Network Analysis 2017
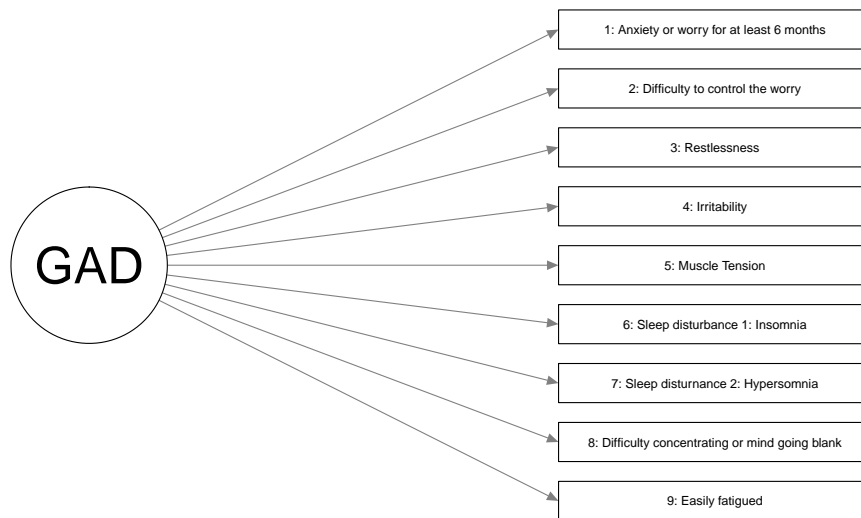
Assignment 4

Conceptual Questions

Suppose the folowing model is the *true* data-generating model of nine symptoms:

GAD	→	1: Anxiety or worry for at least 6 months
	→	2: Difficulty to control the worry
	→	3: Restlessness
	→	4: Irritability
	→	5: Muscle Tension
	→	6: Sleep disturbance 1: Insomnia
	→	7: Sleep disturnance 2: Hypersomnia
	→	8: Difficulty concentrating or mind going blank
	→	9: Easily fatigued

The circle indicates that generalized anxiety disorder (GAD) is latent and thus not observed.

Question 1 (2 points)

Theoretically, what should the Gaussian graphical model (GGM; a network of partial correlation co-efficients) of the nine observed symptoms look like? Plot or draw the expected network (use only the numbers as node labels). Tip: Note that GAD is not observed and thus (a) not in the GGM and (b) not a node we can condition on.

■

Suppose we are interested in investigating the network structure of people suffering from GAD. We follow the DSM and select only participants in our dataset that have at least 5 of the 9 symptoms.

Question 2 (2 points)

Can you explain why such a selection method might be problematic?

■

Question 3 (2 points)

One often-used interpretation of edges in a Markov random field is that these may be indicative of potential causal pathways. Do you agree? Why (not)? Note that there is no right or wrong answer to this question and the question will be rated based on the quality of your arguments.

■

Partial correlation networks

Run the following codes in R:

```r
library("psych")
data(bfi)
bfiNoNA <- na.omit(bfi)
bfiSumScores <- data.frame(
  Agreeableness = rowSums(bfiNoNA[,1:5]),
  Conscientiousness = rowSums(bfiNoNA[,6:10]),
  Extraversion = rowSums(bfiNoNA[,11:15]),
  Neuroticism = rowSums(bfiNoNA[,16:20]),
  Openness = rowSums(bfiNoNA[,21:25])
)
```

This creates a data frame `bfiSumScores` that contains the factor scores of the big 5 inventory data (we used in assignment 1).

Question 4 (1 point)
Use the `partial.r` function from the *psych* package to compute the partial correlations, conditioning on all remaining variables, between (round your results to two digits):
- Agreeableness and Conscientiousness
- Agreeableness and Extraversion

We can compute the sample variance–covariance matrix in R using the `cov` function:

```
covMat <- cov(bfiSumScores)
round(covMat,2)

##                    Agreeableness Conscientiousness Extraversion Neuroticism
## Agreeableness              13.23              0.53         2.92       -2.76
## Conscientiousness           0.53              7.56         1.34        4.14
## Extraversion                2.92              1.34         7.16        0.49
## Neuroticism                -2.76              4.14         0.49       35.73
## Openness                    1.71              1.69         1.75        2.65
##                  Openness
## Agreeableness        1.71
## Conscientiousness    1.69
## Extraversion         1.75
## Neuroticism          2.65
## Openness             7.03
```

These covariances are hard to interpret and we rather standardize the results to *correlation coefficients*. Let R denote the correlation matrix and S the variance–covariance matrix. Then, we can obtain the correlation between variables i and j, element r_{ij} of the matrix R, as follows:

$$r_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}.$$

Question 5 (1 point)
Use the formula above to compute the correlation between Agreeableness and Conscientiousness. Then, use the `cov2cor` function to compute all correlations.

We can use `solve(covMat)` to *invert* a variance–covariance matrix:

$$K = S^{-1}.$$

This *precision matrix* $\mathbf{K}$ (kappa) can be standardized to obtain the partial correlation coefficient matrix $\boldsymbol{P}$ (the partial correlation between variable i and j after conditioning on all other variables in the dataset):

$$p_{ij} = \begin{cases} -\dfrac{\kappa_{ij}}{\sqrt{\kappa_{ii}}\sqrt{\kappa_{jj}}} & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}.$$

> **Question 6 (2 points)**
>
> Use the formula above to compute the partial correlation between Agreeableness and Conscientiousness. Then, compute all partial correlations (note the similarity between the expression above and standardizing covariances to correlations). Plot the resulting partial correlation matrix as a network using the *qgraph* function, and compare your results to your answer in question 1.

We can also perform multiple regressions. For example, we can regress Agreeableness on all other variables:

```r
# Fit regression model:
fitA <- lm(Agreeableness ~ Conscientiousness + Extraversion + Neuroticism + Openness,
           data = bfiSumScores)

# Obtain coefficients:
coefA <- coef(fitA)

# Obtain residual standard deviation:
sdA <- sigma(fitA)
```

> **Question 7 (2 points)**
>
> Perform a multiple regression of Conscientiousness on all other variables, and confirm:
>
> $$\rho_{ij} = \frac{\gamma_{ij}\,\mathrm{SD}\,(\varepsilon_j)}{\mathrm{SD}\,(\varepsilon_i)} = \frac{\gamma_{ji}\,\mathrm{SD}\,(\varepsilon_i)}{\mathrm{SD}\,(\varepsilon_j)}$$
>
> Tip: If $i = 1$ indicates Agreeableness, then γ_{ij} is the R object `coefA[2]` and $\mathrm{SD}\,(\varepsilon_i)$ is the R object `sdA`

> **Question 8 (1 point)**
>
> Using `summary` on the results of both regression models you ran above, do you judge the partial correlation between Agreeableness and Conscientiousness significant?

To make things simpler, we can use the *bootnet* package:

```r
library("bootnet")

## Loading required package:  ggplot2
##
## Attaching package:  'ggplot2'
## The following objects are masked from 'package:psych':
##
##     %+%, alpha
## This is bootnet 1.0.1
## For questions and issues, please see github.com/SachaEpskamp/bootnet.

# Estimate network:
Network <- estimateNetwork(bfiSumScores, default = "pcor")
```
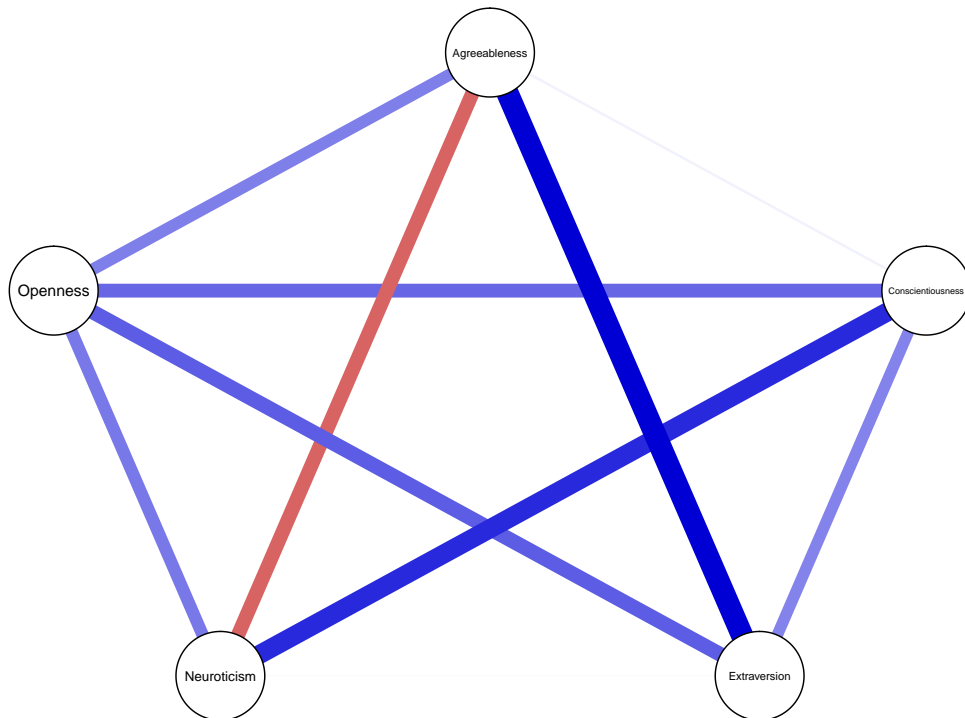
```
## Estimating Network.   Using package::function:
##  - qgraph::qgraph(..., graph = 'pcor') for network computation
##  - qgraph::cor_auto for correlation computation
##     - using lavaan::lavCor
##  - psych::corr.p for significance thresholding

# Weights matrix:
Network$graph

##             Agr        Cns         Ext          Nrt        Opn
## Agr   0.0000000 0.0140859 0.266929542 -0.163656312 0.1344642
## Cns   0.0140859 0.0000000 0.129378030  0.225074249 0.1585909
## Ext   0.2669295 0.1293780 0.000000000  0.004107258 0.1708875
## Nrt  -0.1636563 0.2250742 0.004107258  0.000000000 0.1412214
## Opn   0.1344642 0.1585909 0.170887513  0.141221411 0.0000000

# Plot network (defaults to colorblind theme)
plot(Network, layout = "circle")
```


Question 9 (1 point)

Look at the help page of `estimateNetwork` and find a way to remove partial correlations that are not significant after a bonferroni corrected significance test. How many edges are then removed? Plot the resulting network.

Ising models

Load the dataset we used last week in the assignment:

```
datafile <- read.table(file='http://borsboomdenny.googlepages.com/datafile.txt')
```

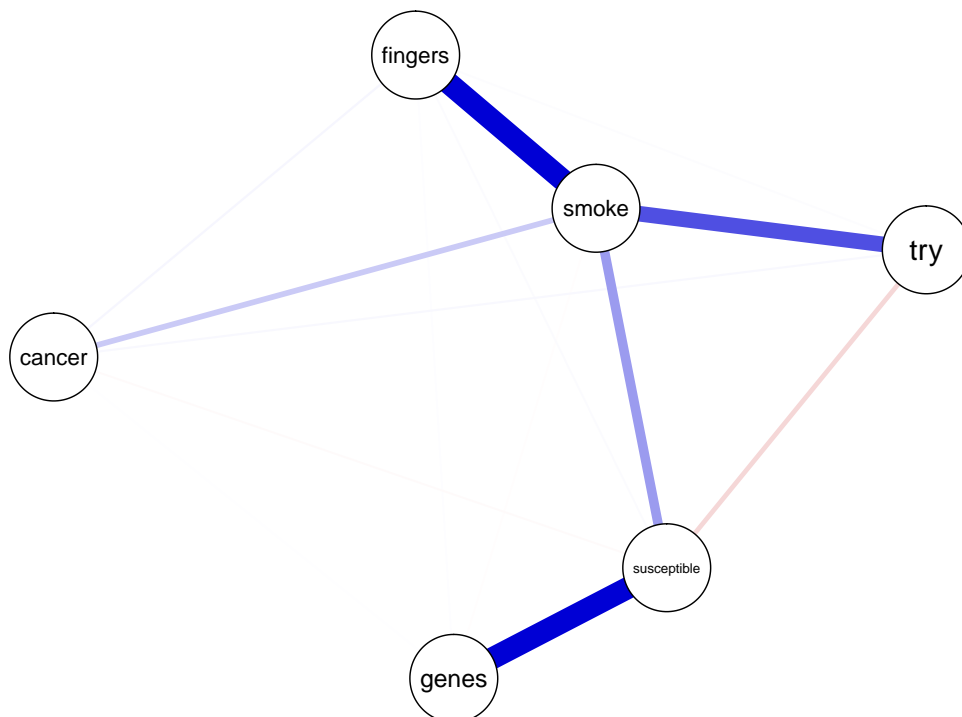We can perform a logistic regression of smoke on all other variables as follows:

```
fit_try <- glm(smoke ~ try +  susceptible  + fingers + genes + cancer,
               family = "binomial", data = datafile)
```

> **Question 10 (2 points)**
> Perform nodewise regressions for each variable given all other variables to obtain all edge weights of the Ising model. You get two estimates for each edge, which you should average to obtain a single estimate per edge. Combine these in a matrix, and plot the corresponding graph using *qgraph*.

We can obtain the Ising model also using bootnet:

```
IsingNet <- estimateNetwork(datafile, default = "IsingSampler", method= "uni")

## Estimating Network.   Using package::function:
##  - IsingSampler::EstimateIsing for network computation

plot(IsingNet)
```


> **Question 11 (1 point)**
> Compare the estimated Ising model to the solutions of last week. Does the Ising model look like you would expect? Can you explain why there is a negative edge between try and susceptible?

Challenge question (1 bonus point)

We can obtain a table with the probability of all possible response patterns using the *IsingSampler* package:

```
library("IsingSampler")

## Loading required package: Rcpp

Table <- IsingLikelihood(IsingNet$graph,IsingNet$intercepts, beta = 1)
head(Table)

##    Probability Var1 Var2 Var3 Var4 Var5 Var6
## 1  0.18055426    0    0    0    0    0    0
## 2  0.04266869    1    0    0    0    0    0
## 3  0.03755325    0    1    0    0    0    0
## 4  0.15045443    1    1    0    0    0    0
## 5  0.04398324    0    0    1    0    0    0
## 6  0.01034250    1    0    1    0    0    0
```

> **Question 12**
>
> Compute the probability of:
> - A random person in the sample having cancer
> - A random person in the sample having cancer, given that we know that person smokes
> - A random person in the sample having cancer, given that we know that person smokes and has stained fingers
>
> Did the probability substantially (more than 1%) chance after conditioning on smoking? And when we also conditioned on stained fingers? Can you explain this result?