



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

Network Analysis 2022

Kyuri Park (12183881)

December 22, 2022

Word count: 994

Table of Contents

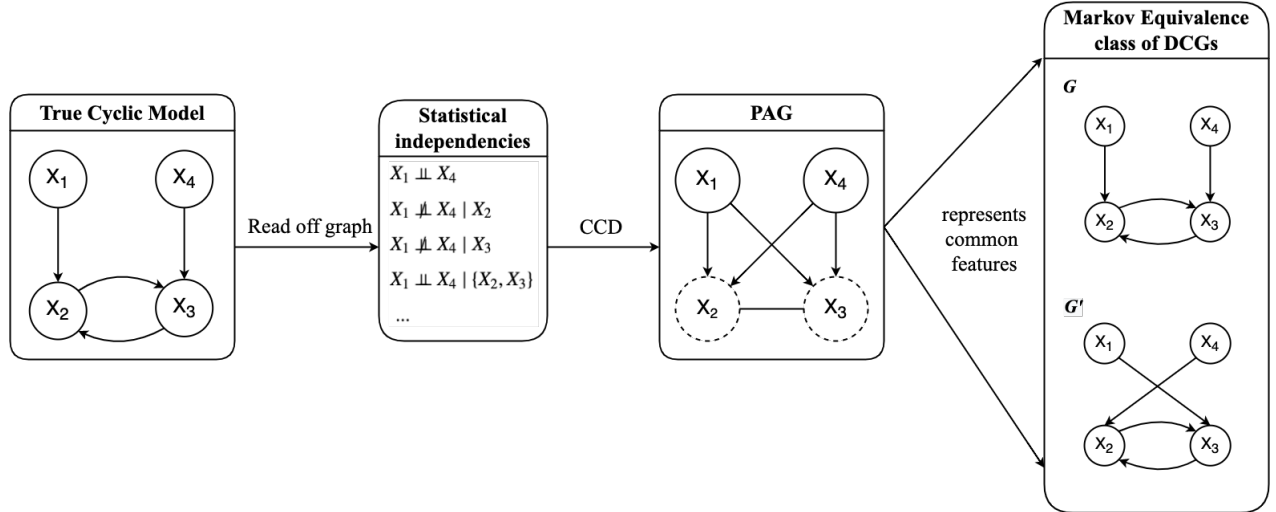
1	Summary of Final Project	2
2	Reflection on Limitations	3
3	Further Extensions	3
3.1	Variation in Density of DCGs	3
3.2	Variation in Degree Centrality of DCGs	5
4	Conclusion	6
5	References	6
6	Appendix	7

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 *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 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.

¹See 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 class thus reflects the uncertainty in its estimation. 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, implying a high uncertainty/instability in estimation.
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 data and utilized in psychological research in practice is yet questionable.

3 Further Extensions

To examine the overall uncertainty and variation in estimation as to the first limitation, we check the size of equivalent class from each of the considered 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 in practice, we refer the interested readers to Appendix (section 6).

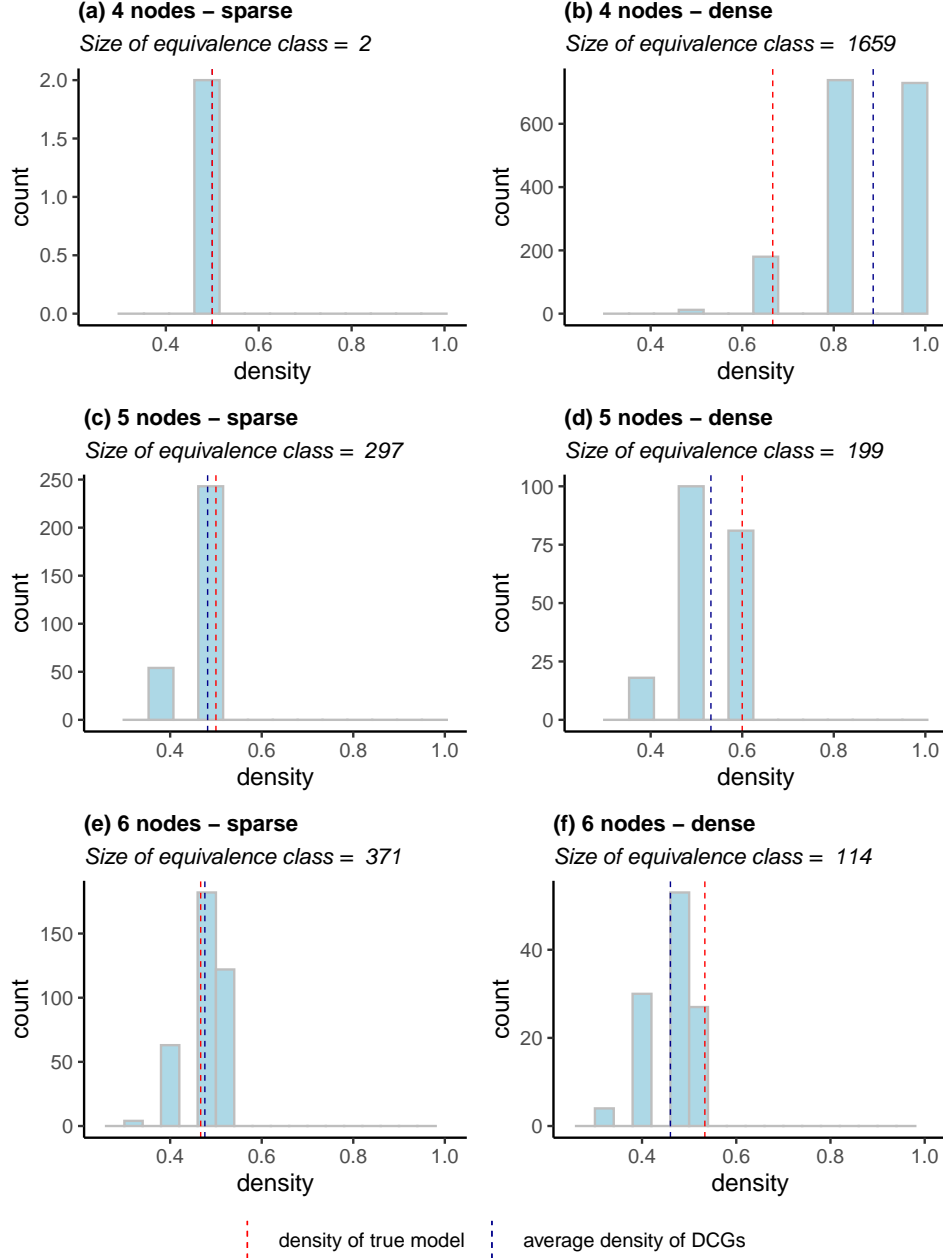
3.1 Variation in Density of DCGs

Figure 2 shows the size of equivalence class and the distribution of density per the set of DCGs from each of the considered models. First thing that stands out is that dense models (right column of Figure 2) seem to have a slightly higher variation in density (i.e., spread of distribution is wider) and accordingly the discrepancy between the true and average density of DCGs is larger in dense models than in the sparse models. This indicates that when the variation is high in the estimated densities, the average density of DCGs is unlikely to approximate the true density well. Hence, it is not probable that the average density of DCGs follow the true density closely, while having a high variance in the set of densities.

Secondly, it could be seen that the size of equivalence class is overall quite large, except for the *sparse model with 4 nodes* (Figure 2 (a)). There is no specific pattern observed between the size of equivalence class and the types of models, but these overall fairly large equivalence classes suggest the lack of certainty in

estimation with CCD, which is actually known as one of the weaknesses of the causal discovery algorithms (Bongers et al., 2016).

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



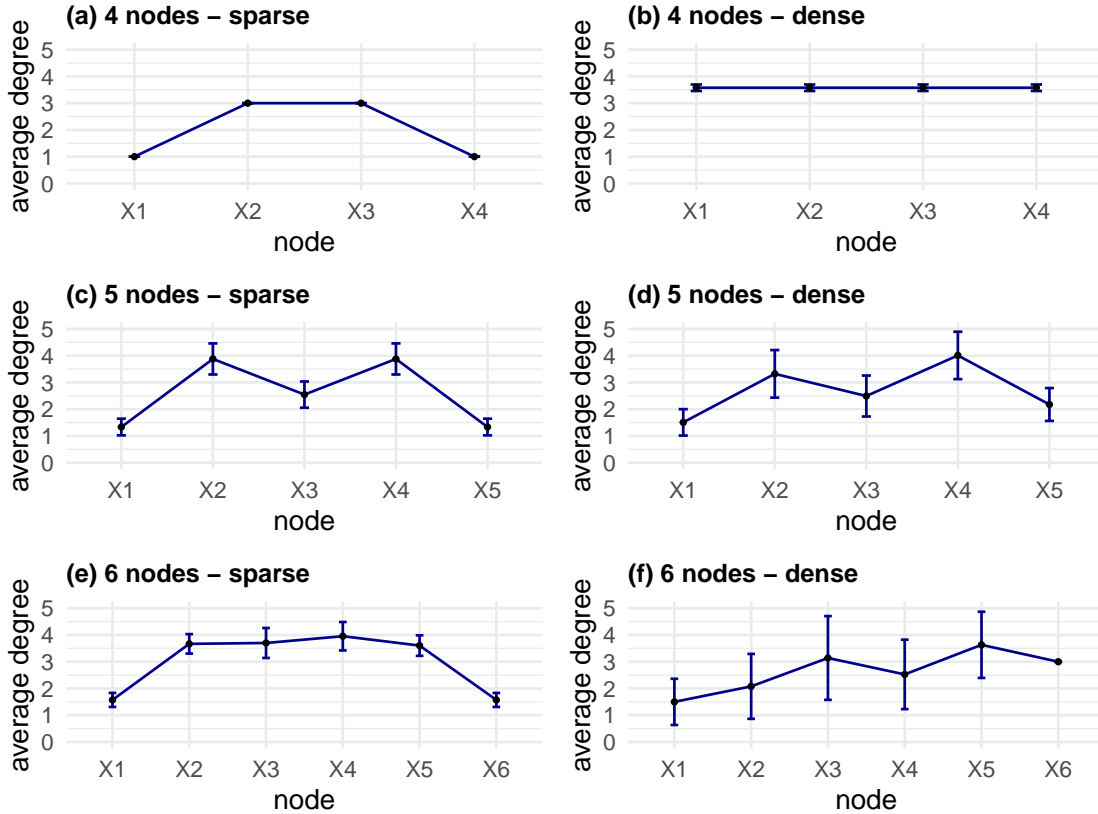
Note. The scale of x-axis is fixed the same across the graphs so that the comparison between distributions (e.g., spread) can be easily made. In (a), both DCGs in the equivalence class have the same density that is identical to the true density (the two lines overlap). The dense model with 4 nodes (b) 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. In general, CCD struggles more to estimate the dense models, since in the dense cases, there are more variations in densities with larger discrepancies in comparison to the sparse models.

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 dense models are relatively wider than the ones in the sparse models, which indicate that there are more uncertainty 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 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. But as the variation becomes larger, the average density/degree also deviates more from the true value. Thus, it is not likely that the average density/degree closely follows the true density/degree in the cases with high variance. However, we still need to keep in mind the rather general limitation that there typically exists a considerable size of equivalence class and correspondingly need to think of what that implies for one's analysis, 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 values of SE are also very small (0.01 - 0.01), which makes it hard to visualize 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

In this project, we aimed to clarify the role of statistical network models (GGM) as causal discovery tools in cyclic settings compared to the directed cyclic graph models (DCG) estimated by CCD algorithm. Overall, the results indicate that GGMs are not suitable for inferring causal structure with cycles. Even though, CCD comes with a couple of limitations as previously discussed, the limitation is not deemed to be detrimental and it is still considered more useful when it comes to modeling a causal structure, as it can approximate the causal relations more accurately.

Note that it does not state that the statistical network models are completely uninformative in this regard. Statistical models can provide additional information such as the sign and strength of relationships (Epskamp et al., 2018) and also visualize clustering structures nicely (Golino & Epskamp, 2017), which can be useful. However, when it comes to identifying causal relations, the 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 such as DCGs rather than statistical network models.

5 References

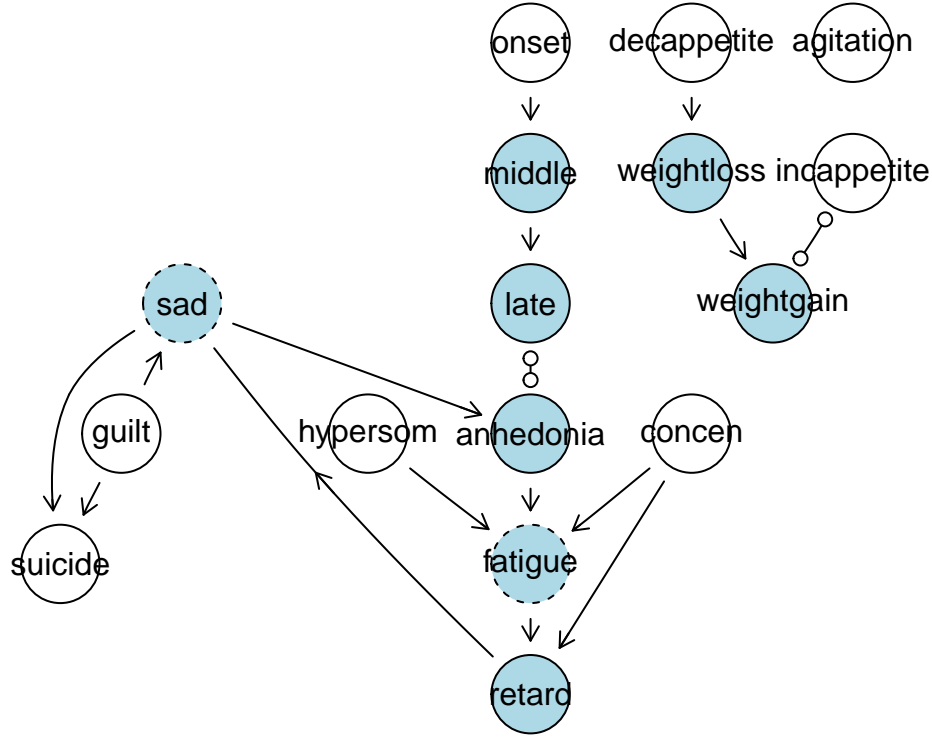
- Bongers, S., Peters, J., Schölkopf, B., & Mooij, J. (2016, November). *Theoretical Aspects of Cyclic Structural Causal Models*. <https://www.semanticscholar.org/paper/Theoretical-Aspects-of-Cyclic-Structural-Causal-Bongers-Peters/264c2fc543e02fac69a892c2c01c29256422de2f>
- Constantin, M., Schuurman, 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., Möttus, 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** \rightarrow **fatigue** \rightarrow **retard** \rightarrow **sad** \rightarrow **anhedonia**, which seems reasonable in a substantive sense. Here, we do not know the true model 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 in 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 can be used in a triple of nodes: solid underlining (A - B - C) and dotted underlining (A - B - 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.

