Network Analysis 2017

Location 1
Grading 1
Learning Goals 2
Outline 3
  Week 1 – Introduction to R 3
  Week 2 – Introduction to network analysis 3
  Week 3 – Causality 4
  Week 4 – Markov random fields I 4
  Week 5 – Markov random fields II 5
  Week 6 – Directed acyclic graphs 5
  Week 7 – Time-series analysis 6
Final Project 7
  Possible theoretical final project topics 8
    Within- and between-subject variance 8
    Replicability of networks 8
    Factor models and network model equivalences 9
    Large-scale simulation studies 9
    Review of topic-specific literature 9
    References 10

Location

Every Monday:

- Lecture
- 13:00 – 15:00 in REC-G -1.11

Every Thursday:

- Practical (may include lecture)
- 15:00 – 17:00 in REC-JK B.11

Grading

- Individual assignments: 60% of final total grade
    - Week 1 and 7: 5%, weeks 2 – 6: 10%
- Final project: 40% of final total grade

Both average grades must have an average grade of at least 5.5 to pass the course.

Learning Goals

An encompassing goal of the course is to prepare the student for academic discourse (PhD). That is, the final learning goals align with the Research Master in general, preparing the student to…

- Critically asses academic literature
- Work through the empirical cycle, connecting previous literature to current data analysis
- Communicate both your findings and thoughts through academic writing
- Communicate to an audience of your peers through presenting

To accomplish these general goals, weekly assignments will sometimes feature an **essay question**, which is judged on writing quality, in addition to if the question is answered. Throughout the course students will work on a **final project**, which requires the student to critically evaluate literature or to perform their own analysis and to communicate the results both in writing and presenting at the end of the course.

Further aims of this course are to prepare the student to…

- Work with R and Rstudio (week 1)
- Analyze social networks and understand the difference between social and psychological networks (week 2)
- Understand the interventionist approach to causality (week 3)
- Estimate undirected network models from cross-sectional data (week 4)
- Perform model selection, stability checks and further inferential methods on undirected network models (week 5)
- Estimate directed acyclic networks from cross-sectional data (week 6)
- Estimate graphical vector-autoregression models from time-series data (week 7)

In order to accomplish these goals a mixture of conceptual knowledge and practical skills is required. To this end, every Monday there is a **conceptual lecture** and every Thursday there is a **practical**. Skills will be assessed in a weekly assignment, consisting of both **conceptual questions** and **practical questions**.

Specific learning goals to each week follow on the next pages.

Outline

Week 1 – Introduction to R

- Practical (02/11)
  - Introduction to R
- Assignments
  - Start Assignment 1 (5%)
  - Start Final Assignment (40%)
- Learning goals
  - Know how to work with R and RStudio
  - Create a matrix in R
  - Load and inspect data in R
  - Select rows and columns of a dataset
  - Compute a variance—covariance matrix in R
  - Install and load packages in R
  - Find help on R functions

Final project: groups of 2 students each apply the discussed methodology on a dataset and write a short paper interpreting the results.

Week 2 – Introduction to network analysis

- Lecture (06/11)
  - Introduction to Network Analysis
- Practical (09/11)
  - Creating graphs in R
  - Social Network Analysis
- Assignments
  - Hand in Assignment 1 (5%)
  - Start Assignment 2 (10%)
- Learning goals
  - Understand the theoretical foundations of the network perspective to Psychology
  - Explain the difference between social network analysis and network psychometrics
  - Visualize and analyze (social) networks using the qgraph package for R

Week 3 – Causality

- Lecture (13/11)
    - Causality (guest lecture by Denny Borsboom)
- Practical (16/11)
    - Causality
- Assignments
    - Hand in Assignment 2 (10%)
    - Start Assignment 3 (10%)
- Learning goals
    - Understand the basic principles of the interventionist approach to causality
    - List implied conditional independence relationships through d-separation
    - Investigate conditional independence relationships in data

Week 4 – Markov random fields I

- Lecture (20/11)
    - Markov Random Fields I
- Practical (23/11)
    - Markov Random Fields I
- Assignments
    - Hand in Assignment 3 (10%)
    - Start Assignment 4 (10%)
- Learning Goals
    - Interpret Markov random fields (Gaussian graphical models and Ising models)
    - Draw an implied Markov random field given a certain causal structure
    - Estimate Markov random fields from cross-sectional data

Week 5 – Markov random fields II

- Lecture (27/11)
  - Markov Random Fields II
- Practical (30/12)
  - Markov Random Fields II
- Assignments
  - Hand in Assignment 4 (10%)
  - Start Assignment 5 (10%)
- Learning Goals
  - Apply LASSO regularization to perform model selection in Markov random fields
  - Compute centrality indices and understand how these can be calculated by hand
  - Apply bootstrapping procedures to check network models for accuracy and stability
  - Apply permutation tests to check for differences between two groups

Week 6 – Directed acyclic graphs

- Lecture (04/12)
  - Directed network discovery
- Practical (07/12)
  - Directed network discovery
- Assignments
  - Hand in Assignment 5 (10%)
  - Start Assignment 6 (10%)
- Learning goals
  - Estimate directed acyclic graphs (DAG) using R packages pcalg and bnlearn, and interpret the results
  - Explain the assumptions behind DAG estimation
  - Explain the concept of equivalent models

Week 7 – Time-series analysis

- Lecture (11/12)
    - Time-series Analysis
    - Start Assignment 7 (5%)
- Practical (14/12)
    - Hand in Assignment 6 (10%)
- Learning goals
    - Estimate graphical vector auto-regression models from n = 1 time-series data, and interpret the results
    - Estimate multi-level graphical vector auto-regression models from n > 1 time-series data, and interpret the results
    - Explain the difference between within-subject effects and between-subject effects
    - Explain the value and interpretation of cross-sectional data analysis

Week 8 – Final projects

- Lecture (18/12)
    - Project presentations
- Practical (21/12)
    - Project presentations
- Assignments (22/12)
    - Hand in Final Assignment (40%) and assignment 7 (5%)

Final Project

The final projects are to be done in groups of two, so form a group early! You may do a project by yourself, but it will not be graded differently meaning you will do twice as much work! Post on blackboard to claim your topic as soon as possible. For each topic, you are required to present in the last week (10 minutes max including questions and changing between groups) and to write a scientific report. The report must be written in APA Style as a scientific paper (word limit: 3000).

1) If you have your own data, perform a network analysis and write down your findings in a paper. Only choose this option if your data are really best analyzed via network analysis, you plan to publish this analysis or some version of it on future data, and your analysis is otherwise "real" and not contrived.

2) Gather data and perform a network analysis. For example, you can gather time-series data of yourself during the course, or you can gather cross-sectional data amongst people doing the course, friends and online. This project requires time to collect the data. Contact me about it as soon as possible to discuss specifics.

3) Re-analyze an existing dataset. For example, factor analysis papers often report a variance—covariance matrix, which may be used to perform a network analysis. Raw data is also shared more often now, which is better as it allows for more sophisticated analyses as well as stability/accuracy checks. Critically discuss in the paper the original work and how/if your re-analysis gives new insights.

4) Review the literature and write a critical report on a current topic of debate in the field of network analysis. I have listed several topics on the next pages, but you may also choose a topic yourself. Only one topic may be investigated per group, so be sure to claim your topic on blackboard!

Grading for the final project will be based on your presentation and your written report. The difficulty of your project will be taken into account.

A very important note on plagiarism: When working in pairs, it sometimes happens that one partner copies some text from an article, then the other partner puts that text into the paper thinking that the first person wrote it, and they end up with a partially plagiarized paper. DO NOT LET THIS HAPPEN TO YOU. Plagiarism is a very big deal, and any cases will be caught and sent to the examination committee. It's best not to copy whole sentences from papers to begin with (take notes in your own words!), but if you do, be smart enough to put them in quotation marks and mark them with the source so you don't confuse yourself or your partner later. Both partners will be punished when plagiarism is detected.

Possible theoretical final project topics

Please contact me for any references you cannot find yourself. Note: these projects required you to critically study the literature beyond the scope of the course, some of which is hard and technical. Your report should be critical too. While some cited work below directly opposes work from our group, and we have strong and clear opinions on some of these topics in writing and teaching, you do not have to agree with us and have the freedom to support any opinion you want.

These topics feature prominent discussions of the field, and may therefore lead to interesting papers to others. If your report is good enough, it may be worth sharing online and/or expanding to a publishable paper.

Within- and between-subject variance

Most methods discussed in the course rely on an assumption of independence of cases (Epskamp & Fried, 2017). Such is the case in cross-sectional data, where the responses of one person are likely unrelated from the responses of another person, but may not be the case in time-series data. To this end, we will mostly use cross-sectional data in the course. Many authors criticize cross-sectional data analysis however, especially in the context discovering psychological dynamics (Bos et al., 2017). This line of reasoning can be traced back to the work of Molenaar (2004), who argues that cross-sectional results may not align with analyses based on time-series data. We argue that cross-sectional networks can be interpreted as between-subject effects and may also offer valuable insight (Epskamp, Waldorp, Mõttus, & Borsboom, 2017).

Replicability of networks

Very recently, a group of authors viciously attacked network analysis (Forbes, Wright, Markon, & Krueger, 2017a). By analyzing two similar data-sets, the authors claim to have found "evidence that psychopathology symptom networks do not replicate have limited replicability". We responded both on our blog[1] as well as through an invited commentary (Borsboom et al., 2017), and argue that the authors made severe errors in three out of four network models (all models except Ising), and that further investigation of the Ising model shows stunning replicability (edge parameters correlate with 0.95 and the NetworkComparisonTest shows no significant differences). In addition, we argue that the data are unsuited for this project due to a problematic missing data imputation strategy. Still, the authors argue in their rebuttal that network models "have limited replicability and utility" (Forbes, Wright, Markon, & Krueger, 2017b), now referring to an overview of PTSD studies and a second commentary on their original paper (Steinley, Hoffman, Brusco, & Sher, 2017). We respond in blog posts and a new commentary[2] that Steinley et al. (2017), too, make severe errors and that PTSD replicability has been well studied by our group already (Fried et al., 2017).

[1] http://psychosystems.org/psychopathology-networks-replicate-with-stunning-precision/

[2] Not online yet, but available on request

Factor models and network model equivalences
For over a century, factor models have been applied to psychological questionnaire datasets, with varying degrees of success. Often, a factor model fits decently to well. To this end, any statistical model used on questionnaire data should lead to data on which factor models will also fit. The network models used in this course can all be shown to be mathematically equivalent to some factor model: a fully connected network of edges that are about the same strength corresponds to a uni-dimensional factor model. This was already shown in the first paper of our group (der Maas et al., 2006), and later mathematically proven for the Ising model (Epskamp, Maris, Waldorp, & Borsboom, 2016; Kruis & Maris, 2016; Marsman et al., in press; Maarten Marsman, Maris, Bechger, & Glas, 2015), the Gaussian graphical model (Epskamp, Rhemtulla, & Borsboom, 2017; Golino & Epskamp, 2017), and the graphical vector auto-regression model.[3] This equivalence leads to important consequences. For example, if networks and factor models are equivalent, how can we assess which is the likely data-generating model?

Large-scale simulation studies
Studying a methodology, such as developing network psychometrics, requires (a) mathematical proofs and (b) simulation studies assessing the methodology. Epskamp & Fried (2017) describe a way in which the bootnet package can be used to perform simulation studies, and the new parSim package can further be used to setup more advanced simulation studies.[4] Many questions still remain which may be investigated in large-scale simulation studies. For example, accuracy checks of Epskamp, Borsboom, & Fried (2017) have not been validated yet for regularized Ising models, it is not clear how many observations are needed for graphical vector auto-regression models, the *Network Comparison Test* has not yet been validated for a wider array of network models (Borkulo et al., 2017), more simulations are needed to compare EBIC model selection to cross-validation model selection in the MGM package, the impact of measurement error on all network models is yet unclear, and the best way to handle ordinal data may require more investigation. You may pick a topic of interest and set up a large-scale simulation study. More groups can potentially do this topic as long as the simulation studies differ enough.

Review of topic-specific literature
Various fields of study now feature several publications in which network models are used. The most notable fields are depression, post-traumatic stress disorder (PTSD), schizophrenia, personality and attitude research, which all feature multiple interesting publications. One possible topic is to pick such a field and write a review of network-related publications. Multiple groups can potentially do this topic as long as the specific topic differs for each group (e.g., depression, PTSD). Contact me for starting points in the literature.

[3] Manuscript available on request
[4] https://github.com/SachaEpskamp/parSim

References

Borkulo, C. van, Boschloo, L., Kossakowski, J., Tio, Pi., Schoevers, R., Borsboom, D., & Waldorp, L. (2017). Comparing network structures on three aspects: A permutation test. http://doi.org/10.13140/RG.2.2.29455.38569

Borsboom, D., Fried, E., Epskamp, S., Waldorp, L., Van Borkulo, C., Van Der Maas, H., & Cramer, A. (2017). False alarm? A comprehensive reanalysis of "Evidence that psychopathology symptom networks have limited replicability" by Forbes, Wright, Markon, and Krueger. *Journal of Abnormal Psychology*. http://doi.org/10.17605/OSF.IO/TGEZ8

Bos, F. M., Snippe, E., de Vos, S., Hartmann, J. A., Simons, C. J. P., van der Krieke, L., … Wichers, M. (2017). Can We Jump from Cross-Sectional to Dynamic Interpretations of Networks? Implications for the Network Perspective in Psychiatry. *Psychotherapy and Psychosomatics*, *86*(3), 175–177. http://doi.org/10.1159/000453583

der Maas, H. L. J., Dolan, C. V, Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842–861.

Epskamp, S., Borsboom, D., & Fried, E. I. (2017). Estimating psychological networks and their accuracy: a tutorial paper. *Behavior Res*. http://doi.org/10.3758/s13428-017-0862-1

Epskamp, S., & Fried, E. I. (2017). A tutorial on Estimating Regularized Psychological Networks. Retrieved from http://arxiv.org/abs/1607.01367

Epskamp, S., Maris, G., Waldorp, L. J., & Borsboom, D. (2016). Network Psychometrics. In P. Irwing, D. Hughes, & T. Booth (Eds.), *Handbook of Psychometrics*. New York, NY, USA: Wiley. Retrieved from https://arxiv.org/abs/1609.02818

Epskamp, S., Rhemtulla, M. T., & Borsboom, D. (2017). Generalized Network Psychometrics: Combining Network and Latent Variable Models. *Psychometrika*. http://doi.org/10.1007/s11336-017-9557-x

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2017). Discovering Psychological Dynamics: The Gaussian graphical model. Retrieved from http://arxiv.org/abs/1609.04156

Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2017a). Evidence that Psychopathology Symptom Networks have Limited Replicability. *Journal of Abnormal Psychology*.

Forbes, M. K., Wright, A. G. C., Markon, K. E., & Krueger, R. F. (2017b). Further evidence that psychopathology networks have limited replicability and utility: Response to Borsboom et al. and Steinley et al. *Journal of Abnormal Psychology*.

Fried, E., Eidhof, M. B., Palic, S., Costantini, G., Huisman-van Dijk, H. M., Bockting, C. L. H., … Karstoft, K.-I. (2017). Replicability and generalizability of PTSD networks: A cross-cultural multisite study of PTSD symptoms in four trauma patient samples. *Clinical Psychological Science*. http://doi.org/10.17605/osf.io/2t7qp

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. http://doi.org/10.1371/journal.pone.0174035

Kruis, J., & Maris, G. (2016). Three representations of the Ising model. *Scientific*

Reports, *6*, 34175. http://doi.org/10.1038/srep34175

Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., Bork, V., Waldorp, L., … Marsman, M. (n.d.). An Introduction to Network Psychometrics: Relating Ising Network Models to Item Response Theory Models. *Multivariate Behaviorial Research*. Retrieved from http://maartenmarsman.com/wp-content/uploads/2017/09/MarsmanEtAlNW.pdf

Marsman, M., Maris, G., Bechger, T., & Glas, C. (2015). Bayesian inference for low-rank Ising networks. *Scientific Reports*, *5*(9050), 1–7.

Molenaar, P. C. M. (2004). A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Measurement: Interdisciplinary Research & Perspective*, *2*(4), 201–218. http://doi.org/10.1207/s15366359mea0204_1

Steinley, D., Hoffman, M., Brusco, M. J., & Sher, K. J. (2017). A Method for Making Inferences in Network Analysis: Comment on Forbes, Wright, Markon, and Krueger (2017). *Journal of Abnormal Psychology*.