Critical Reflection on Comparison of Gaussian Graphical Models (GGM) and Directed Cyclic Graph Models (DCG) as Causal Discovery Tools

Network Analysis 2022

Kyuri Park (12183881)

December 22, 2022

Word count: 998

Table of Contents

1   Summary of Final Project

In our final project, we aimed to investigate the utility of statistical network models as tools for causal discovery in *cyclic* settings compared to the *directed cyclic graph* models (DCG) by means of a simulation study. We used the *cyclic causal discovery* (CCD) algorithm to estimate DCGs (Richardson, 1996b) and compare them to the Gaussian graphical models (GGM). The comparison is made based on the overall density of models and degree centrality per each node in models. Note that the output of CCD is a PAG (*partial ancestral graph*)[1], which represents the *Markov-equivalent* class of DCGs (i.e., it encompasses multiple DCGs that are statistically equivalent under the found conditional independencies in the data). Therefore, we computed the *average* density and *average* degree centrality given all the equivalent DCGs per simulated condition and compared those to the density and degree centrality of GGM and true model, respectively.

The results found that the DCGs approximated the true cyclic models better in terms of both density and degree centrality compared to GGMs. GGMs often overestimated the density and correspondingly resulted in high degrees for almost all nodes in the considered models. The conclusion based on these results could be that statistical network models perform poorly as causal discovery tools in cyclic settings and hence, it shall be preferred to use the purpose-built *causal discovery methods* when one is interested in underlying causal mechanism.

Figure 1: Example partial ancestral graph.
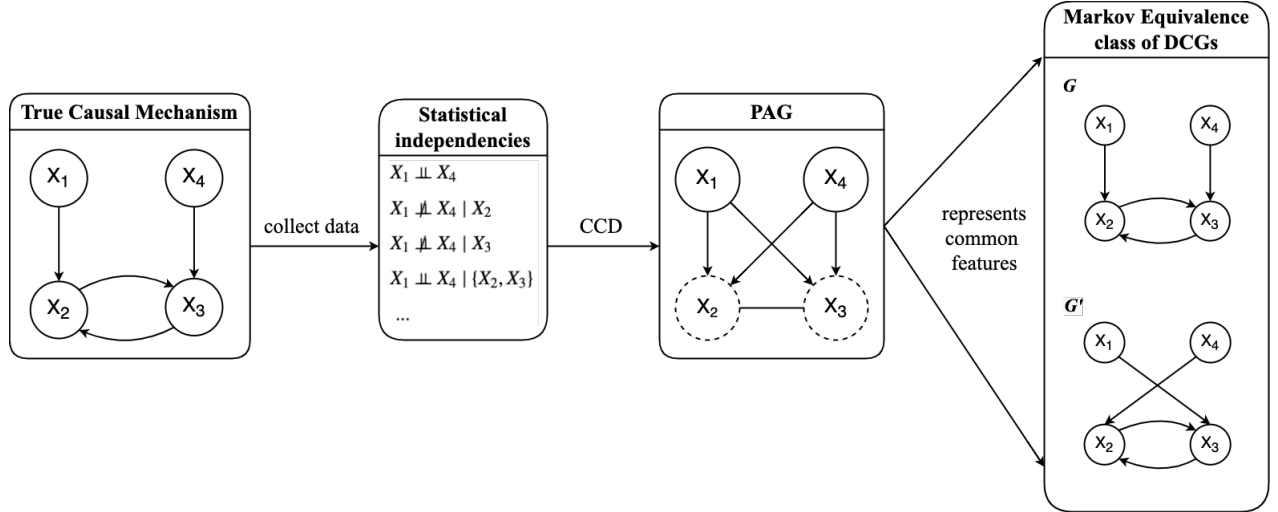
Note. Given the found statistical independencies in the data from the true causal mechanism, CCD outputs a partial ancestral graph (PAG). It can be seen that in this example, the PAG represents two different directed cyclic graphs (DCG), of which we compute the density and degree centrality. Then, the average value of density and degree centrality are subsequently used for the comparison with the GGM.

---

[1]See Figure 1.

2   Reflection on Limitations

However, we cannot just naively draw such conclusion merely based on these results, as there are several limitations in this simulation study. Below, I list two crucial limitations in my opinion.

1. We did not explicitly account for the *uncertainty* in estimation with CCD. As explained above, CCD in general cannot uniquely identify one true graph, but instead provides a set of equivalent DCGs (as a form of PAG). The *size* of the equivalent set reflects the level of uncertainty in its estimation, such that the larger the set is, the higher the uncertainty in estimation. Besides the size, the *variation* in the equivalent set can also imply the instability in estimation; the higher the variation, the more unstable the estimates are. It is a critical aspect to further investigate, as it could be the case that even when the average density and degree centrality are closely aligned to those of the true model, the *size* of the set or the *variation* in the set is large, indicating the lack of precision in its estimates.

2. There is little evidence on practical applicability of CCD, as we only tested it out on rather small simulated models ($p = 4, 5, 6$) with a very large sample size ($N = 10^6$). The typical psychological data, however, usually consist of much less observations while including more variables (Constantin et al., 2021). Thus, whether CCD can be applied to such empirical data and utilized in psychological research in practice is yet questionable.

3   Further Extensions

To examine the overall uncertainty/instability in estimation as to the first limitation, we check the size of equivalent class from each of the simulated models and investigate the variation in the density as well as in degree centrality per each set of DCGs. Regarding the practical applicability of CCD, we test the CCD algorithm on an empirical data (McNally et al., 2017) and check if it produces a reasonable output. Here, we focus on discussing the analysis results only on the first limitation. For the results on the applicability of CCD, we refer the interested readers to Appendix (section 6).

3.1   Variation in Density of DCGs

Figure 2 shows the size of the equivalence class and the distribution of density per set of DCGs from each of the simulated models. First thing that stands out is that the dense models (right column of Figure 2) seem to have a slightly higher variation in density (i.e., the spread of distributions is wider) and accordingly the discrepancy between the true and average density of DCGs is larger in the dense models than in the sparse models. This indicates that CCD tends to be less stable when the causal structure is dense and correspondingly, the resulting estimates are more likely to deviate from the true values.

Secondly, it could be seen that the size of equivalence class is overall quite large across different conditions, except for the *sparse model with 4 nodes* (Figure 2 (a)) case. There is no specific pattern observed between the

size of equivalence class and the types of simulated models, but these overall fairly large equivalence classes suggest the lack of certainty in estimation, which is commonly known as one of the general weaknesses of the causal discovery algorithms (Bongers et al., 2016).

**Figure 2:** Distribution of density per set of DCGs.
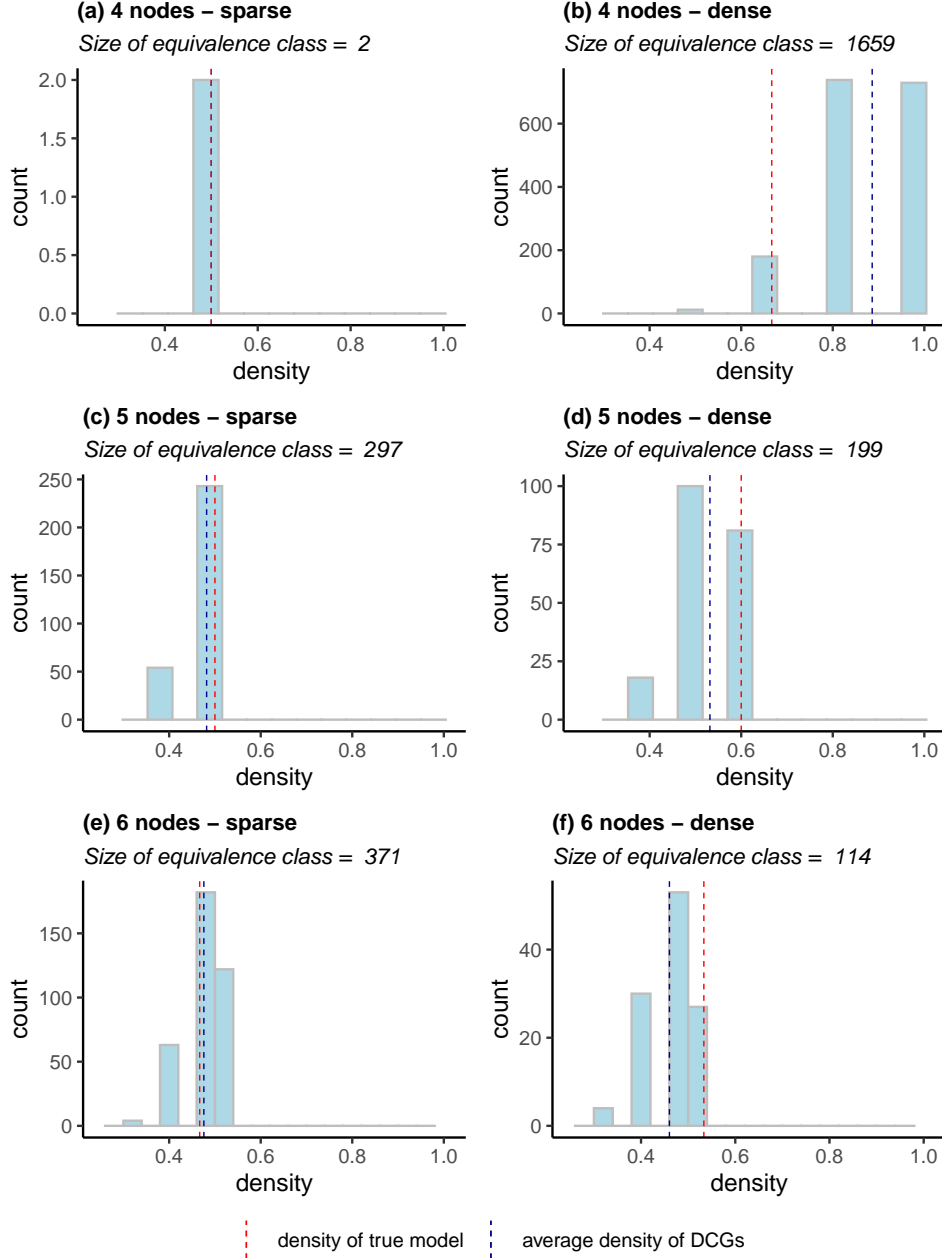
Note. The scale of x-axis is fixed the same across all conditions so that the comparison between distributions can be easily made. In *(a)*, both DCGs in the equivalence class have the same density that is identical to the true density (the two lines overlap). The dense model with 4 nodes in *(b)* seems to be the most difficult case for CCD, given that it has the largest equivalence class (1,659 DCGs) with the highest discrepancy in density among all considered cases. Typically, CCD seems to struggle more to estimate the dense models, given that there are observed more variations with higher deviations from the true values in the dense condition.

3.2 Variation in Degree Centrality of DCGs

Figure 3 shows the average degree per node with 95% confidence interval in each of the simulated models. Here, we see that the 95% confidence intervals in the dense models are relatively wider than the ones in the sparse models, which indicate that there are less stability in degree estimation with dense models. This is in accordance with the variation in density as shown in Figure 2, where the dense models are shown to have more variations in the estimated densities.

Figure 3: Average degree for every node with 95% confidence interval per set of DCGs.
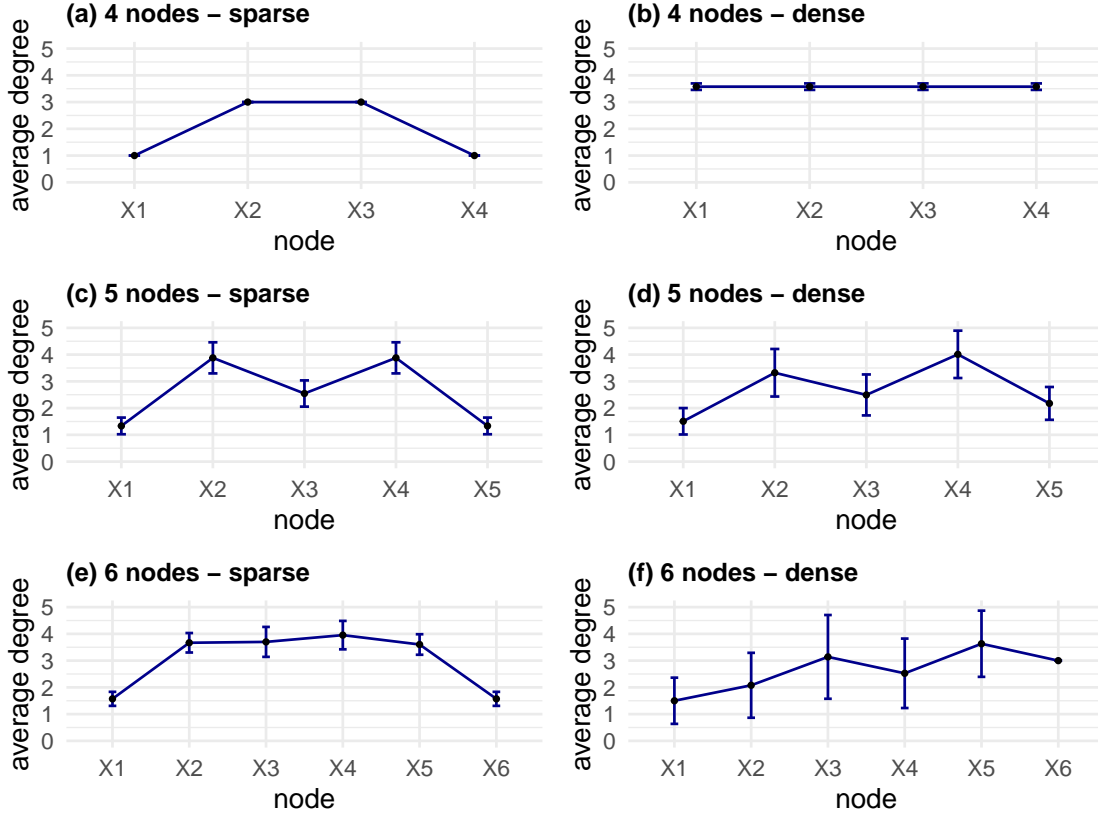
Note. In *(a)*, the standard error (SE) is zero and accordingly the confidence interval (CI) doesn't exist. For the rest, the values of SE are also very small $(.001 - .01)$, which makes it hard to visualize the CIs. For the ease of comparison, the SEs are scaled up. Hence, note that the width of CIs can only be compared across different cases in a relative sense.

4 Conclusion

All in all, with these additional analyses, we learn that there typically exists a considerable size of equivalence class and the variation in estimates is relatively large when the causal structure is dense. These findings imply quite a large uncertainty and instability in estimation with CCD, which raises doubts on estimation precision of causal models.

5

Despite these issues, the core conclusion from our simulation study (i.e., causal models are preferred to statistical network models when causal hypothesis is of interest) remains the same for the following reasons. First, even though the size of equivalence class is generally larger than desired, they are still more informative than the *undirected* statistical network models when it comes to inferring causal relations. Second, the difficulty in estimation with dense structures is equally problematic in statistical network models. In fact, our simulation study showed that causal models still outperformed the statistical network models in terms of approximating the density and degree centrality of true models. Outside of their use for causal discovery, statistical network models can be very useful as *descriptive* tools; to explore multivariate statistical relationships (Epskamp et al., 2018); to visualize clustering structures (Golino & Epskamp, 2017). However, if one is interested in the network approach to search for causal mechanism, then the focus should be on estimating causal models rather than statistical network models.

5 References

Bongers, S., Peters, J., Schölkopf, B., & Mooij, J. (2016, November). *Theoretical Aspects of Cyclic Structural Causal Models.* https://www.semanticscholar.org/paper/Theoretical-Aspects-of-Cyclic-Structural-Causal-Bongers-Peters/264c2fc543e02fac69a892c2c01c29256422de2f

Constantin, M., Schuurman, N. K., & Vermunt, J. (2021). *A General Monte Carlo Method for Sample Size Analysis in the Context of Network Models.* PsyArXiv. https://doi.org/10.31234/osf.io/j5v7u

Epskamp, S., & Fried, E. I. (2018). A Tutorial on Regularized Partial Correlation Networks. *Psychological Methods*, *23*(4), 617–634. https://doi.org/10.1037/met0000167

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The Gaussian Graphical Model in Cross-Sectional and Time-Series Data. *Multivariate Behavioral Research*, *53*(4), 453–480. https://doi.org/10.1080/00273171.2018.1454823

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLOS ONE*, *12*(6), e0174035. https://doi.org/10.1371/journal.pone.0174035

McNally, R., Mair, P., Mugno, B., & Riemann, B. (2017). Co-morbid obsessive–compulsive disorder and depression: A bayesian network approach. *Psychological Medicine*, *47*(7), 1204–1214. https://doi.org/10.1017/S0033291716003287

Richardson, T. (1996a). A discovery algorithm for directed cyclic graphs. *Proceedings of the Twelfth International Conference on Uncertainty in Artificial Intelligence*, 454–461.

Richardson, T. (1996b). *Discovering cyclic causal structure.* Carnegie Mellon [Department of Philosophy].

6 Appendix

Figure 4 shows the partial ancestral graph (PAG) for depression symptoms estimated by CCD algorithm based on the data provided by McNally et al. (2017). A couple of features are apparent. First, there are two clusters (islands), one comprising symptoms related to sleeping problems, and the other one comprising symptoms related to appetite issue. Secondly, there exists a cycle (feedback loop): `anhedonia` → `fatigue` → `retard` → `sad` → `anhedonia`, which seems reasonable in a substantive sense. Here, we do not know the true causal structure, but the overall findings from this causal model is deemed rather informative and sensible. Given this, it can be concluded that CCD is indeed applicable in practice to typical psychological data and can be used to help discovering some interesting causal dynamics in psychological processes.

Just for a comparison, the statistical network model (GGM) is also estimated using graphical LASSO (Epskamp & Fried, 2018) as shown in Figure 5. In accordance with the causal model, similar clustering structures are observed with symptoms related to sleep and appetite. Even though here with the statistical network model, we could additionally examine the sign (positive/negative) and the strength of relations between symptoms, we can hardly infer any causal relations or detect the presence of feedback loops, as we did with the causal model.

Figure 4: PAG estimated by CCD algorithm for depression symptoms.
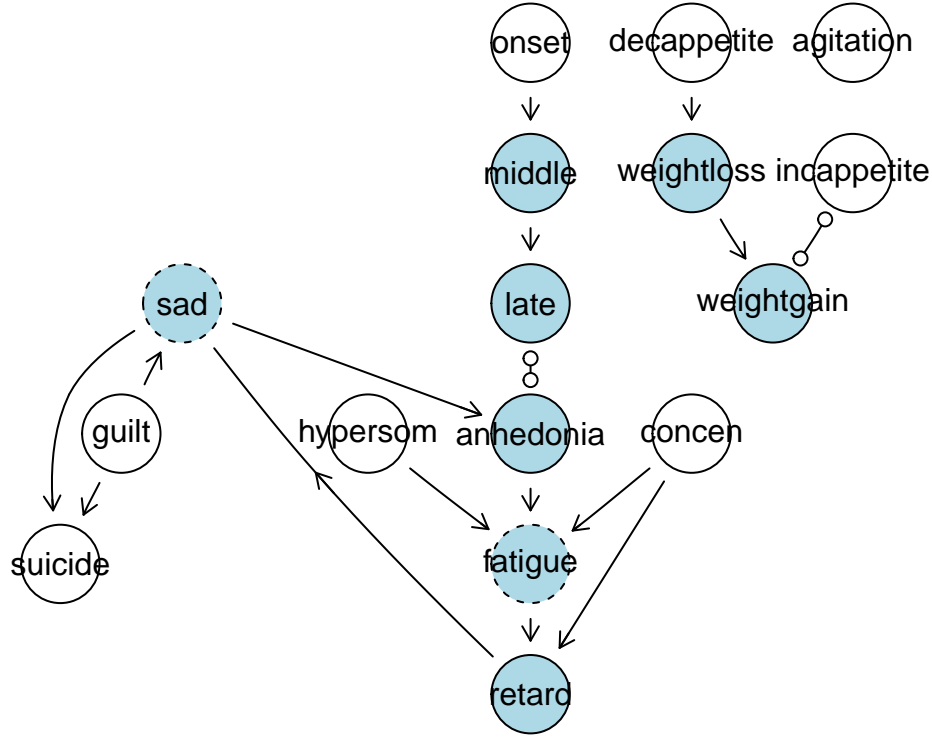
Note. In the PAG representation, there exists two types of underlining that are used in a triple of nodes: solid underlining (A - B̲ - C) and dotted underlining (A - ◌̲B̲ - C). The colored nodes (in blue) refer to the presence of the solid underlinings and the dashed nodes refer to the presence of dotted underlinings on the corresponding nodes. These underlinings are used to further orient the edges in a PAG. For more information, see Richardson (1996a).

Figure 5: Statistical network model constructed via graphical LASSO for depression symptoms.

ons: onset
mdd: middle
lat: late
hyp: hypersom
sad: sad
dcp: decappetite
inc: incappetite
wghtl: weightloss
wghtg: weightgain
cnc: concen
glt: guilt
scd: suicide
anh: anhedonia
ftg: fatigue
rtr: retard
agt: agitation