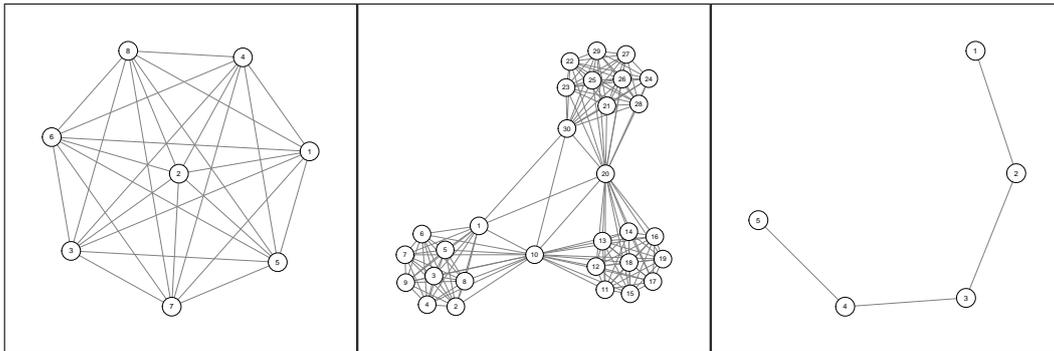


Exercises 3

Network Analysis

Exercise 1

See the following graphs:



Which of these graphs have the highest:

1. Diameter
2. Density
3. Global clustering coefficient

(do not compute these properties)

Exercise 2

In this assignment you will do some of the analyses done in the “Small world of psychopathology” paper by ? (?):

<http://www.plosone.org/article/info:doi/10.1371/journal.pone.0027407>

First load the data in your workspace:

```
load("DSMdata.RData")
```

This gives you three objects: `Adj`, `Layout` and `SymxDis`. The `Adj` object is the adjacency matrix of the graph with symptom names as column and rownames (extract them with `colnames` and `rownames` functions). The `Layout` object contains the layout for the graph, which can be passed to the `layout` argument in the `igraph` plotting function. Finally, the `SymxDis` object contains the original data: a matrix with on each row a symptom, on each column a disorder and for each element a 1 indicating a symptom is in a disorder and a 0 otherwise.

(1) Optional Explain what this network represents and how it was constructed.

(2) Without using `igraph` test if the graph is connected: can you walk from any node to any other node in the graph?

(3) Use `igraph` to plot the graph with the supplied layout. Make sure the graph looks interpretable (the vertices should not be too big). Color the nodes white with black borders and do not use labels.

(4) Use the `clusters` function in `igraph` to compute which nodes belong to the giant component. Subset `Adj`, `Layout` and `SymxDis` to new objects that only include symptoms that are in the giant component, make a new graph object for this component and plot the results again using the subsetted layout object (the layout for the giant component should be the same as in the full graph).

(5) Compute for the giant component the following measures:

- The names of the 5 most central symptoms per each of the discussed centrality measures
- The diameter of the graph
- The density of the graph
- The two discussed versions of the global clustering coefficient

Bonus point: Reproduce the graphs used to visualize centrality in the lecture. This includes varying color of the nodes corresponding to the centrality measure and plotting the symptom names in a wordcloud using the `wordcloud` package

(6) Test if this graph exhibits a small world

(7) Reproduce the two graphs of the degree distribution in figure 4 of the paper (not the network pictures).

(8) Compute the average shortest path length between "Major Depressive Disorder, Single Episode" and "Generalized Anxiety Disorder". How much symptoms do these disorders share?

(9) **Optional** Look at the `scg` function in R and play around with it. Look at the help page and especially the example section. Can you use it to produce a plot that colors at least 3 clear clusters correctly different from the rest of the graph? (tip: Look at the strongest eigenvectors of the adjacency matrix or the weakest eigenvectors of the Laplacian).

References

- Borsboom, D., Cramer, A. O., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PloS one*, *6*(11), e27407.