Individual Report

Critical Reflection on Comparison of GGM and Directed Cyclic Graph Model

Network Analysis 2022

Kyuri Park (12183881)

December 22, 2022

Word count: 994

Table of Contents

1 Summary of Final Project

In our final project, we aimed to investigate the utility of statistical network models as a tool for causal discovery in *cyclic* settings compared to the *directed cyclic graph* models (DCG) by means of a simulation study. We used *cyclic causal discovery* (CCD) algorithm to estimate DCGs (Richardson, 1996b) and compare it to the Gaussian graphical model (GGM). The comparison is made based on the overall density of the model and degree centrality of each node. 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 per node for all equivalent DCGs 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. GGM often overestimated the density and correspondingly resulted in high degree for almost all nodes in the considered models. The conclusion based on these results could be that statistical network model performs poorly as causal discovery tools in cyclic settings and hence, it shall be preferred to use a purpose-built *causal discovery methods* when one is interested in underlying causal mechanism.

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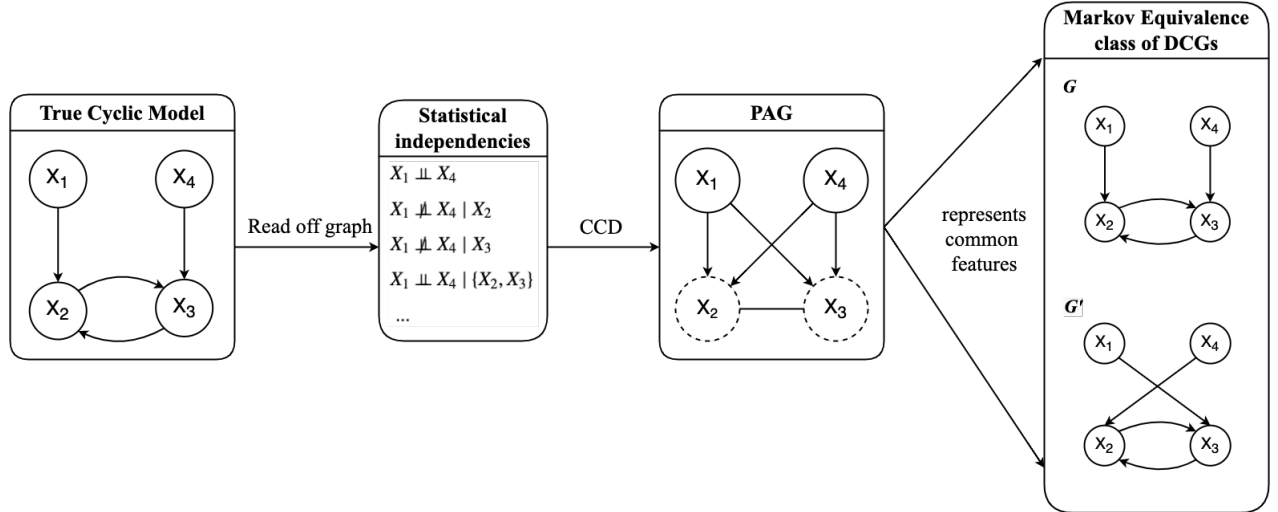
Note. Given the found statistical independencies from the true graph, CCD outputs a partial ancestral graph (PAG). It can be seen that in this example, the PAG represents two different directed cyclic graphs (DCG), of which we compute the density and degree centrality. Then, the average value of density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

---

[1]See Figure 1.

2 Reflection on Findings

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 class thus reflects the uncertainty in its estimation. But this was not investigated thoroughly in our simulation study. I believe that it is a critical aspect to further investigate, as it could be the case that even when the average density and average degree centrality are closely aligned to those of the true model, the *variation* in the set is huge, which implies a high uncertainty/instability in estimation.

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

3 Further Extension

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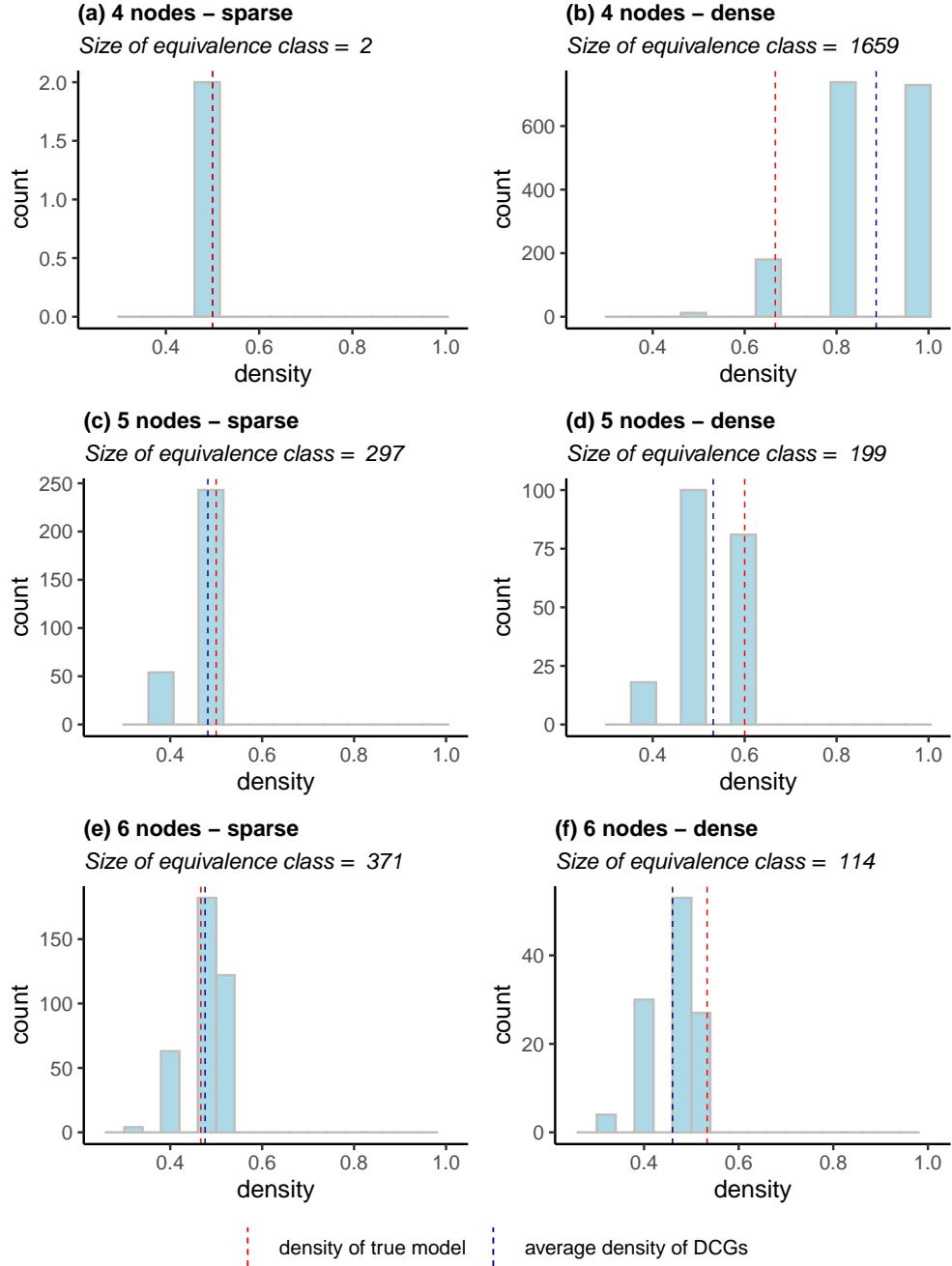
3.1 Variation in Density of DCGs

Figure 2 shows the size of equivalence class and distribution of density per set of the DCGs from each of considered models. First thing that stands out is that in general high-density models seem to have a slightly higher variation in density (i.e., spread of distribution is wider)[2] and accordingly the discrepancy between true density and average density of DCGs is larger compared to the low-density models. This indicates that when variation is high, the average density of DCGs is unlikely to align well with the true density. Secondly, it could be seen that the size of equivalence class is overall quite large, except for the *4 nodes model with low density* case. There is no specific pattern observed between the size of equivalence class and types of models, but these overall fairly large equivalence classes suggest the lack of precision in estimation with CCD, which could be seen as a weakness of causal discovery algorithm.

[2]The scale of x-axis is fixed the same across the graphs.

Figure 2: Distribution of density per the set of DCGs.

(a) 4 nodes – sparse
Size of equivalence class = 2

(b) 4 nodes – dense
Size of equivalence class = 1659

(c) 5 nodes – sparse
Size of equivalence class = 297

(d) 5 nodes – dense
Size of equivalence class = 199

(e) 6 nodes – sparse
Size of equivalence class = 371

(f) 6 nodes – dense
Size of equivalence class = 114

density of true model   average density of DCGs

3.2   Variation in Degree Centrality of DCGs

Figure 3 shows the average degree per node with 95% confidence interval in each of the considered models. Here, we see that the 95% confidence intervals in the high-density models are relatively wider than the ones in the low-density models, which indicate that there are more uncertainty in degree estimation with high-density models. This is in accordance with the variation in density as shown in Figure 2, where the high-density models are shown to have more variation in the overall density estimation. All in all, with these additional analyses, we learn that the variation in estimation with CCD tends to become higher when the models are dense, and in general, the size of equivalence class of DCGs is fairly large (> 100), which needs to be accounted for, when using the causal discovery algorithms.

Figure 3: Average degree per node with 95% confidence interval.
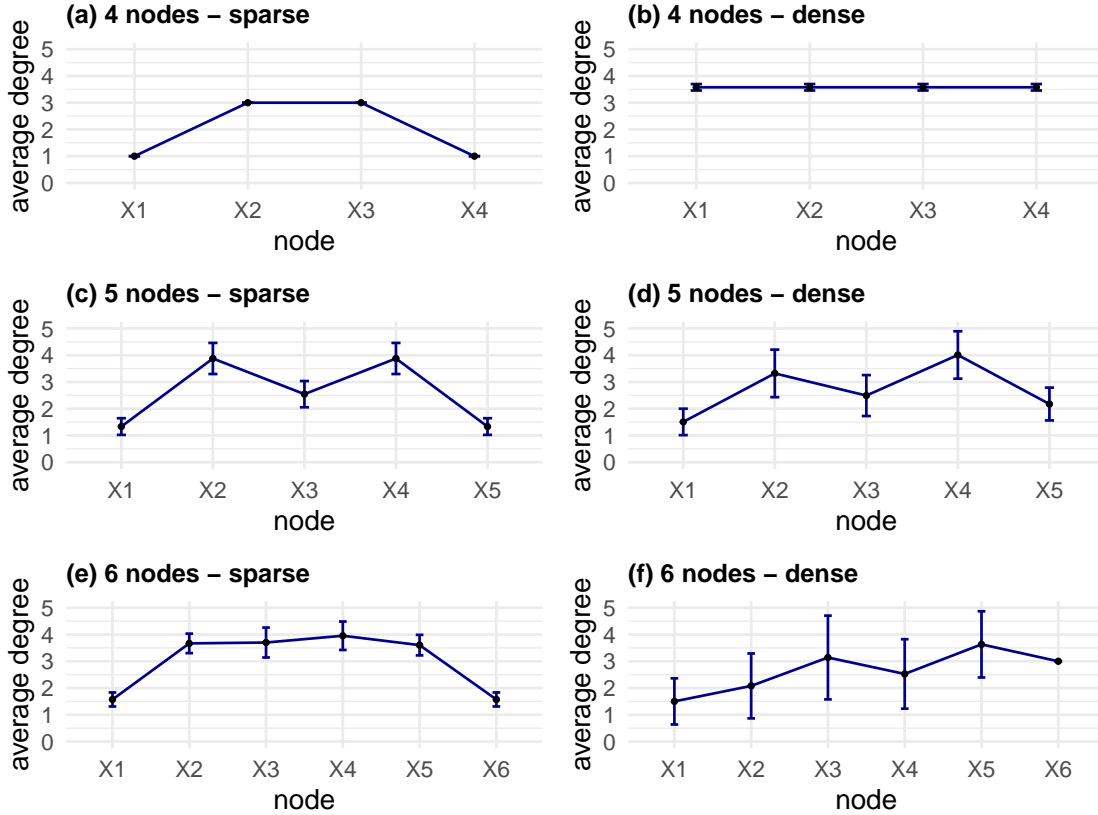
Note. In (a), the standard error (SE) is zero and accordingly confidence interval (CI) doesn't exist. For the rest, the value of SE is also very small (0.01 - 0.01), which makes it hard to visualize CIs. For the ease of comparison, the SE is scaled up. Hence, note that the width of CIs can only be compared in a relative sense.

4 Conclusion

In this project, we aimed to clarify the role of statistical network models (GGM) as causal discovery tools compared to the causal discovery method (CCD) in cyclic settings. Overall, the results indicate that GGM is not suitable for inferring causal structure with cycles. Even though, CCD comes with a couple of limitations as previously discussed (e.g., not able to uniquely identify one true graph), the limitation is deemed not detrimental and it is still considered to be more useful when it comes to generating causal hypotheses based on observational data. Note that it does not state that the statistical network models are completely uninformative. Outside of their use for causal discovery, there are other attractive reasons to use statistical network models. For instance, they are very useful as descriptive tools to summarize the multivariate statistical relationships (Epskamp et al., 2018) and to visualize the clustering structure (Golino & Epskamp, 2017). However, when it comes to identifying causal relations, these statistical network models come up short as our simulation study showed. Accordingly, when one is interested in the network approach to search for causal mechanism, then the focus should be on estimating a causal model rather than statistical network models.

5 References

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 comprising symptoms related to appetite issue. Secondly, there exists a cycle (feedback loop): `anhedonia` → `fatigue` → `retard` → `sad` → `anhedonia`, which is interesting and seems reasonable. Here, we do not know the true model but the overall findings from this causal model is considered rather informative and sensible. Therefore, it is concluded that CCD can be actually applied in practice to typical psychological data and be used to help discovering some interesting causal dynamics between variables.

For the 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 is again observed between symptoms related to sleep and appetite. Even though here with the statistical network model, we could additionally examine the sign of relations (positive/negative) and the strength of relations between symptoms, we cannot directly infer any causal relations, as expected.

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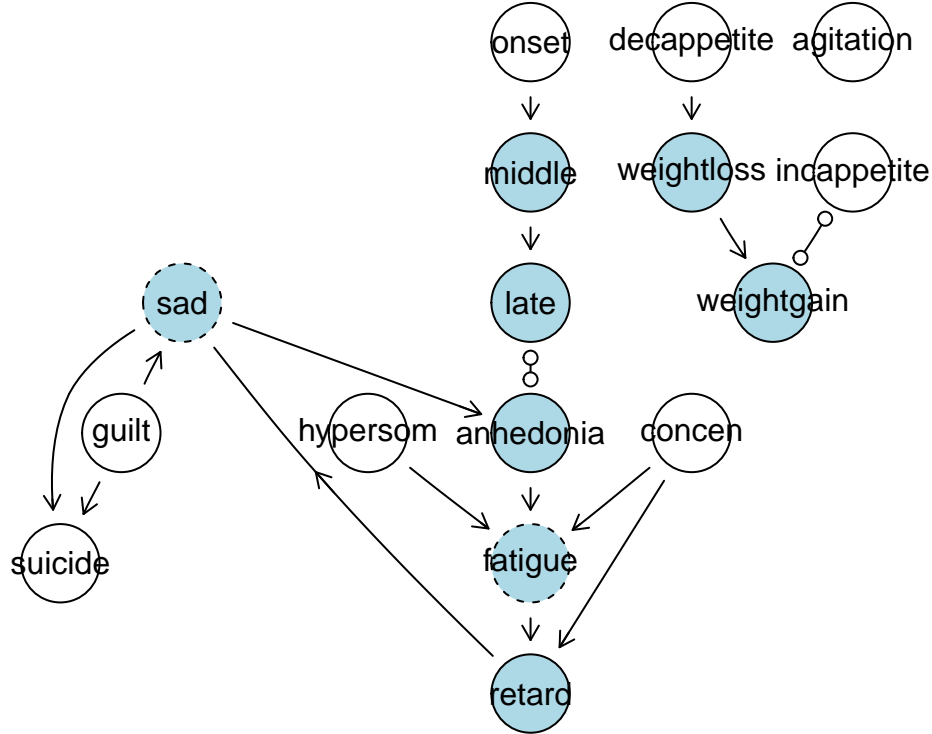
Note. In the PAG representation, there exists two types of underlining that can be used in a triple of nodes: solid underline (A - B̲ - C) and dotted underline (A - ̤B̤ - C). The colored nodes (in blue) refer to the presence of the solid underlining and the dashed nodes refer to the presence of the dotted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