Bayesian analysis of cross-sectional networks — Assignment in R with Solutions

Maarten Marsman & Nikola Sekulovski

Learning goals

After completing this assignment, you will be able to:

1. Test for edge presence or absence with the inclusion Bayes factor.
2. Estimate a cross-sectional network in a Bayesian framework with or without edge selection.
3. Visualize the results from the Bayesian analysis with a network plot and an edge evidence plot.
4. Interpret the results from a Bayesian analysis, such as the Bayes factor and credible interval.
5. Convey to others the advantages of using the Bayesian approach to network analysis.

Background Literature

For more background information, you can have a look at:

Huth, K. B. S., de Ron, J., Goudriaan, A. E., Luigjes, J., Mohammadi, R., van Holst, R. J., Wagenmakers, E.-J., & Marsman, M. (2023). Bayesian analysis of cross-sectional networks: A tutorial in R and JASP. *Advances in Methods and Practices in Psychological Science, 6*(4).

Huth, K., Keetelaar, S., Sekulovski, N., van den Bergh, D., & Marsman, M. (2024). Simplifying Bayesian analysis of graphical models for the social sciences with easybgm: A user-friendly R-package. *Advances in Psychology*, e66366.

Sekulovski, N., Keetelaar, S., Huth, K. B. S., Wagenmakers, E.-J., van Bork, R., van den Bergh, D., & Marsman, M. (in press). Testing Conditional Independence in Psychometric Networks: An Analysis of Three Bayesian Methods. *Multivariate Behavioral Research, 59*, 913-933.

Setup

Before starting the task in R, you need to install and load some packages. We have created the package `easybgm` to help you easily fit and visualize the results of your analyses. For an example analysis, you can take a look at the GitBook provided in the PSYNETS workshop materials. To avoid possible installation problems, please make sure you have the latest version of R installed. As of January 2025, the latest version of R is 4.4.2.

```r
# Install the required packages with install.packages("package") and
# load them using library("package")
library(readr)
library(BDgraph)
library(qgraph)
library(bgms)
library(easybgm)
```

Part 1: Analyze an Ordinal Network Model

The Data

This data is a subset from an online version of the Sexual Compulsivity Scale (Kalichman & Rompa, 1995), which is openly available at https://openpsychometrics.org/_rawdata/. The items were rated on a Likert scale (1 = Not at all like me, 2 = Slightly like me, 3 = Mainly like me, 4 = Very much like me):

- Q1: My sexual appetite has gotten in the way of my relationships.
- Q2: My sexual thoughts and behaviors are causing problems in my life.
- Q3: My desires to have sex have disrupted my daily life.
- Q4: I sometimes fail to meet my commitments and responsibilities because of my sexual behaviors.
- Q5: I sometimes get so horny I could lose control.
- Q6: I find myself thinking about sex while at work.
- Q7: I feel that sexual thoughts and feelings are stronger than I am.
- Q8: I have to struggle to control my sexual thoughts and behavior.
- Q9: I think about sex more than I would like to.
- Q10: It has been difficult for me to find sex partners who desire having sex as much as I want to.

Kalichman, S. C., & Rompa, D. (1995). Sexual Sensation Seeking and Sexual Compulsivity Scales: Validity, and Predicting HIV Risk Behavior. *Journal of Personality Assessment, 65*(3), 586–601.

```r
# load the data
data <- read_csv("data_scs.csv")
```

Part 1: Testing for conditional (in)dependence (Is there an effect?)

Task 1.1: Analyze the graphical model

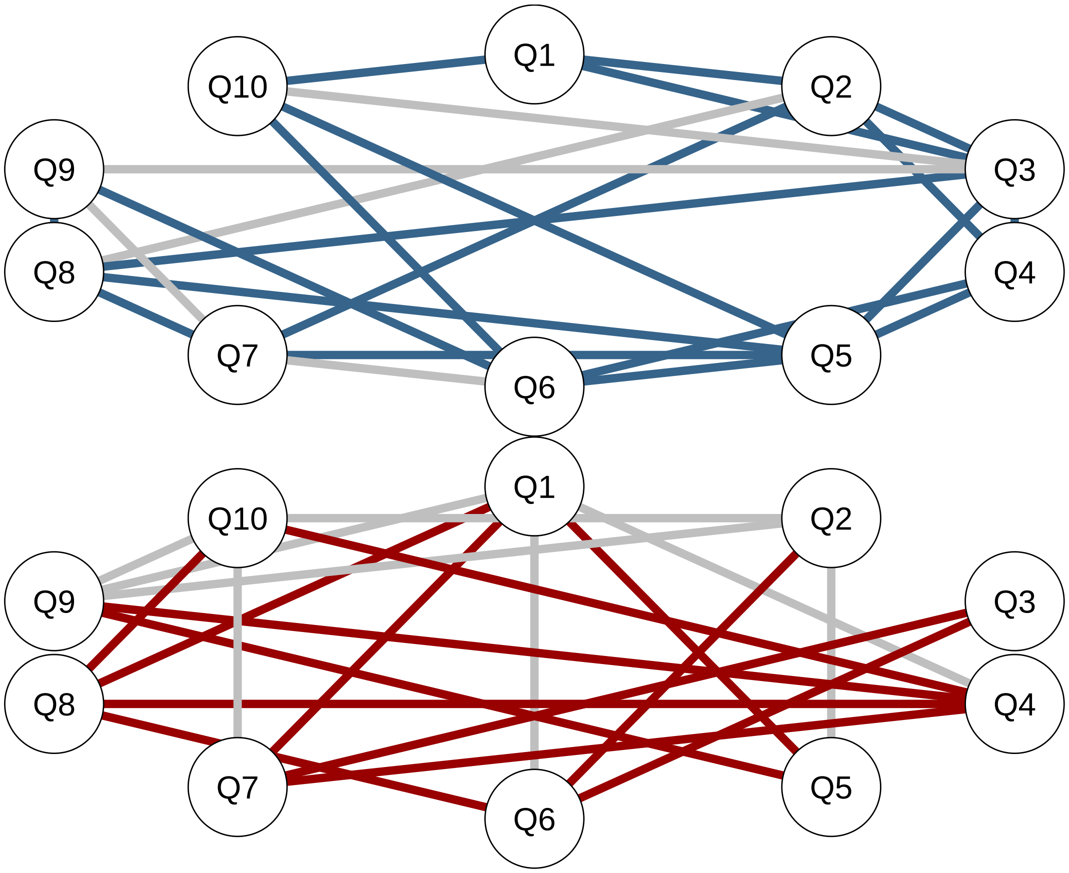
See the help file of the `easybgm` package to familiarize yourself with the input arguments. Estimate a network model with `easybgm`. The data are ordinal, so be sure to specify this in the `type` argument.

```r
fit <- easybgm(data = data,
type="ordinal",
iter = 1e4,
save = TRUE
)
```

Task 1.2: Extract the edge evidence plot and interpret the results

Compute the edge evidence plot with `plot_edgeevidence`. Take a look at the `split` argument, which may help with the interpretation. The `plot_edgeevidence` function also takes the same arguments as `qgraph`.

```
par(mfrow=c(2,1))
plot_edgeevidence(fit,
                  edge.width = 3,
                  split = T,
                  legend = F,
                  layout = "circle")
```



- What does the edge evidence plot show?
- How do you interpret an edge inclusion Bayes factor? What is the default threshold value used in the `plot_edgeevidence function`?

The Bayes factor compares the relative evidence for two competing hypotheses. Is the data more likely under one of the two hypotheses? The Inclusion Bayes factor pits the hypothesis that the edge should be included in the network against the hypothesis that the edge should be excluded. The default value used in plot_edgeevidence is 10. An inclusion Bayes factor of 10 indicates that the observed data is ten times more likely to come from a network that includes the edge than from a network that excludes the edge. Conversely, an inclusion Bayes factor of 1/10 indicates that the observed data are ten times more likely to come from a network that excludes the edge than from a network that includes the edge. If the Bayes factor equals one, then the data are equally likely under both hypotheses.

Part 2: Estimation (What is the size of the effect?)

Once we have tested for the presence of edges, the next step is to estimate the parameters for the present edges. We can do this in one of two ways. The first way is consistent with the way we tested for conditional independence and takes into account both structure and parameter uncertainty by using Bayesian variable selection methods. The second method estimates the parameters given only a single network structure — the fully connected structure. For more details on Bayesian model averaging, of which Bayesian variable selection is a special case, you can take a look at the following paper:

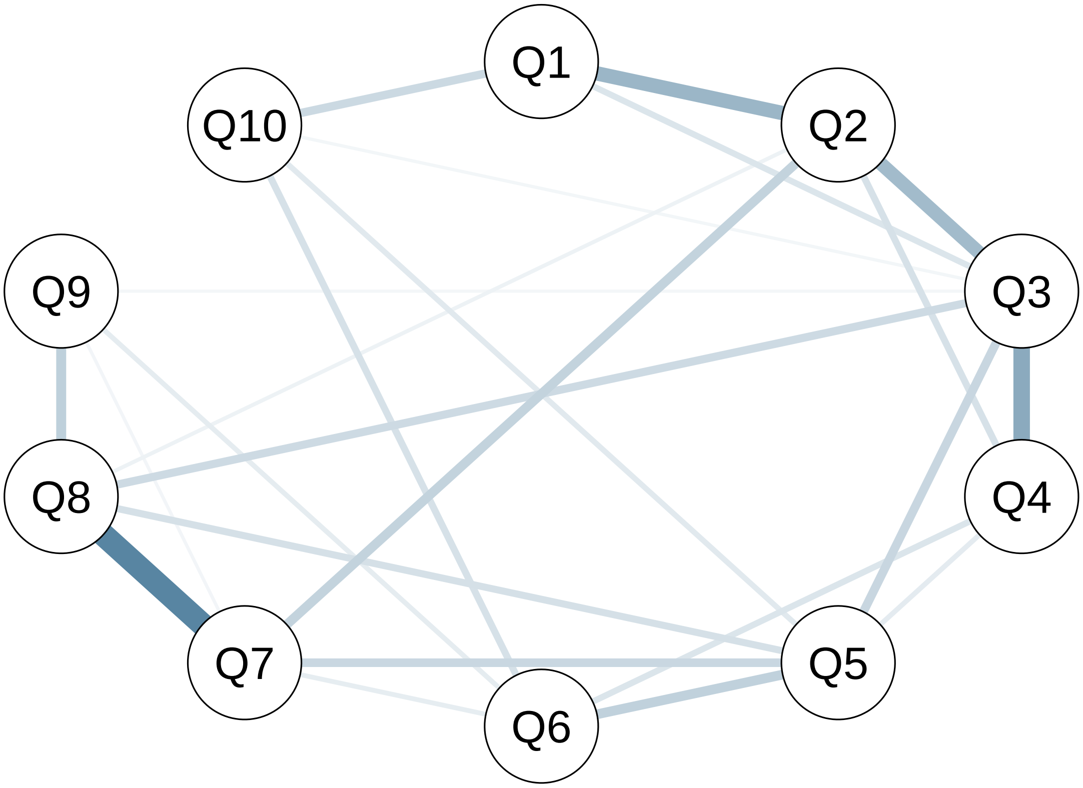
Hinne, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E. J. (2020). A Conceptual Introduction to Bayesian Model Averaging. *Advances in Methods and Practices in Psychological Science, 3*(2), 200-215.

Task 2.1: Estimate the parameters with edge selection

The object with the results from Part 1 already contains the averaged model parameters using edge selection, therefore you can use this object to answer the following questions.

- Visualize the network using `plot_network`, which, by default, will give you the median probability model.

```
plot_network(fit, legend = FALSE, layout = "circle")
```



- What conclusions can you draw from the estimated network?
- What does it mean that this network is based on the median probability model, how does this relate to the inclusion Bayes factors which you have calculated in Part 1? *The network shows positive connections, with all nodes connected to at least one other node. The strongest link is between Q7 and Q8. The network appears not to be too densely connected, which is consistent with the edge evidence plots from the previous part.*
- You can also use the `summary` function to get a convenient table with the estimated values of these parameters as well as the values of the inclusion Bayes factors and posterior probabilities for edge inclusion. There will be `inf` values for the inclusion Bayes factors for some of the edges; this simply

means that the *estimated* posterior inclusion probability is 1, so the Bayes factor calculation is undefined since the denominator is zero.

```r
summary(fit)
```

```
##
##  BAYESIAN ANALYSIS OF NETWORKS
##  Model type: ordinal
##  Number of nodes: 10
##  Fitting Package: bgms
## ---
##  EDGE SPECIFIC OVERVIEW
##  Relation Estimate Posterior Incl. Prob. Inclusion BF     Category
##     Q1-Q2    0.564                 1.000          Inf     included
##     Q1-Q3    0.200                 0.966       28.070     included
##     Q2-Q3    0.526                 1.000          Inf     included
##     Q1-Q4    0.017                 0.154        0.182 inconclusive
##     Q2-Q4    0.229                 0.996      269.270     included
##     Q3-Q4    0.655                 1.000          Inf     included
##     Q1-Q5    0.000                 0.013        0.013     excluded
##     Q2-Q5   -0.039                 0.278        0.385 inconclusive
##     Q3-Q5    0.313                 1.000          Inf     included
##     Q4-Q5    0.153                 0.916       10.891     included
##     Q1-Q6    0.017                 0.166        0.199 inconclusive
##     Q2-Q6    0.000                 0.017        0.018     excluded
##     Q3-Q6    0.000                 0.008        0.008     excluded
##     Q4-Q6    0.203                 0.958       23.038     included
##     Q5-Q6    0.356                 1.000          Inf     included
##     Q1-Q7    0.000                 0.014        0.015     excluded
##     Q2-Q7    0.340                 1.000          Inf     included
##     Q3-Q7   -0.006                 0.056        0.060     excluded
##     Q4-Q7    0.000                 0.013        0.013     excluded
##     Q5-Q7    0.307                 1.000          Inf     included
##     Q6-Q7    0.140                 0.790        3.762 inconclusive
##     Q1-Q8    0.001                 0.020        0.020     excluded
##     Q2-Q8    0.101                 0.582        1.393 inconclusive
##     Q3-Q8    0.285                 1.000          Inf     included
##     Q4-Q8    0.006                 0.064        0.069     excluded
##     Q5-Q8    0.238                 1.000          Inf     included
##     Q6-Q8    0.001                 0.015        0.015     excluded
##     Q7-Q8    0.942                 1.000          Inf     included
##     Q1-Q9    0.013                 0.147        0.173 inconclusive
##     Q2-Q9    0.050                 0.408        0.691 inconclusive
##     Q3-Q9    0.067                 0.511        1.045 inconclusive
##     Q4-Q9    0.000                 0.007        0.007     excluded
##     Q5-Q9   -0.001                 0.016        0.017     excluded
##     Q6-Q9    0.147                 0.982       54.866     included
##     Q7-Q9    0.074                 0.593        1.457 inconclusive
##     Q8-Q9    0.367                 1.000          Inf     included
##    Q1-Q10    0.296                 1.000          Inf     included
##    Q2-Q10   -0.032                 0.266        0.362 inconclusive
##    Q3-Q10    0.072                 0.537        1.159 inconclusive
##    Q4-Q10    0.000                 0.014        0.015     excluded
##    Q5-Q10    0.172                 0.987       78.365     included
##    Q6-Q10    0.232                 1.000         Inf      included
##    Q7-Q10   -0.023                 0.219       0.280 inconclusive
##    Q8-Q10    0.001                 0.023       0.024      excluded
##    Q9-Q10    0.011                 0.148       0.174 inconclusive
##
##  Bayes Factors larger than 10 were considered sufficient evidence for the classification
##  Bayes factors were obtained using Bayesian model-averaging.
##  ---
##  AGGREGATED EDGE OVERVIEW
##  Number of edges with sufficient evidence for inclusion: 19
##  Number of edges with insufficient evidence: 13
##  Number of edges with sufficient evidence for exclusion: 13
##  Number of possible edges: 45
##
##  ---
##  STRUCTURE OVERVIEW
##  Number of visited structures: 2116
##  Number of possible structures: 3.518437e+13
##  Posterior probability of most likely structure: 0.027
## ---
```

Task 2.2: Estimate the parameters without edge selection

For this you need to specify the argument `edge_selection = FALSE` in the `easybgm` function. You can again use the `summary` function to see the estimated parameters. Notice that this table only contains the values of the estimated parameters, and not the inclusion Bayes factors and posterior inclusion probabilities. Can you explain why this is the case?

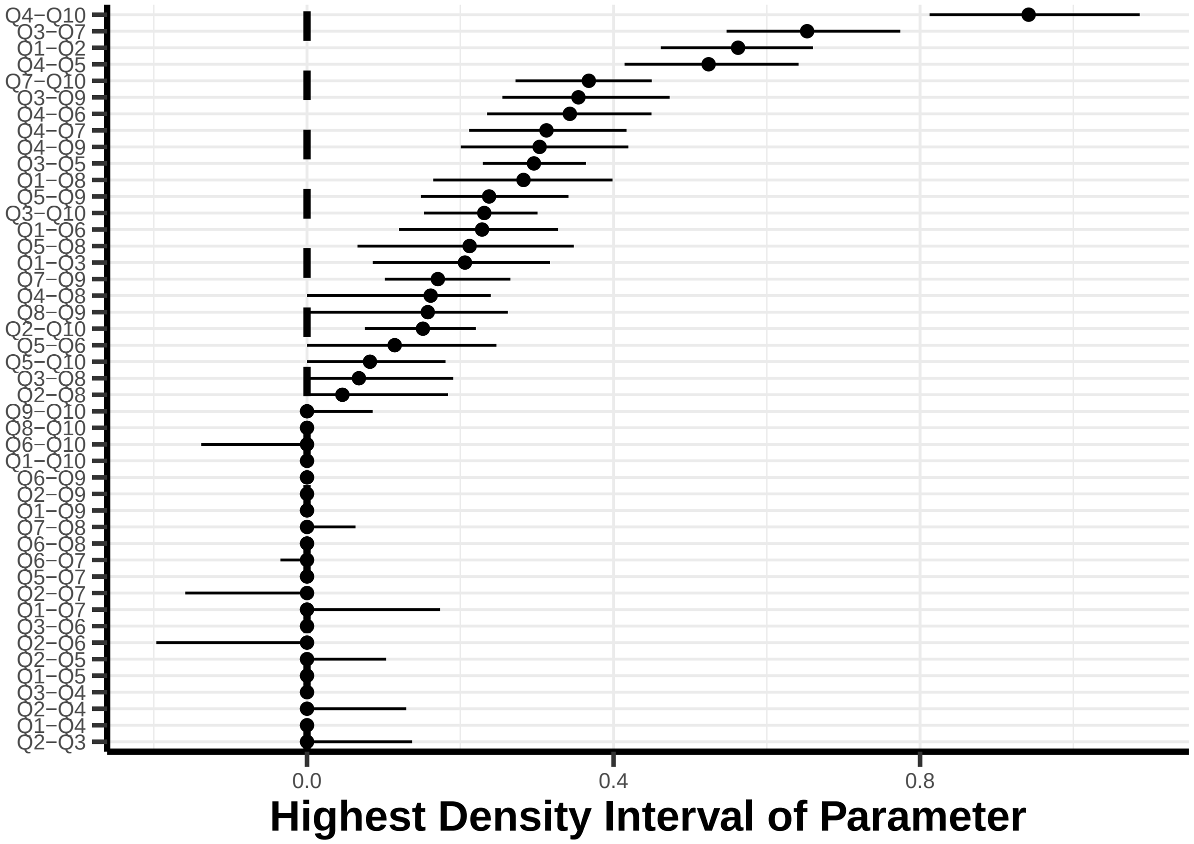
```r
fit_ne <- easybgm(data = data,
type="ordinal",
iter = 1e4,
save = TRUE,
package = "bgms",
edge_selection = FALSE # estimate parameters given a single network structure
)
summary(fit_ne)
```

```
##
##  BAYESIAN ANALYSIS OF NETWORKS
##  Model type: ordinal
##  Number of nodes: 10
##  Fitting Package: bgms
## ---
##  EDGE SPECIFIC OVERVIEW
##  Relation Parameter
##     Q1-Q2     0.557
##     Q1-Q3     0.156
##     Q2-Q3     0.567
##     Q1-Q4     0.098
##     Q2-Q4     0.221
##     Q3-Q4     0.628
##     Q1-Q5    -0.039
##     Q2-Q5    -0.104
##     Q3-Q5     0.356
##     Q4-Q5     0.178
##     Q1-Q6     0.119
##     Q2-Q6    -0.037
##     Q3-Q6    -0.033
##     Q4-Q6     0.200
##     Q5-Q6     0.355
##     Q1-Q7     0.048
##     Q2-Q7     0.349
##     Q3-Q7    -0.072
##     Q4-Q7    -0.064
##     Q5-Q7     0.362
##     Q6-Q7     0.197
##     Q1-Q8    -0.015
##     Q2-Q8     0.164
##     Q3-Q8     0.241
##     Q4-Q8     0.108
##     Q5-Q8     0.222
##     Q6-Q8    -0.012
##     Q7-Q8     0.950
##     Q1-Q9     0.043
##     Q2-Q9     0.076
##     Q3-Q9     0.110
##     Q4-Q9    -0.022
##     Q5-Q9    -0.055
##     Q6-Q9     0.117
##     Q7-Q9     0.135
##     Q8-Q9     0.326
##    Q1-Q10     0.299
##    Q2-Q10    -0.100
##    Q3-Q10     0.156
##    Q4-Q10    -0.030
##    Q5-Q10     0.184
##    Q6-Q10     0.222
##    Q7-Q10    -0.140
##    Q8-Q10     0.054
##    Q9-Q10     0.077
```

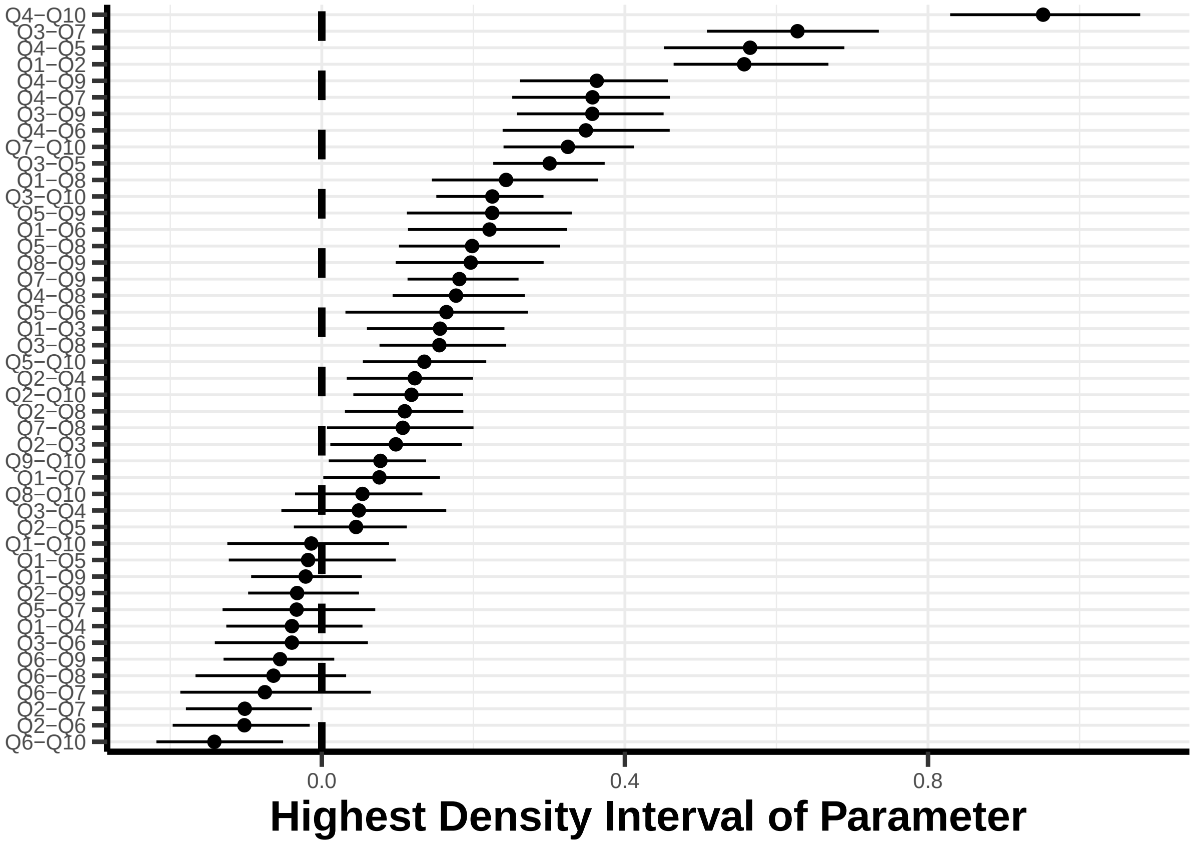
Task 2.3: Quantify the uncertainty for the interaction parameters & compare the results with those from Task 2.1 and Task 2.2.

- Visualize the uncertainty for the interaction parameters using the parameter forest plot. You can generate the plot with the `plot_parameterHDI` function. Do this for both the model with edge selection and the model without edge selection.

```
par(mfrow=c(1,2))
plot_parameterHDI(fit)
```



```r
plot_parameterHDI(fit_ne)
```



- How do you interpret credible intervals, especially the highest density interval? If you are unsure, there is more information on this in the tutorial paper or in the JASP guidelines for reporting a Bayesian analysis:

Van Doorn, J., Van Den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., . . . & Wagenmakers, E. J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychon Bull Rev 28*, 813–826

Credible intervals are used in Bayesian statistics to convey the degree of confidence we have in individual parameter values after seeing the data. And x% credible interval covers x% of the posterior distribution of the parameter. The highest density interval (HDI) is the shortest credible interval containing x% of the posterior probability mass.

Part 3: Reflection and Conceptual Understanding

Throughout the talk (and in detail in the tutorial paper), we mention the practical benefits of analyzing networks using a Bayesian approach. What are these benefits?

- *Obtaining evidence for edge inclusion and exclusion*
- *Quantifying structure uncertainty*
- *Quantifying parameter precision*

Optional Part 4: Prior robustness

Reviewer 2 criticizes you for using a Bayesian approach because they find the prior choices arbitrary. You want to be more certain that your results are not just based on your specific prior choices. You perform a prior robustness check. For more details on the priors used in the analysis of these models you can take a look at:

Sekulovski, N., Keetelaar, S., Haslbeck, J. M. B., & Marsman, M. (2024). Sensitivity Analysis of Prior Distributions in Bayesian Graphical Modeling: Guiding Informed Prior Choices for Conditional Independence Testing. *Advances .in/psychology*, e92355.

- What prior settings could you change to conduct a prior robustness check? Have a look at the help file for the `easybgm` function to see what prior choices the three underlying packages allow. You can also explore the help files of the underlying packages themselves (in this case for the `bgm` function of the `bgms` package).

In Bayesian analysis of MRF models, there are two sets of prior distributions. (1) Priors on the network structure: in bgms, users can specify independent Bernoulli distributions on each edge with a specified inclusion probability, or Beta-Bernoulli prior, which gives a beta hyperprior on the edge inclusion probability. (2) Priors on the edge weight parameters, currently users can specify a Cauchy distribution with an adjustable scale parameter.

- Imagine that you have substantive or theoretical reasons to believe that there are edges that are more likely to be present and edges that are more likely to be absent. Pick some of these edges, adjust the arguments for the prior distributions accordingly, and re-estimate the network. How do the different prior choices affect your conclusions?

One way this can be achieved is by using the inclusion_probability argument.

Optional Part 5: Test a Binary Network Model

Data background: A group of clinical researchers has contacted you to help them evaluate two recently collected data sets on symptoms of alcohol use disorder (AUD). One dataset is a large-scale survey collected from the general public. It consists of 15,000 individuals. The second dataset consists of individuals who have all been diagnosed with an alcohol use disorder in the past 12 months. Unfortunately, the clinical data consists of only 500 individuals because it is much more difficult to collect data on this subgroup.

The research group wants to assess the structure of AUD symptoms in both subgroups to determine the mechanisms that lead from recreational alcohol use to harmful alcohol dependence.

Both data sets include six symptoms of AUD, namely: (1) time spent drinking and obtaining alcohol, (2) tolerance (needing more alcohol for the same effect), (3) loss of control, (4) emotional problems related to alcohol use, (5) work problems due to alcohol use, and (6) engaging in risky activities while under the influence of alcohol.

Load the `data_clinical.csv` and `data_population.csv` data. Note that access to the real data is prohibited, so we use simulated data in this example. The datasets that you will use are simulated based on the characteristics of the original data.

```r
#data_clinical <- read_csv("data_clinical.csv")
#data_population <- read_csv("data_population.csv")
```

Task 5.1: Estimate a Binary Network Model

Estimate a binary network model with `easybgm`. You have looked at previous network research on AUD and have reason to believe that 70% of the edges are present. How can you incorporate this information into the analysis? Take a look at the additional function arguments of `bgms` to see which additional argument you need to feed `easybgm`.

Task 5.2: Evaluate the Network Structure using the Edge Evidence Plot.

- Compute the edge evidence plots using `plot_edgeevidence`.

```
plot_edgeevidence(res_clinical, edge.width = 3)
plot_edgeevidence(res_population, edge.width = 3)
```

- Interpret the network structure of both subgroups. Do the networks of the clinical and population sample differ and if so, what are the key distinctions?

Task 5.3: Network Comparison Test

Recently we introduced a network comparison test in `bgms` that allows you to estimate and test if there is a difference in the edge weight parameters in the networks for two independent samples (Marsman et al., 2024). If you want to give this a try, you can download the latest version of the `bgms` package (version 0.1.4.2). Feel free to ask one of the teachers or assistants to further explain how to use this method. A code snippet is already provided below. Soon, these functionalities will also be available in the `easybgm` package.

Marsman, M., Waldorp, L., Sekulovski, N., & Haslbeck, J. M. B. (2024). A Bayesian Independent Samples T Test for Parameter Differences in Networks of Binary and Ordinal Variables. doi.org/10.31219/osf.io/f4pk9

```
?bgmCompare
compare_fit <- bgms::compare_networks(data1 = data_clinical,
                                      data2 = data_population,
                                      difference_selection = TRUE)
```

Task 5.4: Reflection and Conceptual Understanding

- During the lecture, we specified the prior probabilities in a way that did not favor one structure over another. But we can also specify these probabilities to match certain expectations we may have about the structure of the network (e.g., the network is densely connected and has only positive relationships). What sources of information could we use to formulate our expectations about the network structure before we see or analyze a new data set?

Subject matter expertise: *If you have a good understanding of the problem at hand and the context in which the data were collected, you can use your expertise to inform the prior distribution. Or you can elicit this information from other experts.*

Previous research: *If there is prior research on the topic, you can use the results of that research to inform the prior distribution. For example, you may find that certain relationships are almost always in the network or almost never; you may find that networks have very many relationships or only a few; you may find that variables are always positively related.*

Empirical data: *If you have some data that is similar or relevant to the data you want to analyze, you can use it to inform the prior distribution.*

Objective priors: *In some cases, you may want to use an objective specification of the prior probabilities that meets certain criteria. For example, in the lecture, we specified the prior probabilities to reflect our indifference to a particular structure.*

- For which of the two datasets or analyses do you think the Bayesian approach is most useful? And why?

In the clinical sample, we have a limited number of observations compared to the population sample, and it would be particularly helpful to know which edges have insufficient evidence to conclude either inclusion or exclusion.