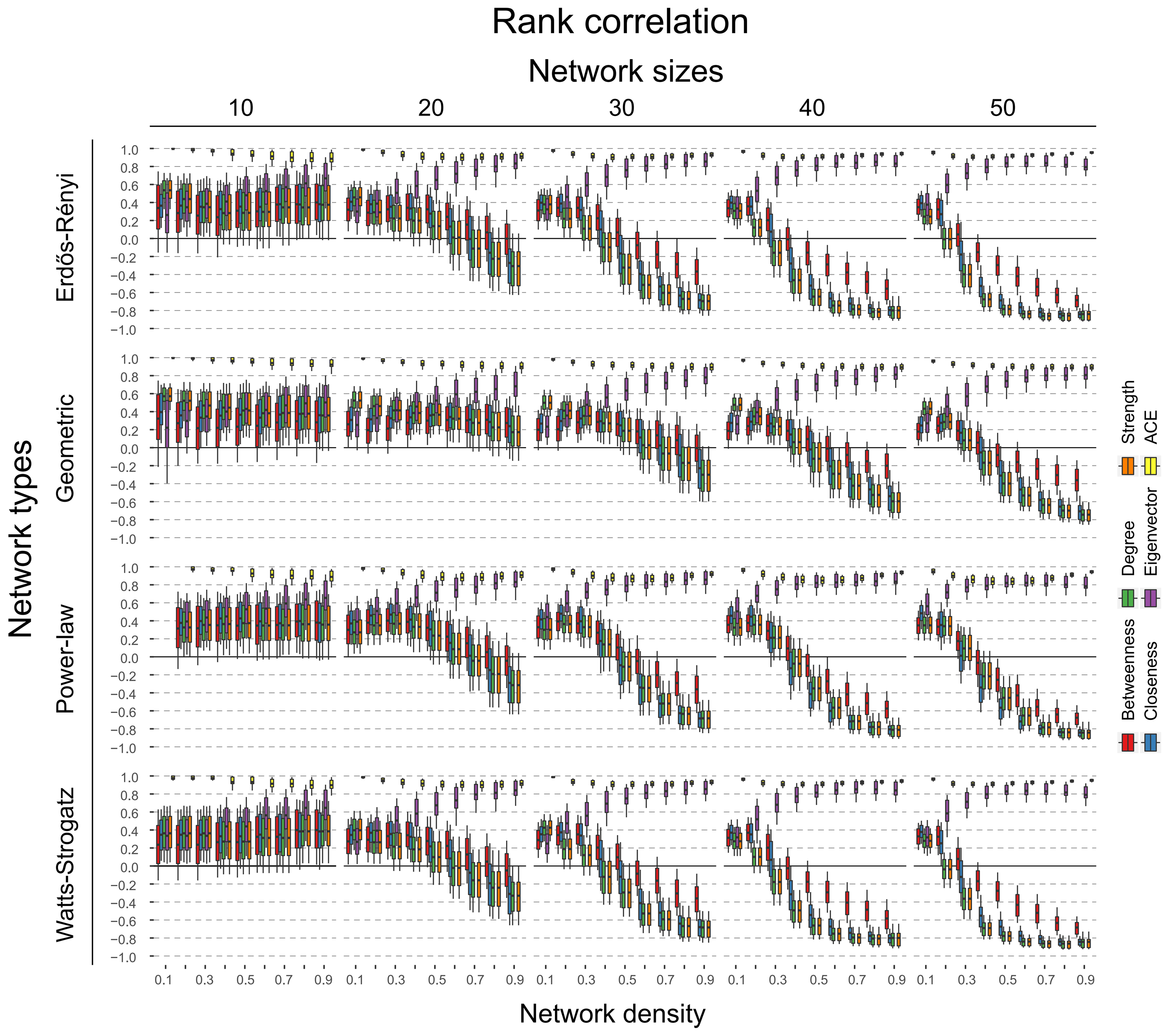

Figure 3: Shows rank-based correlation r_s of CE_{KL} with node centrality measures. Error bars denote 10% and 90% quantiles across the 500 simulations.

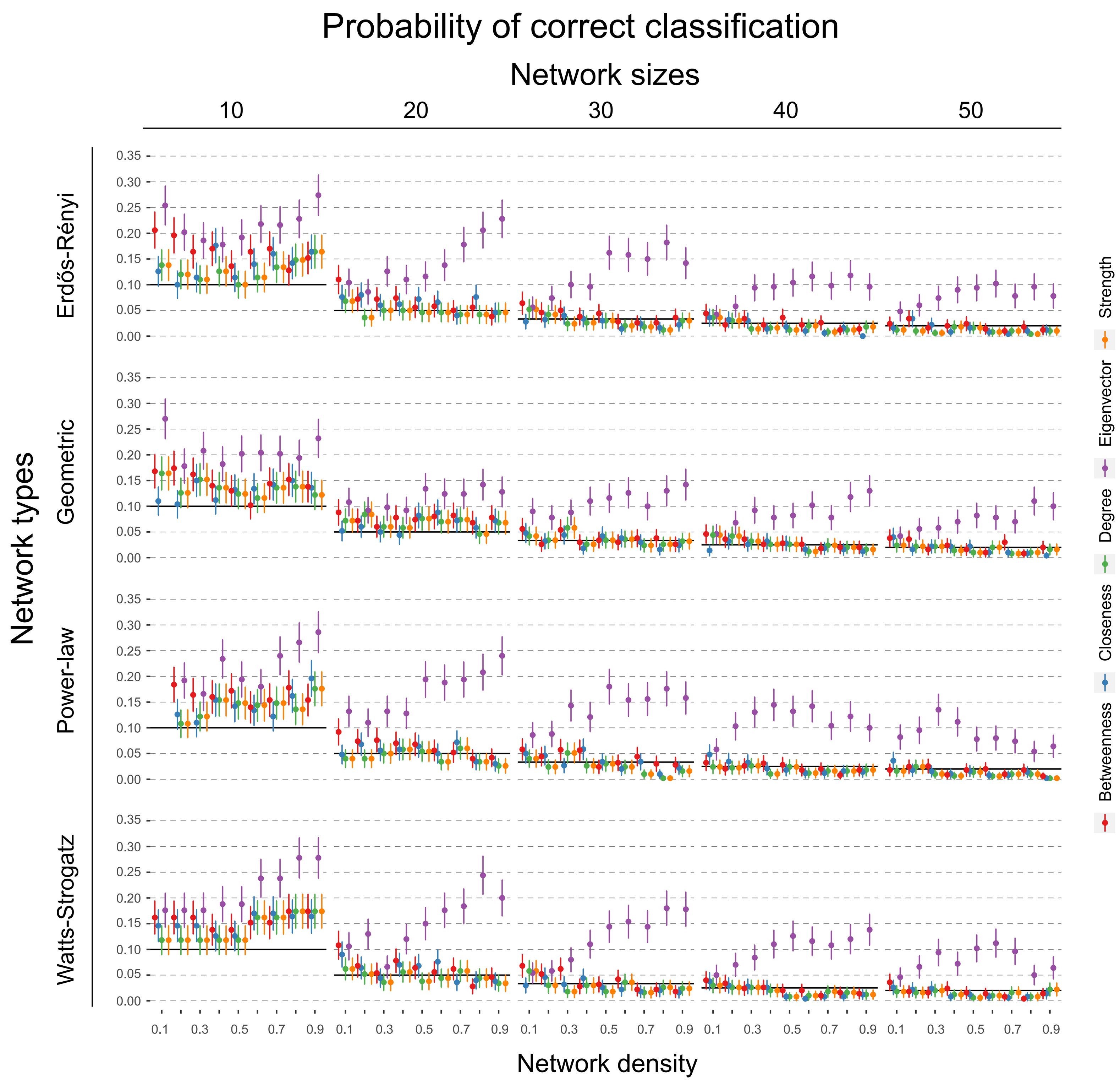

Figure 4: Shows probability of correctly classifying the node with highest CE_{KL} based on centrality measures. Error bars denote 95% confidence intervals across the 500 simulations.

Conclusion

- Overall weak relationship between node centrality and causal influence measures, except for eigenvector-centrality.
- If the underlying real-world system can be modeled as a DAG, but researchers interpret nodes with high centrality as causally important, then this may result in sub-optimal interventions.

References

[1] Denny Borsboom. A network theory of mental disorders. *World Psychiatry*, 16(1):5–13, 2017.

[2] Alba Contreras, Ines Nieto, Carmen Valiente, Regina Espinosa, and Carmelo Vazquez. The study of psychopathology from the network analysis perspective: A systematic review. *Psychotherapy and Psychosomatics*, 16(1):1–13, 2019.

[3] Fabian Dablander and Max Hinne. Node Centrality Measures are a poor substitute for Causal Inference. Preprint: 10.31234/osf.io/nue4z.

[4] Dominik Janzing, David Balduzzi, Moritz Grosse-Wentrup, and Bernhard Schölkopf. Quantifying causal influences. *Annals of Statistics*, 41(5):2324–2358, 2013.

[5] Tore Opsahl, Filip Agneessens, and John Skvoretz. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3):245–251, 2010.