Technical Note: The *ComDet* Package - Iterated Community Detection for the Analysis of Community Membership, its Stability and the Detection of Fuzzy Community Boundaries.

M. A. Werner
24-05-2018

Contents

1.	Psychological Networks	p. 1
2.	Common Causes vs. Communities	p. 1
3.	Community Detection	p. 2
4.	The Extended Function	p. 2
5.	Output of the Extended Function	p. 3
6.	Discussion and Next Steps	p. 5
7.	References	p. 5

1. Psychological Networks

In recent years, network analysis has entered psychological science (Borsboom, 2017). In psychological network models, nodes are variables and edges are estimated relationships between these variables (Epskamp & Fried, 2017; Epskamp, Waldorp, Möttus, & Borsboom, 2018). In other words, psychological networks represent multivariate joint distributions that are visualized as network graphs. Thereby, psychological networks differ from, for instance, social networks, due to the network connections being estimated rather than directly observed (Epskamp, Rhemtulla, & Borsboom, 2017). The estimation, visualization, and analysis of psychological constructs as complex systems has offered new and insightful ways to analyze multivariate psychological data.

Originally, psychological network models were introduced as alternatives to - so called - common cause models (Borsboom & Cramer, 2013). Network psychometricians argued that some psychological variables, that purportedly measure a psychological construct, do not associate with each other by virtue of a shared common cause, but by virtue of potentially causal relations among the psychological variables themselves (Cramer et al., 2012). From a common cause perspective, variable associations are spurious and would disappear if we were to control for inter-individual differences in the underlying, measured construct. Yet, these purportedly spurious relationships build the very structure of a network model, and are regarded as carrying insightful, rather than expendable, information.

2. Common Causes vs. Communities

Besides other known network properties (such as small-worldness), psychological networks can exhibit clustering behavior (Blanken et al., 2018; Bringmann & Eronen, 2018; Golino & Demetriou, 2017; Golino & Epskamp, 2017). That is, some variables cluster together more strongly with each other than with variables outside of that group of variables. For example, symptom networks of the DSM diagnostic categories do show symptom clusters of diagnoses and syndrome groups, but also symptom overlap between diagnostic categories (Borsboom, Cramer, Schmittmann, Epskamp, & Waldorp, 2011). From a common cause perspective, such clusters could imply that there is an underlying shared process that causes clustered symptoms to co-occur and that “cross-loading” symptoms represent nuisance, or what has been called “transdiagnostic factors”. From the network perspective, symptom clusters could represent structural or functional sub-units in the larger network structure and overlapping symptoms represent bridging symptoms between problem areas (Blanken et al., 2018).

The common cause perspective offers established conceptual and statistical models to study such co-occurrence of symptoms (i.e., the conceptual reflective measurement model and the statistical factor model). However, psychological network clustering has not yet been studied and conceptualized as thoroughly. Thus, psychological network science could profit from being able to detect clusters statistically through network analytical techniques. Recently, Blanken et al. (2018) have introduced network communities in the psychological network literature. They suggest that network community detection can help conceptualize and study “hallmark symptoms” and comorbidity. Furthermore, community detection might help to statistically detect sub-domains in other kinds of psychological network models, such as subdomains in personality, representing so-called “mutually reinforcing components” rather than epiphenomena of common causes (Cramer et al., 2012).

In this project, we aim to extend another community detection technique, called spinglass community detection, which has been applied to psychological networks in empirical studies (Csárdi, 2017; Heeren & McNally, 2016). In doing so, we follow Blanken et al. (2018) in their call for the extended application of network analytical techniques onto psychological networks. Such implementations represent the first step in building and vetting a variety of statistical techniques

for detecting psychological network communities (see also Golino & Epskamp, 2017, for another implementation). Our extended community detection function builds upon the existent literature in two ways: 1. it offers summarizing output that communicates the stability and variability of particular community results. 2. Our function can be readily applied to psychological networks (which are commonly estimated in R; R Core Team, 2018) since our function is implemented in R (using the spinglass algorithm implemented in the *igraph*-package; Csárdi, 2017).

3. Community Detection

A network community can be defined as a subgraph of nodes which are more densely connected amongst each other than with nodes outside of the subgraph¹ (Yang, Algesheimer, & Tessone, 2016). Network science offers several ways to detect such communities in network graphs (Csárdi, 2017; Yang et al., 2016). One such method is the spinglass community detection algorithm. The spinglass algorithm uses simulations based on the Potts-model from statistical mechanics (Reichardt & Bornholdt, 2006). Due to the simulations, the algorithm is non-deterministic. In other words, if we run the spinglass algorithm on the same structure, using the same parameters, we might retrieve variable node-community demarcations.

During different network analytical projects, we noted that spinglass resulted in such varying results. For instance, 1. nodes switched community membership across runs and 2. smaller distinct communities would merge into larger communities across runs. Interestingly, these differences of results across runs showed repeating patterns and were, oftentimes, consistent with expectations. In other words, 1. nodes switching community membership across runs could indicate that these nodes belonged to both of those communities and represented bridging nodes. 2. Communities merging across runs could indicate that nodes exhibited overlapping and/or hierarchical clustering. Therefore, rather than dispensing with the spinglass algorithm in psychological network science, we exploited the variability in results in order to 1. detect nodes that might represent bridging nodes between communities and 2. find overlapping and hierarchically organized communities in the overall community structure.

4. The Extended Function

Our function extends the spinglass algorithm as it is currently implemented in the *igraph*-package (Csárdi, 2017) by repeatedly performing the spinglass community detection algorithm for a specified number of runs, quantifying and summarizing the results across runs, and visualizing those results.

When we run the *igraph* spinglass algorithm on a network structure once, *igraph* offers different output, of which the membership vector is the most important for the current purposes. This membership vector encodes which nodes belong to which community by means of numbers: all nodes that belong to the same community will be labelled with the same number (henceforth called, spinglass solution). Repeatedly running the spinglass algorithm on the same network structure can result in differently encoded spinglass solutions, even in cases where the solutions are equivalent (i.e., even though the communities include the very same nodes, a community would be labelled “1” in the first run, whereas the equivalent community would be labelled “2” in a following run). In the extended function, we therefore recode the spinglass solutions in order to meaningfully summarize the solutions from repeated spinglass runs.

¹ This definition is a conceptual rather than mathematical definition but should suffice for the current aims. It is surprisingly difficult to give a complete and unambiguous definition of communities (Hennig, 2015).

5. Output of the Extended Function

The extended function provides the researcher with the following information that can currently not be obtained with (spinglass) community detection:

1. A heatmap (see Figure 1) with the nodes on the y- and x-axis representing the frequency (across the runs) with which two nodes are assigned to the same community, ordered according to their clustering (using the *pheatmap* package; Kolde, 2018). This heatmap summarizes which nodes group together most frequently, regardless of the observed community structures (i.e., the heatmap summarizes how often two nodes group together across potentially different spinglass solutions).

Figure 1. Heatmap with nodes on x- and y-axis. Boxes' saturation indicates the frequency with which two nodes ended up in the same community across 1000 spinglass runs on a simulated network structure.

2. All unique spinglass solutions (see Figure 2) in descending order of frequency. For instance, below, the first solution was produced in 60 % of 1000 runs, the second in 32 % of 1000 runs etc., and there were 8 unique spinglass solutions across 1000 runs. This stacked visualization aims to compile the solutions across the runs to retrieve the most frequently occurring spinglass solution, i.e. how many unique spinglass solutions are there for a given number of runs, and in which frequency do they occur?

Figure 2. Panel A depicts all unique spinglass solutions in order of frequency across 1000 spinglass runs on a simulated network structure. Panel B shows the two most frequently occurring solutions which are also plotted in the first row of panel A.

Crucially, the output allows researchers to gain insight into the stability and variability of node-membership and communities:

1. The output can signal how stable the number and borders of communities are. First, we would have much confidence in the stability of a particular solution if it came up in the large majority of all runs. Also, we can see how borders migrate across similarly frequent solutions. For instance, in the above example we could be confident that the lower two communities (V5-V9 and V10-V18) cluster together consistently (since they occurred in more than 92 % of all solutions).

2. The output may also suggest which nodes bridge particular communities. Even though no clear cut-offs exist, a node that occurs 50% of the runs within one community of nodes and 50% of runs with another community of nodes (which by themselves occur together more than, for instance, 90% of runs) could probably be considered a bridging node between these communities. In the above example, V3 most likely represents such a bridging node as it switches membership between its own, the V1-V2-V4, and the V10-V18 community (note, however, that

the frequency with which V3 belongs to the V10-V18 community is probably not high enough to consider it a bridging node to that community with much certainty).

6. Discussion and Next Steps

The package is not yet fully developed and we are still in the process of validation by using simulations and benchmark networks (Lancichinetti & Fortunato, 2009). However, spinglass community detection has already been validated on other kinds of networks, suggesting that the algorithm itself performs adequately (Yang et al., 2016).

Nevertheless, it remains to be seen whether the rationale of the spinglass algorithm can be applied properly to gain insight into psychological networks. Specifically, we need to establish whether all psychological networks can be validly conceptualized as “spinglasses” (i.e., extensions of Ising models; Reichardt & Bornholdt, 2006). More generally, we need to establish whether we can use the structural topography of a network (i.e., the constellation and weight of edges between nodes) to infer community structure validly in psychological networks. The latter point has recently been raised in criticizing the use of centrality indices for pinpointing hallmark variables in psychological networks (Fried et al., 2018); the rationale of this critique seems to apply to community detection as well. Regarding our extended function, we need to assure whether it is sensible to exploit the variability of the spinglass algorithm in order to find bridging nodes in (psychological) community structures. Finally, we think it is important to note that current visualizations of community structure, such as those presented in Figure 2, sometimes display “overlapping clouds”, falsely implying that a node which is covered by two clouds belongs to two communities. This is not the case. We hope that future network visualization packages draw community-clouds without such misleading overlap.

This technical note will be updated with new insights in the near future. The final aim of this project is to provide researchers with a freely available R-package that runs the *ComDet* function (possibly including other non-deterministic detection algorithms, in addition to spinglass community detection).

7. References

- Blanken, T. F., Deserno, M. K., Dalege, J., Borsboom, D., Blanken, P., Kerkhof, G. A., & Cramer, A. O. J. (2018). The role of stabilizing and communicating symptoms given overlapping communities in psychopathology networks. *Scientific Reports*, 8(1), 5854. <https://doi.org/10.1038/s41598-018-24224-2>
- Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16(1), 5–13. <https://doi.org/10.1002/wps.20375>
- Borsboom, D., & Cramer, A. O. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, 9, 91–121. <https://doi.org/10.1146/annurev-clinpsy-050212-185608> [doi]
- Borsboom, D., Cramer, A. O. J., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The Small World of Psychopathology. *PLoS ONE*, 6(11). <https://doi.org/10.1371/journal.pone.0027407>
- Bringmann, L. F., & Eronen, M. I. (2018). Don't blame the model : Reconsidering the network approach to psychopathology. *Psychological Review*.
- Cramer, A. O. J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., ... Borsboom, D. (2012). Measurable like temperature or mereological like flocking? On the nature of personality traits. *European Journal of Personality*, 26(4), 451–459. <https://doi.org/10.1002/per.1879>
- Csárdi, G. (2017). Package ‘igraph’ - Routines for simple graphs and network analysis. It can

- handle large graphs very well and provides functions for generating random and regular graphs, graph visualization, centrality methods and much more. Retrieved from <https://cran.r-project.org/web/packages/igraph/igraph.pdf>
- Epskamp, S., & Fried, E. (2017). A Tutorial on Regularized Partial Correlation Networks. *ArXiv Preprint*, 1–20. <https://doi.org/10.1037/met0000167>
- Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized Network Psychometrics: Combining Network and Latent Variable Models. *Psychometrika*, 82(4), 904–927. <https://doi.org/10.1007/s11336-017-9557-x>
- Epskamp, S., Waldorp, L. J., Möttus, R., & Borsboom, D. (2018). The Gaussian Graphical Model in Cross-sectional and Time-series Data. *ArXiv Preprint*, 1–31. Retrieved from <http://arxiv.org/abs/1609.04156>
- Fried, E. I., Eidhof, M. B., Palic, S., Costantini, G., Huisman-van Dijk, H. M., Bockting, C. L. H., ... Karstoft, K.-I. (2018). Replicability and Generalizability of Posttraumatic Stress Disorder (PTSD) Networks: A Cross-Cultural Multisite Study of PTSD Symptoms in Four Trauma Patient Samples. *Clinical Psychological Science*, 1–17. <https://doi.org/10.1177/2167702617745092>
- Golino, H. F., & Demetriou, A. (2017). Estimating the dimensionality of intelligence like data using Exploratory Graph Analysis. *Intelligence*, 62, 54–70. <https://doi.org/10.1016/j.intell.2017.02.007>
- 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). <https://doi.org/10.1371/journal.pone.0174035>
- Heeren, A., & McNally, R. J. (2016). An integrative network approach to social anxiety disorder: The complex dynamic interplay among attentional bias for threat, attentional control, and symptoms. *Journal of Anxiety Disorders*, 42, 95–104. <https://doi.org/10.1016/j.janxdis.2016.06.009>
- Hennig, C. (2015). What are the true clusters? *Pattern Recognition Letters*, 64, 53–62. <https://doi.org/10.1016/J.PATREC.2015.04.009>
- Kolde, R. (2018). Package “pheatmap” - Implementation of heatmaps that offers more control over dimensions and appearance. Retrieved from <https://cran.r-project.org/package=pheatmap>
- Lancichinetti, A., & Fortunato, S. (2009). Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Physical Review E*, 80(1), 016118. <https://doi.org/10.1103/PhysRevE.80.016118>
- R Core Team. (2018). R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from <https://www.r-project.org/>
- Reichardt, J., & Bornholdt, S. (2006). Statistical Mechanics of Community Detection. *Physical Review E*, 74, 016110. <https://doi.org/10.1103/PhysRevE.74.016110>
- Yang, Z., Algesheimer, R., & Tessone, C. J. (2016). A Comparative Analysis of Community Detection Algorithms on Artificial Networks. *Scientific Reports*, 6(1), 30750. <https://doi.org/10.1038/srep30750>