Assignment 3 - Network Analysis 2022

Kyuri Park, 12183881

December 01, 2022

Contents

Conceptual Questions

Question 1 (1 point)

Are the following statements true or false? Explain why (0.5 point per statement).

1. If ergodicity holds, results from between-person analysis are expected to equal results from within-person analysis.

   Yes, if ergodicity holds, there should be no difference in between-person level, and correspondingly it aligns with the within-person analysis of every individual (Molenaar, 2004).

2. It is generally recommended to always remove trends (such as linear trends) prior to analyzing your N = 1 time series.

   No, even though some simulation stuides show detrending helps increasing the performance of estimated networks (Epskamp et al., 2018), it is often the case that these changing trends are of main interest. So it is not correct to say it is *always* recommended to detrend (Isvoranu et al., 2022).

Question 2 (1.5 points)

Suppose a therapist measures a patient about 75 times on a set of depression symptoms, including a question on "suicidal thought". Suppose that you estimate a graphical VAR model from this data and find that the node "suicidal thought" is not connected to any of the other nodes in your network, neither in the temporal nor in the contemporaneous network.

List three potential reasons why the node "suicidal thought" may be disconnected in the resulting network.

1. Probably the patient answered "no" every time on "suicidal thought, since it is quite an extreme symptom to develop. It is likely that the patient never developed"suicidal thought" and that leads to no variance, which will result in no connections in the network.

2. It could be the case that the data is missing completely on "suicidal thought". For example, the patient felt pressure or some sorts of discomfort to answer the question on "suicidal thought" and left the question always unanswered. Then, the item would have only *NAs*, which would likely to result in no variation on that item.

3. Lastly, another potential scenario is that the measurement period was too short to capture the variance in "suicidal thought". For example, the 75 times of measurement was performed every 2 hrs in a day over 2 weeks or so, which is quite a short period of time. It could be that the patient only started having "suicidal thought" after the measurement period and hence left the item score on "suicidal thought" constant.

Question 3 (1 point)

Give an example of a relationship that can only be studied at the *between-person* level.

Things that can only be studied at the *between-person* level have to be something that stay stable and do not change over time (i.e., no deviations / fluctuations over time). One example relationship that can be studied only in *between-person* level is: the relationship between the education level and different races. (place of birth, sex at birth, gene?..., nationality)

Question 4 (2 points)

During the lecture we discussed multiple challenges regarding time-series modeling in the network approach. Pick your favorite challenge and explain this challenge in your own words. Make sure to not only explain what the challenge is, but also why this is a challenge: in what way may this challenge impact your results (i.e., the network model you estimate) and how may this jeopardize your conclusions? You can use additional literature to back up your argument, but this is not required. If you choose to use additional literature, make sure to include a reference list (max 250 words).

One of the challenges in time-series network modelling is the difficulty of incorporating variables that operate on different time-scale. For example, the symptom network (e.g., network consist of symptoms) evolves relatively fast, as symptoms generally fluctuate over days or weeks. Whereas, psychological resilience, which is an important protective factor against developing mental disorder, evolves rather slowly and gradually (Lunasky, 2020). Obviously, resilience is part of the dynamic of psychopathology, but there exists no such model that can incorporate variables operating on the different time-scale and enable us to analyze the whole network that integrates the slow-changing variable (e.g., resilience) with the fast-changing symptoms. Note that Lunasky et al.(2020) look into the interaction between slow and fast network processes but it was studied in such a way that have the slow changing variable influence the node in the symptom network but not incorporating them all in one network model, per se.

According to Borsboom (2021), omitting this relevant factor that operate on slower time-scale from the symptom network could lead to a much denser network. The reasoning is as follows. If the slow-changing variable is omitted in the network model, then the edges will be estimated without controlling for the SV. Assuming that SV has consistent influences on the symptoms, not controlling for the SV is likely to exaggerate the edges (i.e., partial correlations) between the symptoms. Therefore, it is expected that the symptom network model excluding the SV would appear denser (i.e., many thicker edges) than the integrated network model including the SV.

Hence, basing the network model only on the symptoms without incorporating the relevant SV can result in misleading model that overestimates the strength of relationships between symptoms. This could be especially detrimental to studies that look into a particular connection between two symptoms. If a SV that has a considerable influence on the symptom dynamics is not included in the model, the researcher may draw a conclusion based on one of the spurious connections in the network, which are resulted from overlooking the effects of the SV.[1]

Question 5 (2 points)

In a recent study by Haslbeck et al., (preprint) it was shown that use of a VAS or Likert scale has an affect on the observed distributions. Take a look at their article. Give (a) a short summary of the problem described in their paper, (b) reflect on this problem (e.g., why is this an issue and how does it affect the interpretation of our results?) and (c) think of a study to specifically test whether this phenomenon (different scales lead to different distributions) is a methodological artifact or a "true" phenomenon (max 250 words).

Question 6 (1 point)

What do the *day* and *beep* arguments in the packages `graphicalVAR`, `psychonetrics`, and `mlVAR` do? Suppose you have a dataset containing only one observation per day for every weekday (but not weekends) in a year long study. Which argument would you use (and why) to make sure that measures from Mondays are not regressed on measures from Fridays?

---

[1] ***Note.*** Reference list can be found at the end of the document.

Practical Questions

```r
# load data
load("clean_network.RData")

# Variables to investigate:
vars <- paste0("Q",1:18)
# Labels:
varLabs <- c("Relax","Irritable","Worry","Nervous","Future","Anhedonia",
"Tired","Hungry","Alone","Angry","Social_offline","Social_online",
"Music","Procrastinate","Outdoors","C19_occupied","C19_worry",
"Home")

# Rename columns in data:
names(Data2)[names(Data2) %in% vars] <- varLabs

# Remove items:
Data2 <- Data2 %>% select(-Hungry,-Angry,-Music,-Procrastinate)
varLabs <- varLabs[!varLabs %in% c("Hungry","Angry","Music","Procrastinate")]
```

N =1 time series

```r
student_number <- 12183881 # Enter student number here
set.seed(student_number)
subject <- sample(Data2$id, 1)
my_data <- Data2[which(Data2$id == subject),]
```

Question 7 (1.5 points)

Estimate a *saturated* (no model selection) GVAR model on your sampled subject using the 3 or 4 variables you chose above using `psychonetrics`. Plot the estimated temporal (partial directed correlations) and contemporaneous networks with the same layout (1 point). Which edges are significant (0.5 point)?

Tips:

- You can safely ignore warnings on non-positive definite matrices as long as the parameters you obtain look reasonable (e.g., no partial correlations near −1 and 1).

- The psychonetrics package models temporal effects using a matrix called beta, but also returns a standardized form back called PDC (partial directed correlations). Note that the PDC matrix is the standardized transpose of the beta matrix, allowing it to be plotted as a directed network. The significance of parameters in the beta matrix corresponds to the significance of parameters in the PDC matrix.

Question 8 (1 point)

Estimate a GVAR model on your sampled subject using the 3 or 4 variables you chose above using the graphicalVAR package (use $\gamma = 0$), and compare your results to the results of the previous question.

4

Question 9 (1.5 points)

We can use the time variable (in this case we use `conc` as a proxy for the time of measurement) to check for trends in the data. To test if the variable "C19_worry" features a trend, we can run a linear regression on time:

```
lm_c19_worry <- lm(C19_worry ~ conc, data = my_data)
summary(lm_c19_worry)
```

```
Call:
lm(formula = C19_worry ~ conc, data = my_data)

Residuals:
     Min       1Q   Median       3Q      Max
-1.03552 -0.30191 -0.01976  0.16423  1.32564

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.103235   0.152314   13.81  < 2e-16 ***
conc        -0.022572   0.004542   -4.97 8.58e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5011 on 49 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.3351,    Adjusted R-squared:  0.3216
F-statistic:  24.7 on 1 and 49 DF,  p-value: 8.576e-06
```

This effect is significant. We can now "detrend" the variable worry as follows:

```
my_data$C19_worry[!is.na(my_data$C19_worry)] <- residuals(lm_c19_worry)
```

Test for significant trends ($\alpha = 0.05$) for your selected variables and detrend these variables if the trends are significant. Then, re-estimate the networks using either `graphicalVAR` or `psychonetrics`. Did your estimated networks change?

Note: if the participant you picked did not show any significant trends for the variables you picked, either pick a different participant or pick new variables that do show a significant trend.

Question 10 (1 point)

For question 8 you estimated a graphical VAR network containing 3 or 4 variables for a sampled subject using the `graphicalVAR` package. Repeat this process for another subject, i.e., estimate an additional graphical VAR network containing the same 3 or 4 variables as before for another subject (for example, use the student number of a friend). Note here that you are not required to detrend the data prior to estimating the network structures. Now that you have estimated two graphical VAR models on non-detrended data, plot the resulting networks and inspect the similarities and differences between these estimated graphical VAR networks. Can you conclude there is heterogeneity between these two people? Why (not)?

N > 1 time-series

Question 11 (1 point)

Look at the help file for the `mlVAR` function and estimate a multilevel GVAR model on the entire dataset on 4 to 6 variables of your choice.

Question 12 (1 point)

Look at the mlVAR plot method help file (`?plot.mlVAR`), and plot the estimated fixed-effect temporal, contemporaneous and between-subjects networks. Plot all networks with the same layout (circle layout or average layout), and hide (threshold) edges that are not significant at $\alpha = 0.05$. For the contemporaneous and between-subjects networks, use an "and"-rule to minimize type-1 error rate in showing edges.

Question 13 (2 points)

Plot the estimated individual differences for both the temporal and the contemporaneous network. Explain what the edges in these networks represent. Inspect the plotted networks, what conclusion can you draw about heterogeneity within this sample based on these network structures?

Tip: Take a look at Bringmann et al., (2013)

Essay Question

Question 14 (2 points)

Let's revisit centrality measures: As we already explored in previous assignments, `qgraph` implements a number of centrality measures for networks. In addition, the Brain Connectivity Toolbox (see https://sites.google.com/site/bctnet/home) contains measures to inspect complex network structures. Inspect these network measures and choose one that would be interesting to compute on one of the estimated networks from any of the previous questions. Report your chosen measure and what it represents for the network you have chosen, explain your choice, and what research question it could answer in the form of a short essay (max 250 words).

Note: you do not need to compute the metric as not all of the centrality measures are also implemented into R (you are of course free to compute the measure nonetheless!).

References

Epskamp, S., Borkulo, C. D. van, Veen, D. C. van der, Servaas, M. N., Isvoranu, A.-M., Riese, H., & Cramer, A. O. (2018). Personalized network modeling in psychopathology: The importance of contemporaneous and temporal connections. *Clinical Psychological Science*, *6*(3), 416–427.

Isvoranu, A.-M., Epskamp, S., Waldorp, L., & Borsboom, D. (2022). *Network psychometrics with r: A guide for behavioral and social scientists*. Routledge.

Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*(4), 201–218.