Assignment 1 - Network Analysis 2022

Kyuri Park, 12183881

November 16, 2022

Contents

Recap

Question 1 (3 points)

1. *FALSE*: since the network models are not equivalent to network theory, one does not necessarily need to assume network theory is true in order to apply network models. One can estimate a network model successfully when the network theory in fact does not hold.
2. *FALSE*: the links between nodes in a network model represent statistical relationships, which cannot be directly translated into causal relationships.
3. *TRUE*: the statement is true, as it follows from the PMRF definition.
4. *TRUE*: for an undirected network, the possible number of edges (m) can be computed as $\binom{n}{2} = n(n-1)/2$.
5. *TRUE*: the statement is true as PMRFs may be suggestive of causal structure, even though they cannot establish causal relations directly.
6. *FALSE*: using `minimum` argument, we can *hide* the edges under the specified value in the graph but they are not in fact *removed*.

Network Appraoches, Theory, and Models

Question 2 (2 points)

Write a short scientific essay on the network theory of mental disorders, answering the following questions. What is the external field of a mental disorder and how do you think this would impact it in the common cause framework and in the network framework? What is an implication that follows from the network perspective towards the diagnosis and treatment of mental disorders?.

When it comes to explaining mental disorder and its corresponding symptoms, the common cause framework used to be a dominant view, which states that symptom arises from a single underlying cause. As an alternative, Borsboom (2008) proposed to apply the network framework, which conceptualizes mental disorders as a network of interacting symptoms that reinforce each other.

One of many implications of adopting network approach is that in network framework, we can naturally move our focus beyond the symptom network, considering the environmental factors that are relevant to symptoms (Boer et al., 2021). The conditions that encompass these environmental factors that our outside of symptom network, yet having influences on the symptoms, form the *external field* (Borsboom, 2008). For example, adverse life events such as losing a loved one can activate a symptom, which might propagate throughout the network system. It is very plausible and often aligned with the clinical evidences (Borsboom & Cramer, 2013), but under the common cause framework, we cannot incorporate these external factors as part of explanation on developing mental disorders, while the network frame allows us to do so.

Another implication of applying network approach to psychopathology concerns the diagnosis and treatment of mental disorders. Under the traditional common cause framework, the diagnosis tends to follow from the set of symptoms that are defined by DSM-V, after which the diagnosis is used to choose a treatment protocol. In the network framework, however, the mental disorder manifests itself by a cluster of symptoms that are strongly connected to each other and correspondingly, the treatment can follow by identifying and targeting the central symptom nodes that could help breaking the strong connectivity of network system (Borsboom & Cramer, 2013). Besides, the network approach accords well with the comorbidity concept, where the relations between symptoms consist of some pathways that connect different disorders (i.e., bridge symptoms). And these so-called bridge symptoms could be promising treatment targets. Lastly, network approach also allows to study individuals' unique dynamics that may enable us to personalize interventions (Piccirillo & Rodebaugh, 2019).[1]

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

Question 3 (2 points)

Choose a paper which makes use of network analysis. Summarize the paper and explain the authors' choice for performing network analysis (if you don't agree with the choice, explain why).

- Chosen paper: *Co-morbid obsessive-compulsive disorder and depression: a Bayesian network approach (2017)* by *R.J, McNally, P.Mair, B.L. Mugno, and B.C. Rieman.*

According to Pinto et al. (2006), about 67.2% of obsessive-compulsive disorder (OCD) patients are diagnosed with major depression. Having observed that many OCD patients become depressed, McNally et al. (2017) investigated the comorbidity between OCD and depression using network analysis. The aim of their study was to apply network analysis in order to characterize the functional relationships among symptoms of OCD and depression in the patients who are diagnosed with primary OCD. They collected data on the severity of OCD as well as depression symptoms from 408 OCD patients. To accomplish the study aim, they estimated a network based on a graphical Gaussian model, whereby the edges signify the partial correlations between the pair of nodes while controlling for all the other variables. In addition, they applied a regularization method via running graphical LASSO (glasso). After they obtained their network model, they computed the strength and betweenness centrality to measure the importance of nodes (symptoms). They found that some of the nodes had the greatest strength centrality as well as highest betweenness. Accordingly, they concluded that those symptoms would have high clinical relevance and thus important when it comes to the comorbidity between OCD and depression.

Overall, I agree with the analytical methods these authors used: fitting graphical LASSO network and computing the centrality/betweenness measures. Fitting a GGM aligns well with the type of data they used (i.e., cross-sectional data with continuous variables) and I do think that using glasso regularization was appropriate given that they had quite many variables (i.e., 26 nodes), which are likely to result in some false-positive edges.

Question 4 (2 points)

Is it worth the hype? Critically reflect on the network approach to psychopathology. Think of one (conceptual or methodological) limitation of the network approach/network analysis in Psychology.

I believe that the network theory has promoted pioneering psychological research, especially in psychopathology since Borsboom first introduced it in 2008. The claim that the mental disorders emerge from causal interactions among symptoms (Borsboom & Cramer, 2013) has gained a substantial amount of attention over the last two decades as it aligns so well with the intuition on how mental disorder develops.

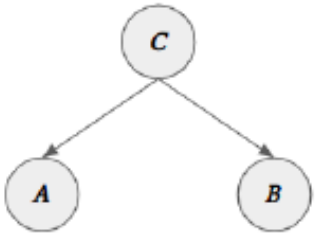
A lot of empirical researchers have tried fitting statistical network models to gain insight into these dynamics between symptoms as per the network theory posits (Robinaugh et al., 2020). Even though the network theory and network model are not equivalent, it is unfortunately often the case that researchers tend to use two terms interchangeably and draw causal inferences based on the estimated network model structure (Ryan & Dablander, 2022). One of the limitations of the statistical network model is that the edges in the network only represents the statistical relations, and they cannot be directly translated into causal relations. However, due to the confusion between two concepts, it has been observed that some unjustified causal interpretations were made based on statistical network models (Bringmann et al., 2022).

Recently though, there has been active research going on estimating causal network graphs, which actually signify causal relations (Kossakowski et al., 2021). I expect that with a set of methodologies to discovering causal relations that are currently studied, there will be soon an important extension to the network analysis tool that can validate inferring causal relations. As a conclusion, I think that the network theory is indeed worth the hype. Despite the limitations, it has created a great momentum in psychological research and I believe that there will come additional tools to greatly improve the application of network analysis.

Causality, Conditional Independece & PMRFs

Question 5 (1 point)

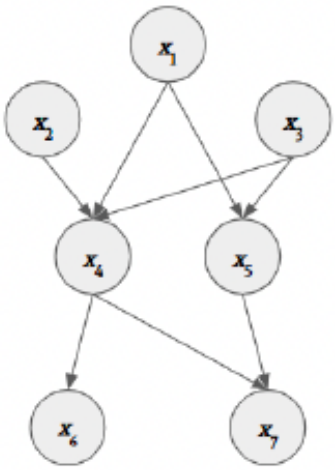
For the indicated graph, explain the dependency structure at hand. What happens if variable C is observed?



- A and B are marginally dependent due to the common cause C : $A \not\perp\!\!\!\perp B$. In other words, C is a confounder, which explains the association between A and B.
- If variable C is observed, A and B are conditionally independent given C: $A \perp\!\!\!\perp B \mid C$. That is, the association between A and B disappears if you condition on the confounder C.

Question 6 (1 point)

For the indicated graph below first indicate for each node x_k its parent nodes pa_k. Using this formula, write down the joint distribution $p(x_1, ..., x_7)$ for the graph below.



- $pa_1 = \emptyset$
- $pa_2 = \emptyset$
- $pa_3 = \emptyset$
- $pa_4 = \{x_1, x_2, x_3\}$
- $pa_5 = \{x_1, x_3\}$
- $pa_6 = \{x_4\}$
- $pa_7 = \{x_4, x_5\}$

Joint distribution for the graph is as follows:

$$p(x_1, x_2, x_3, x_4, x_5, x_6, x_7) = p(x_1) \cdot p(x_2) \cdot p(x_3) \cdot p(x_4|x_1, x_2, x_3) \cdot p(x_5|x_1, x_3) \cdot p(x_6|x_4) \cdot p(x_7|x_4, x_5)$$

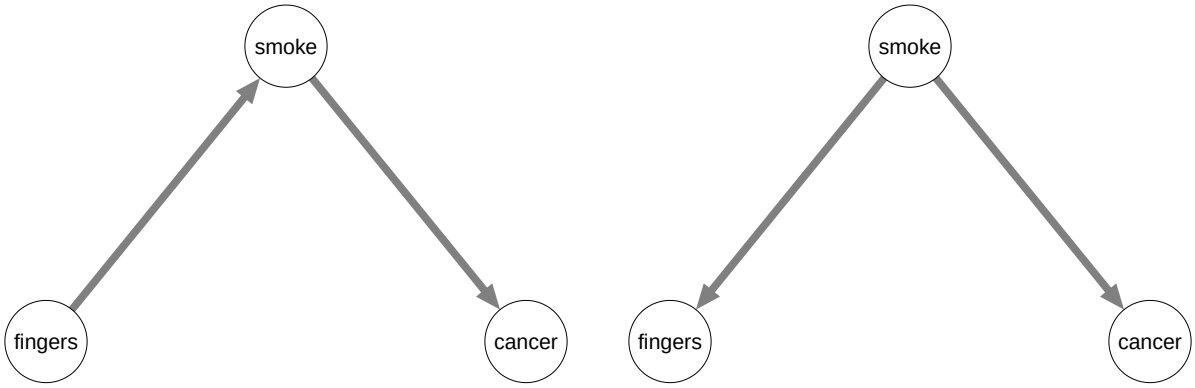
Question 7

Question 7.1 (point)

Consider the three variables `fingers`, `cancer` and `smoke`.

```
## A
ind(fingers, cancer)          # p < 0.5 they are not independent.
## B
cind(smoke, cancer, fingers) # p < 0.5: they are not conditionally independent.
## C
cind(fingers, cancer, smoke) # p > 0.5: they are conditionally independent given smoke.
## D
cind(smoke, fingers, cancer) # p < 0.5: they are not conditionally independent given cancer.
## E: (additional check)
ind(fingers, smoke)           # p < 0.5 they are not independent.
ind(cancer, smoke)            # p < 0.5 they are not independent.
```

A) No, `fingers` and `cancer` are *not* independent.
B) No, `smoke` and `cancer` are *not* conditionally independent given `fingers`.
C) Yes, `fingers` and `cancer` are conditionally independent given `smoke`.
D) No, `smoke` and `fingers` are *not* conditionally independent given `cancer`.
E) `fingers` and `smoke` are *not* independent.
`cancer` and `smoke` are *not* independent.
The causal path that is consistent with the data is:
fingers $\rightarrow$ ***smoke*** $\rightarrow$ ***cancer*** *and* ***fingers*** $\leftarrow$ ***smoke*** $\rightarrow$ ***cancer***

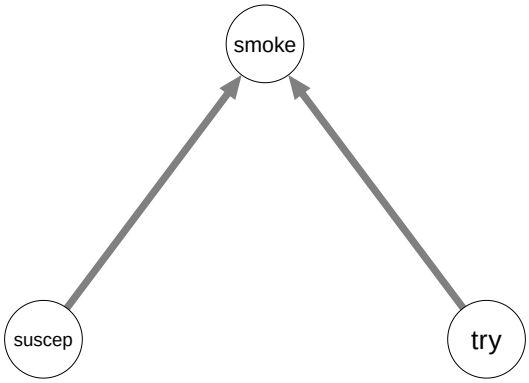


Question 7.2 (1 point)

Consider the three variables `smoke`, `try` and `susceptible`.

```
## A
ind(susceptible, try)          # p > 0.5 they are independent.
## B
cind(susceptible, try, smoke) # p < 0.5: they are not conditionally independent.
## C (additional check)
ind(susceptible, smoke)        # p < 0.5 they are not independent.
ind(try, smoke)                # p < 0.5 they are not independent.
cind(susceptible, smoke, try) # p < 0.5: they are not conditionally independent.
cind(try, smoke, susceptible) # p < 0.5: they are not conditionally independent.
```

A) Yes, `susceptible` and `try` are independent.
B) No, `susceptible` and `try` are *not* conditionally independent given `smoke`.
C) `susceptible` and `smoke` are *not* independent.
`try` and `smoke` are *not* independent.
`susceptible` and `smoke` are *not* conditionally independent given `try`.
`try` and `smoke` are *not* conditionally independent given `susceptible`.
The causal path that is consistent with the data is: ***`susceptible`*** $\rightarrow$ ***`smoke`*** $\leftarrow$ ***`try`***

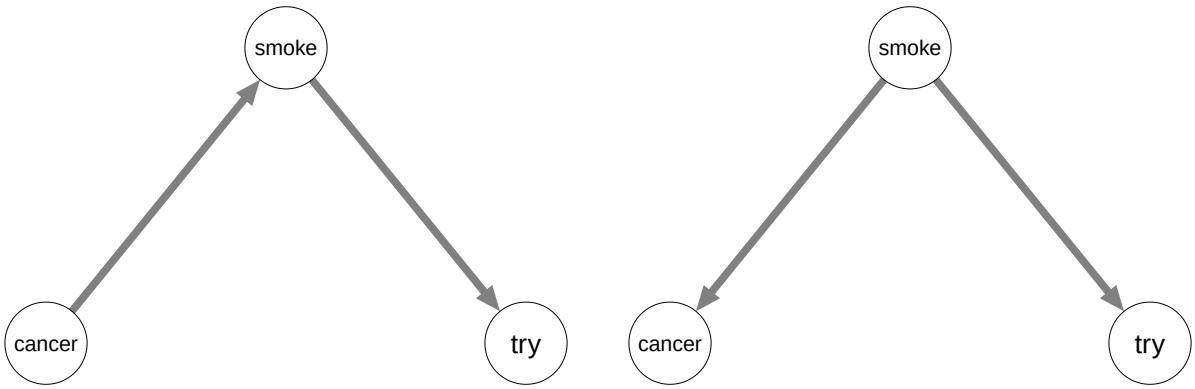


Question 7.3 (1 point)

Consider the three variables `try`, `smoke` and `cancer`.

```
## A
ind(try, cancer)                # p < 0.5 they are not independent.
## B
cind(try, cancer, smoke)        # p > 0.5: they are conditionally independent.
## C (additional check)
ind(try, smoke)                 # p < 0.5 they are not independent.
ind(smoke, cancer)              # p < 0.5 they are not independent.
cind(try, smoke, cancer)        # p < 0.5: they are not conditionally independent.
cind(cancer, smoke, try)        # p < 0.5: they are not conditionally independent.
```

A) No, `try` and `cancer` are *not* independent.
B) Yes, `try` and `cancer` are conditionally independent given `smoke`.
C) `try` and `smoke` are *not* independent.
`smoke` and `cancer` are *not* independent.
`try` and `smoke` are *not* conditionally independent given `cancer`.
`cancer` and `smoke` are *not* conditionally independent given `try`.
The causal path that is consistent with the data is:
`cancer` $\rightarrow$ ***`smoke`*** $\rightarrow$ ***`try`*** *and* ***`cancer`*** $\leftarrow$ ***`smoke`*** $\rightarrow$ ***`try`***

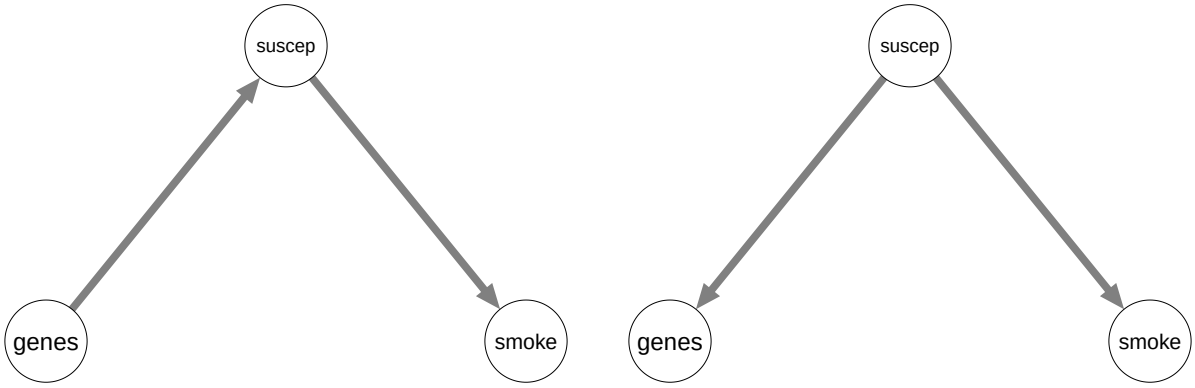


Question 7.4 (1 point)

Consider the three variables `genes`, `susceptible` and `smoke`.

```
## A
ind(smoke, genes)                    # p < 0.5 they are not independent.
## B
cind(smoke, susceptible, genes)      # p < 0.5: they are not conditionally independent.
## C
cind(genes, smoke, susceptible)      # p > 0.5: they are conditionally independent.
## D (additional check)
ind(smoke, susceptible)              # p < 0.5 they are not independent.
ind(genes, susceptible)              # p < 0.5 they are not independent.
cind(genes, susceptible, smoke)      # p < 0.5: they are not conditionally independent.
cind(susceptible, smoke, genes)      # p < 0.5: they are not conditionally independent.
```

A) No, `smoke` and `genes` are *not* independent.
B) No, `smoke` and `susceptible` are *not* conditionally independent given `genes`.
C) Yes, `genes` and `smoke` are conditionally independent given `susceptible`.
D) `smoke` and `susceptible` are *not* independent.
`genes` and `susceptible` are *not* independent.
`genes` and `susceptible` are *not* conditionally independent given `smoke`.
`susceptible` and `smoke` are *not* conditionally independent given `genes`.
The causal path that is consistent with the data is:
`genes` $\rightarrow$ ***`susceptible`*** $\rightarrow$ ***`smoke`*** *and* ***`genes`*** $\leftarrow$ ***`susceptible`*** $\rightarrow$ ***`smoke`***



Question 7.5 (1 point)

Do any additional checks you want on the full dataset.Then draw your best guess of the DAG that created the data. Indicate the evidence you have for your DAG, as well as the evidence that speaks against it (if you have any). Also indicate for which parts of your DAG you have no conclusive evidence and explain why.

I will use the (in)dependence relations I find using conditional independence tests, taking a *brute force* approach. Following two principles below, I will first find the skeleton of the DAG based on the results of (in)dependent tests and then, I will orient the edges upon identifying colliders.

- **Principle 1**: Two variables A and B are directly connected in the DAG (either A →B or B→A) if and only if they are dependent conditional on every possible subset of the other variables.
- **Principle 2**: If the skeleton contains a triplet $A - B - C$, the edge can be orientated as A→B←C if and only if A and C are dependent conditional on every set of variables containing B.

Step1: Perform marginal/conditional independence tests based on correlations using alpha level of 0.05 (assuming normal errors and linear relationships) for every possible pairs and triplets.

Below, you can find the table of marginal relations between pairs. The conditional relations between triplets can be found in the *Appendix* at the end of the document (as it is quite lengthy).

```
### Step1: perform conditional independence test
# get the variable names
variables <- colnames(datafile)
# get all possible combinations of pairs
marginal_string <- t(combn(variables,2))
colnames(marginal_string) <- c("DV1", "DV2")
condition_vars <- list()
for (i in 1:nrow(marginal_string)) {
  condition_vars[[i]] <- variables[!variables %in% marginal_string[i,]]
}
# get all possible combinations of triplets
cond1_string <- matrix("NA",60,3)
colnames(cond1_string) <- c("DV1", "DV2", "Conditional on")
for(i in 1:4){
  cond1_string[seq(i,60+(i-1), 4), c(1,2)] <- marginal_string[, c(1,2)]
}
sequence <- seq(1,60,4)
for (i in 1:nrow(marginal_string)) {
  condition_var <- variables[!variables %in% marginal_string[i,]]
    cond1_string[sequence[i],3] <- condition_var[1]
    cond1_string[sequence[i]+1,3] <- condition_var[2]
    cond1_string[sequence[i]+2,3] <- condition_var[3]
    cond1_string[sequence[i]+3,3] <- condition_var[4]
}
## test the marginal (in)dependency
marg_p <- apply(marginal_string,1,function(r){
  cor.test(datafile[,r[1]], datafile[,r[2]])$p.value
})
## test the conditional (in)dependency
```

```r
c1_p <- apply(cond1_string,1,function(r){
  pcor.test(datafile[,r[1]], datafile[,r[2]], datafile[,r[3]])$p.value
})
## results formatted in a table
alpha = 0.05 # specify alpha level
marg <- cbind(marginal_string,  Relation = ifelse(marg_p < alpha, "Dependent", "Independent"))
cond <- cbind(cond1_string, Relation = ifelse(c1_p < alpha, "Dependent", "Independent"))
pander(marg, caption = "Marginal relations between pairs")
# pander(cond) # find the conditional relations table in the Appendix
```

Table 1: Marginal relations between pairs

DV1	DV2	Relation
genes	susceptible	Dependent
genes	try	Independent
genes	smoke	Dependent
genes	fingers	Dependent
genes	cancer	Independent
susceptible	try	Independent
susceptible	smoke	Dependent
susceptible	fingers	Dependent
susceptible	cancer	Independent
try	smoke	Dependent
try	fingers	Dependent
try	cancer	Dependent
smoke	fingers	Dependent
smoke	cancer	Dependent
fingers	cancer	Dependent

Step2: Using the first principle, I will extract the skeleton of the DAG. I will start by drawing an undirected graph where every variable is connected to every other variables. Then, remove edges between variables if they are either marginally or conditionally independent in any of the tests from the previous step.

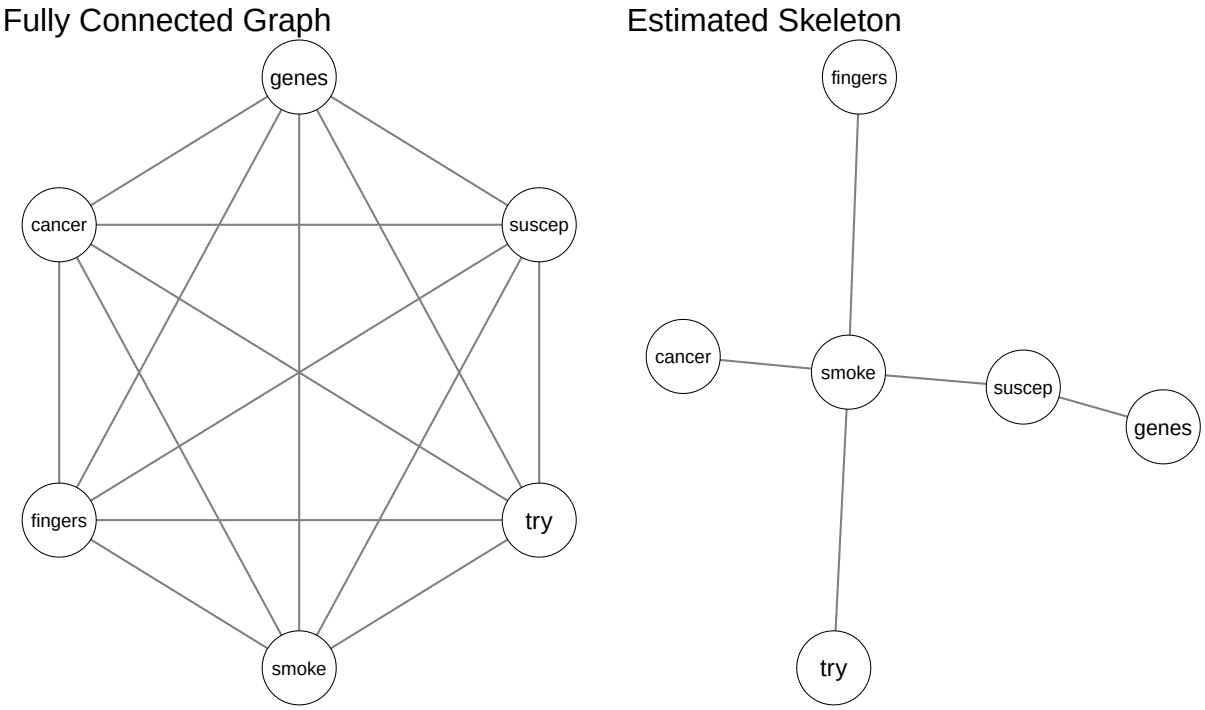
```r
### Step2: extract the skeleton
## find the marginal and conditioanl independencies
marg %>% as.data.frame() %>% filter(Relation == "Independent")
cond %>% as.data.frame() %>% filter(Relation == "Independent")

# get a fully connected matrix
adj_full <- matrix(1,6,6)
diag(adj_full) <- 0

adj <- adj_full
# Remove the edges where relation is independent
adj[1,3] <- adj[3,1] <- 0
adj[1,4] <- adj[4,1] <- 0
adj[2,3] <- adj[3,2] <- 0
adj[2,6] <- adj[6,2] <- 0
adj[1,6] <- adj[6,1] <- 0
adj[1,5] <- adj[5,1] <- 0
```

```
adj[2,5] <- adj[5,2] <- 0
adj[3,5] <- adj[5,3] <- 0
adj[3,6] <- adj[6,3] <- 0
adj[5,6] <- adj[6,5] <- 0

# plot the skeleton
variables[2] <- "suscep"
par(mfrow=c(1,2))
qgraph(adj_full, labels = variables, vsize = 10, title="Fully Connected Graph")
qgraph(adj, labels = variables, vsize = 10, title="Estimated Skeleton")
```

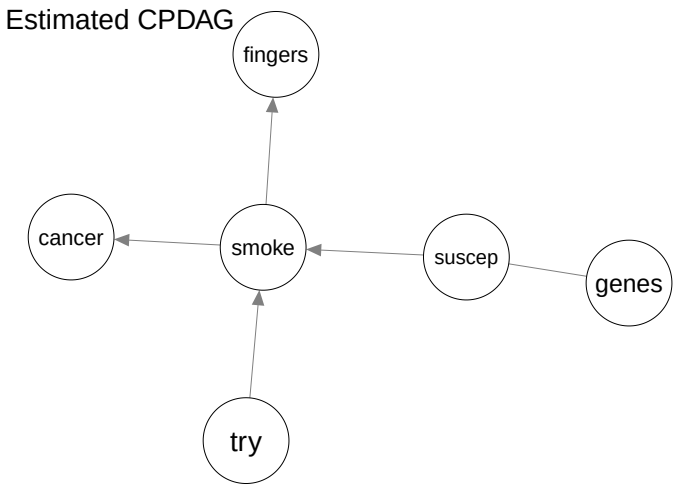


Step3: Using the second principle, I basically perform d-separation rule for colliders and orient the edges. There are six triplets (A - B - C) I need to check if $A \not\perp\!\!\!\perp B \mid C$ holds:
A) `cancer - smoke - fingers`
B) `cancer - smoke - susceptible`
C) `fingers - smoke - susceptible`
D) `smoke - susceptible - genes`
E) `cancer - smoke - try`
F) `try - smoke - susceptible`

As shown in *Table 2* (see Appendix), only `try - smoke - susceptible` has a conditional dependence (`try` $\not\perp\!\!\!\perp$ `susceptible` | `smoke`). Accordingly, we can orient the edge such that `try` $\rightarrow$ `smoke` $\leftarrow$ `susceptible`. Upon identifying only one collider in the structure, we can also orient the following edges: `susceptible` $\rightarrow$ `smoke` $\rightarrow$ `fingers` and `cancer` $\leftarrow$ `smoke` $\leftarrow$ `try` because otherwise they also have to be colliders, which cannot be the case based on the found statistical relations. See below for the resulting *complete partially directed acyclic graph* (CPDAG).

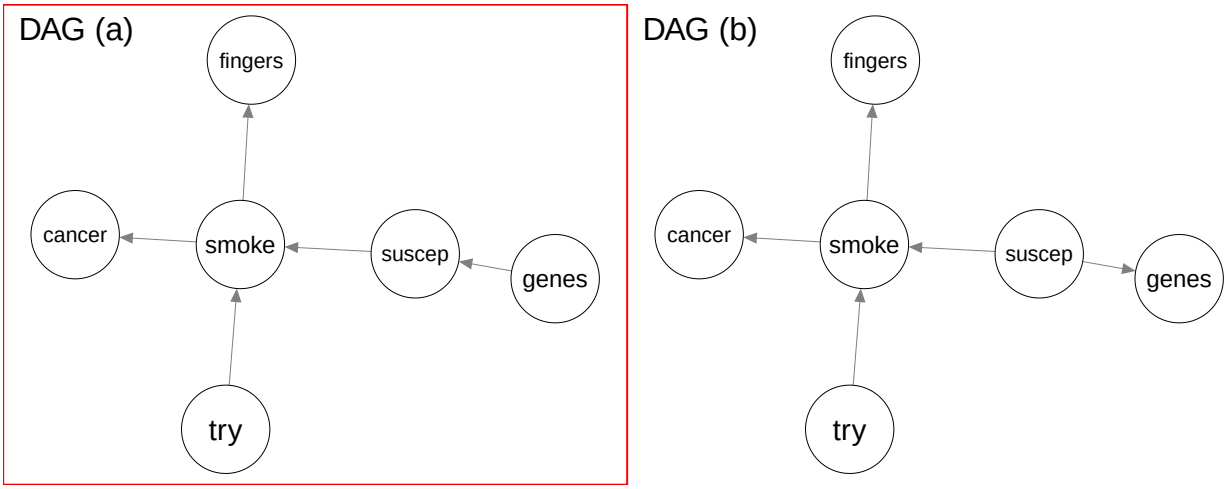
```
### Step3: identify colliders to orient edges
# specify directed (true) and undirected (false) edges
cptf <- matrix(FALSE, 6,6)
cptf[2,4] <- cptf[3,4] <- cptf[4,5] <- cptf[4,6] <- TRUE
```

```r
# plot the CPDAG
qgraph(adj, labels = variables,  directed = cptf, title = "Estimated CPDAG",
       asize = 5, vsize = 13, title.cex = 1.5)
```



Conclusion: Given the estimated CPDAG, we have two DAGs in the equivalence set (see below), as the relation between `susceptible - genes` is *inconclusive* based on the result of (in)dependence tests. I believe that the true DAG is the one on the left, DAG **(a)**, because `genes` can sensically affect `susceptibility` but not the other way around (i.e., susceptibility influences genes is not practically possible).

In addition, there has been found some inconsistencies from the statistical test results; presumed that DAG **(a)** is indeed true, the marginal independencies found between `genes - cancer` and `susceptible - cancer` no longer hold, as they are connected with *chain* structure. However, given that the implied relationships represented by DAG **(a)** make a lot of sense in substantive term (e.g., *being a smoker – causes → yellow-stained finger; being a smoker – causes → cancer; how often a person has tried cigarettes during adolescence – causes → being a smoker; genes – causes → susceptibility; susceptibility – causes → being a smoker*), I would think that the inconsistent marginal relationships are likely to be due to sampling errors. Therefore, my best guess for the true DAG stays the same as DAG **(a)**.



Visualizing Networks

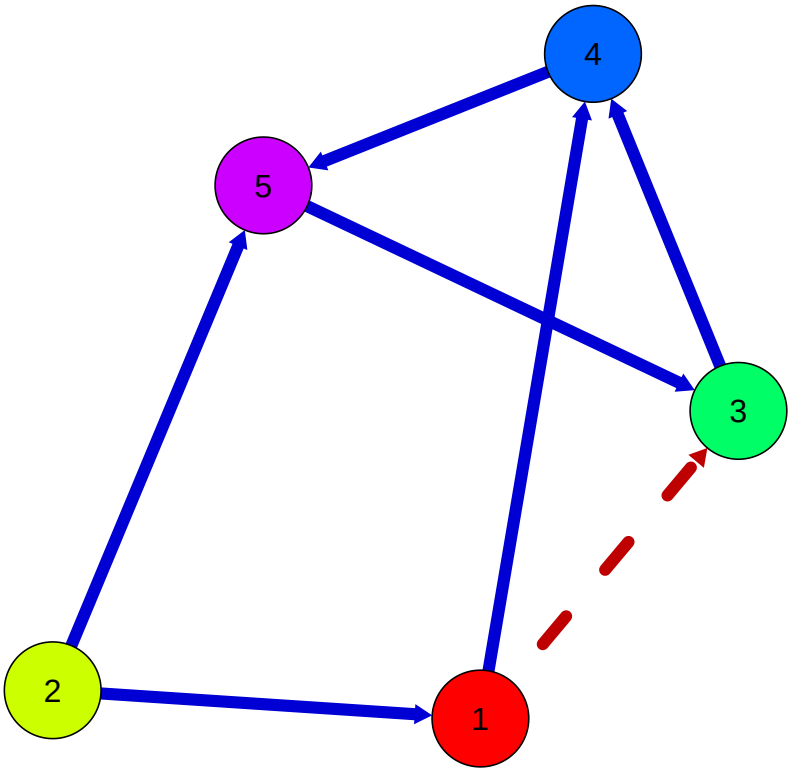
Question 8 (2 points)

Recreate the following networks as close as possible.

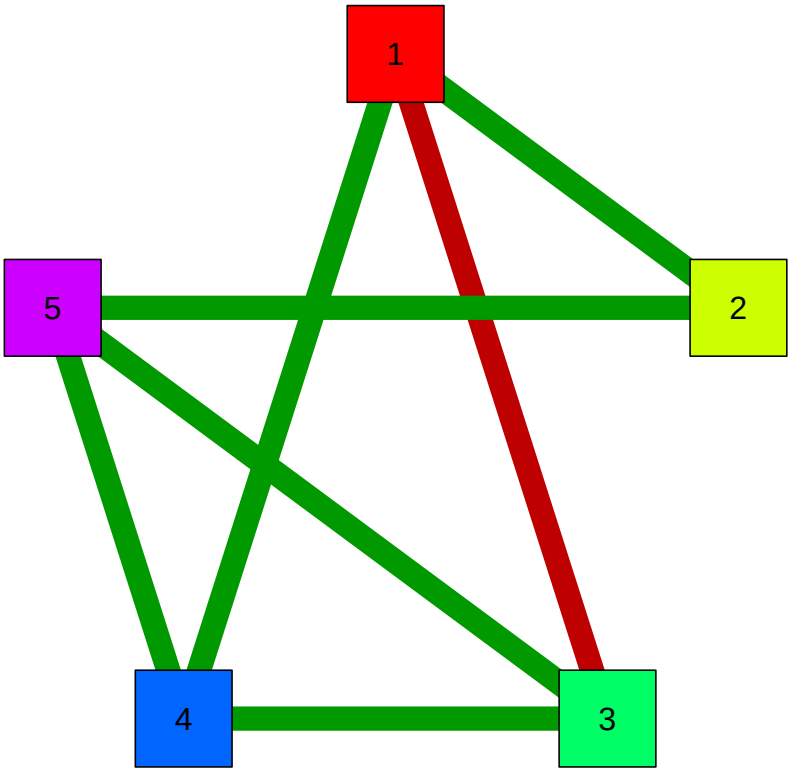
```
par(mfrow=c(2,1))
## first network
# specify the matrix for the first network
mat1 <- matrix(c(0,0,-1,1,0,
                 1,0,0,0,1,
                 0,0,0,1,0,
                 0,0,0,0,1,
                 0,0,1,0,0), 5, 5, byrow=T)
# plot the first network
qgraph(mat1, color = rainbow(5), nodeNames = paste("symptom", 1:5),
       legend.cex = 0.4, theme = "colorblind", negDashed =T)

## second network
# specify the second matrix
mat2 <- matrix(c(0,1,-1,1,0,
                 0,0,0,0,1,
                 0,0,0,1,1,
                 0,0,0,0,1,
                 0,0,0,0,0), 5, 5, byrow=T)

# plot the second network
qgraph(mat2, directed=FALSE, shape = "square", color = rainbow(5),
       nodeNames = paste("symptom", 1:5), legend.cex = 0.4)
```



1: symptom 1
2: symptom 2
3: symptom 3
4: symptom 4
5: symptom 5



1: symptom 1
2: symptom 2
3: symptom 3
4: symptom 4
5: symptom 5

Question 9 (2 points)

Visualize the network in R and write a short report (maximum 300 words) describing what the network represents and how it can be interpreted. Most importantly, explain why you visualized it the way you did – which is the focus of this assignment, so really think about how important information can best be visualized within a network.

- **Data**: I found the data online (link here), which is used in McNally et al. (2017) paper. The data contains information on OCD and depression symptoms that are rated on 5-point Likert scale from 0 to 4. There are in total 408 observations and 26 symptom variables (i.e., dimension: 408 by 26).
- **Analytical Strategy**:

Step1. Presumed that 5-point Likert sacle is sufficient to be treated as interval scales, I decide to use GGM (gaussian graphical model) to estimate the network model of these symptoms.
Step2. Before going about fitting the network, I check the distribution of variables to see if they approximate the normal distribution to some extent. I will transform the data if they deviate much from the normal distribution.
Step3. In order to get a sparse network, I use *glasso* regularization based on EBIC (Extended Bayesian Information Criteria). I tweak the hyperparameter (γ) to find the optimal density.
Step4. Given the final network model I choose, I explain how I plot the network graph and give some subtantive interpretation on what I can infer from the resulting model.

```
## prepartion
# import data
dat <- read.csv("ocd_dep.csv")
# check data
glimpse(dat)
```

```
## check the variable distribution by plotting histograms
dat[,-2] %>%
  pivot_longer(where(is.numeric)) %>%
  ggplot(aes(x = value)) +
  geom_histogram(bins = 10) +
  facet_wrap(~name) +
  theme_minimal()
```

As shown in Figure 1, some variables deviate from the normal distribution quite a bit. Following Sascha's advice on non-normal variables, I will transform the data using `huge.npn` function from `huge` package.

```
## transform data
# some are not normal --> makes the data semiparametric Gaussian using huge package
# following Sacha's advie on non-normal variable
transformed_dat <- huge::huge.npn(dat) %>% as.data.frame()

## check the distributions of transformed data
transformed_dat[,-2] %>%
  pivot_longer(where(is.numeric)) %>%
  ggplot(aes(x = value)) +
  geom_histogram(bins = 10) +
  facet_wrap(~name) +
  theme_minimal()
```

They approximate the normal distribution better after transformation, as shown in Figure 2.

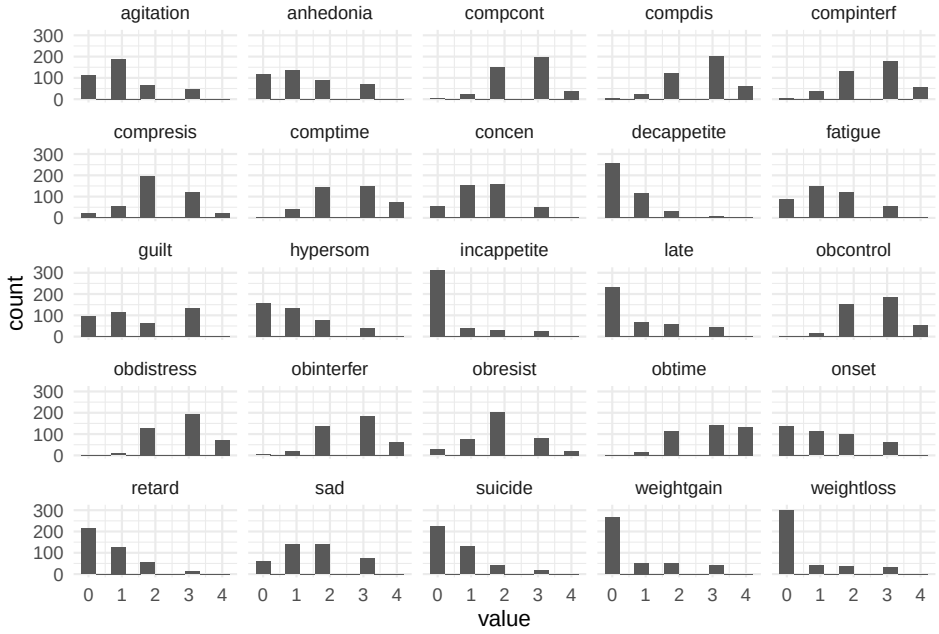


Figure 1: Distribution of original variables

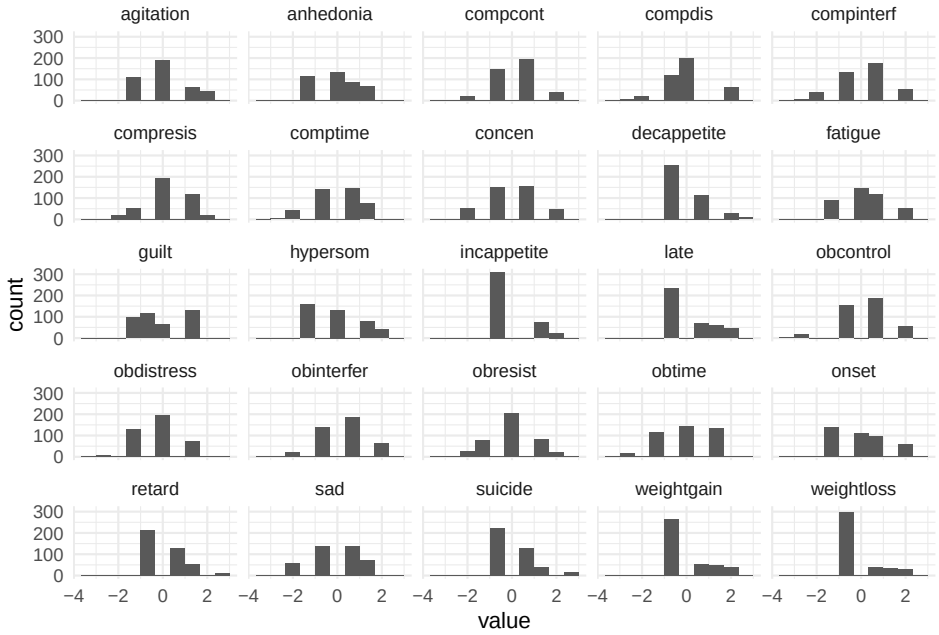


Figure 2: Distribution of transformed variables

```
## compute the correlations
cormat <- transformed_dat %>% cor()

## regularization: find the optimal tuning parameter (gamma)
layout(matrix(c(1,3,2,4),2,2))
list(0.4, 0.5, 0.6, 0.7) %>%
  purrr::map(~ EBICglasso(cormat, n = nrow(transformed_dat), gamma = .x) %>%
               qgraph(layout = "spring", theme= "colorblind", title=paste("gamma", .x))
             )
```

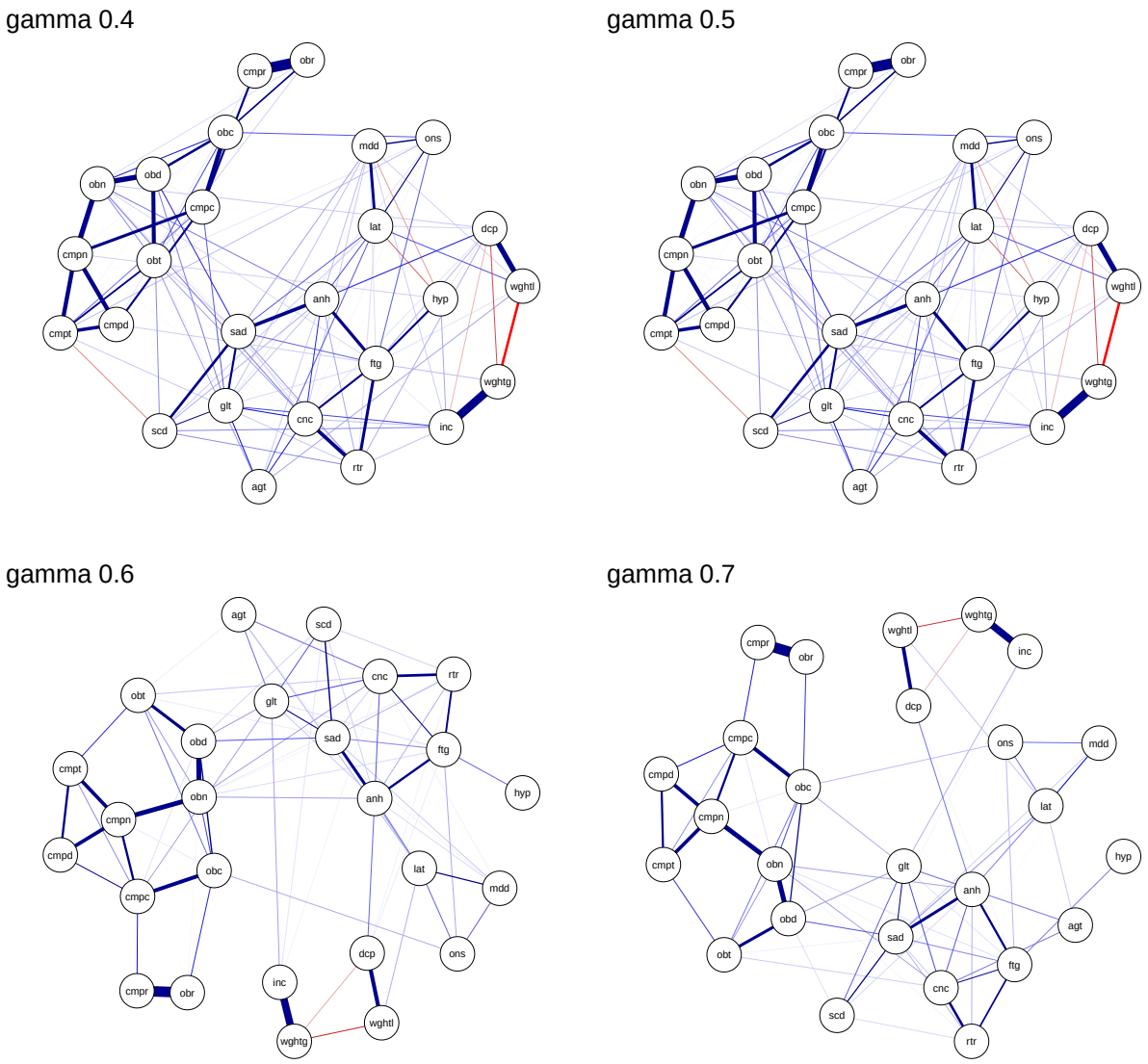


Figure 3: Networks with different density

As there are quite many variables in the dataset, I decide to apply regularization using *glasso* to get a more sparse network. Given that the general suggestion of gamma (γ) value is 0.5 (Foygel & Drton, 2010), I vary the gamma value from 0.4 to 0.7 and check the different resulting networks in terms of density.

Figure 3 shows networks with varying gamma values. I find the optimal density when $gamma = 0.6$. Even though the chosen model might have omitted some true edges (since I apply quite a conservative gamma value), I want to prioritize having less false positives (i.e., less spurious edges) in this case. Hence, my final chosen model is the one with $\gamma = 0.6$ (bottom left in Figure 3).

```r
## final model: pick gamma = 0.6 for the best sparse network
EBICglasso(cormat, n = nrow(transformed_dat), gamma = 0.6) %>%
               qgraph(layout = "spring",
                      theme= "colorblind",
                      groups = list(Depression = 1:16, OCD = 17:26),
                      color = brewer.pal(8, "Pastel2")[c(1,2)],
                      vsize = 5,
                      nodeNames = colnames(transformed_dat),
                      legend.cex=0.35,
                      cut = 0,
                      legend.mode="style1",
                      details=TRUE, label.scale=TRUE)
```

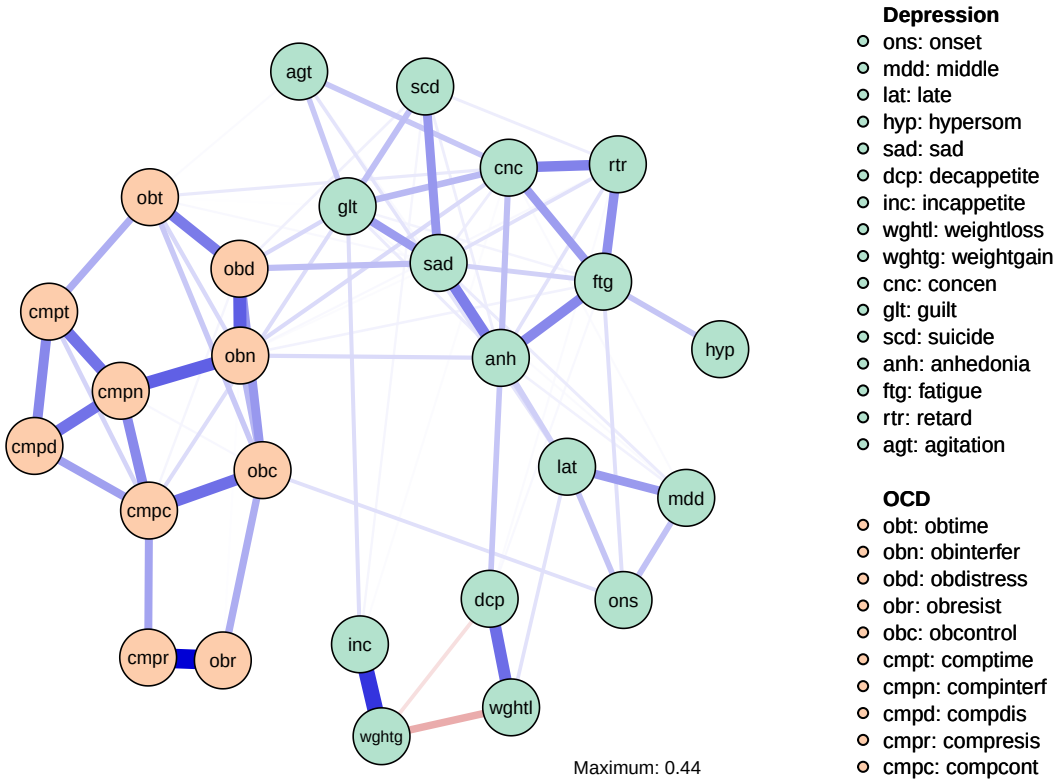


Figure 4: Final model-OCD and depression symptoms

- **Final network model explanation & interpretation**

As explained above, regularization is applied to this final model using `EBICglasso` function, as I wanted to suppress spurious relations. I used gamma value of 0.6, as it gives the optimal density in my perspective (see Figure 3). As my input is a symmetrical correlation matrix (weight matrix), `qgraph` automatically spits out an undirected network, which is appropriate in this case as I don't know the directionality in relations. I used `layout="spring"` to use the force-embedded layout, which improves the interpretability of the network compared to using the default `circular` layout setting. Then, I used `colorblind` theme for the edges to represent the positive/negative edges in blue and red respectively such that it is colorblind-friendly. I specify the `groups` for the symptoms as per its belonging disorder so that the nodes can be colored correspondingly, which thus helps identifying the possible clusters. As the original symptom names are quite long, I let `qgraph` use the abbreviated names and specify the `nodeNames` so that the full names can be shown in the legend. In addition, I used `cut = 0` to disable the cutoff

such that all edges can vary in width and color (saturation) depending on their weights, because this enhances interpretation of the edges. Lastly, I used `maximum = 0.44`, which is the default setting, that is the strongest edge weight in the network. I did not change to 1 (the possibly maximum correlation), as I am mostly interested in the relative strength of relations in this network, not necessarily comparing it to the absolute maximum strength. There are some other minor arguments that I used for aesthetic reasons (e.g., node size, color choice, legend size).

Figure 4 shows the final network model of OCD and depression symptoms. The network represents the regularized partial correlations (edges) between the symptoms of OCD and depression (nodes). What stands out at a first peek of this network is the clustering of symptoms. The OCD symptoms seem to cluster together so do the depression symptoms. And there is seemingly a sub-cluster in the depression symptoms including `inc`, `wghtg`, `wghtl` and `dcp`[2]. Some of the strong edges are observed between `cmpr` & `obr`, `inc` & `wghtg`, and `obd` & `obn`[3] , indicating that they are likely to occur together. Note that given this undirected partial correlation network, the directionality between the symptoms cannot be inferred. In addition, another interesting thing that can be found from this network is the connections between the OCD cluster and depression cluster. There are quite a lot of edges between them and especially the node `obn` (interference due to obsessions) has a lot of connections with the depression symptoms. This may indicate that `obn` is a critical symptom when it comes to the comorbidity between OCD and depression.

[2] `inc` = increased appetite; `wghtg` = weight gain; `wghtl` = weight loss; `dcp` = decreased appetite

[3] `cmpr` = difficulty resisting compulsion; `obr` = difficulty resisting obsessions; `inc` = increased appetite; `wghtg` = weight gain; `obd` = distress caused by obsession; `obn` = interference due to obsessions

References

Boer, N. S. de, Bruin, L. C. de, Geurts, J. J. G., & Glas, G. (2021). The Network Theory of Psychiatric Disorders: A Critical Assessment of the Inclusion of Environmental Factors. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.623970

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, *64*(9), 1089–1108. https://doi.org/10.1002/jclp.20503

Borsboom, D., & Cramer, A. (2013). Network Analysis: An Integrative Approach to the Structure of Psychopathology. *Annual Review of Clinical Psychology*, *9*, 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Bringmann, L. F., Albers, C., Bockting, C., Borsboom, D., Ceulemans, E., Cramer, A., Epskamp, S., Eronen, M. I., Hamaker, E., Kuppens, P., Lutz, W., McNally, R. J., Molenaar, P., Tio, P., Voelkle, M. C., & Wichers, M. (2022). Psychopathological networks: Theory, methods and practice. *Behaviour Research and Therapy*, *149*, 104011. https://doi.org/10.1016/j.brat.2021.104011

Foygel, R., & Drton, M. (2010). Extended bayesian information criteria for gaussian graphical models. *Advances in Neural Information Processing Systems*, *23*.

Kossakowski, J. J., Waldorp, L. J., & Maas, H. L. J. van der. (2021). The search for causality: A comparison of different techniques for causal inference graphs. *Psychological Methods*, *26*, 719–742. https://doi.org/10.1037/met0000390

McNally, R. J., Mair, P., Mugno, B. L., & Riemann, B. C. (2017). Co-morbid obsessive-compulsive disorder and depression: A Bayesian network approach. *Psychological Medicine*, *47*(7), 1204–1214. https://doi.org/10.1017/S0033291716003287

Piccirillo, M. L., & Rodebaugh, T. L. (2019). Foundations of idiographic methods in psychology and applications for psychotherapy. *Clinical Psychology Review*, *71*, 90–100. https://doi.org/10.1016/j.cpr.2019.01.002

Robinaugh, D. J., Hoekstra, R. H. A., Toner, E. R., & Borsboom, D. (2020). The Network Approach to Psychopathology: A Review of the Literature 2008–2018 and an Agenda for Future Research. *Psychological Medicine*, *50*(3), 353–366. https://doi.org/10.1017/S0033291719003404

Ryan, O., & Dablander, F. (2022). *Equilibrium causal models: Connecting dynamical systems modeling and cross-sectional data analysis*.

Appendix

Table 2: Condtional relations of all possible triplets

DV1	DV2	Conditional on	Relation
genes	susceptible	try	Dependent
genes	susceptible	smoke	Dependent
genes	susceptible	fingers	Dependent
genes	susceptible	cancer	Dependent
genes	try	susceptible	Independent
genes	try	smoke	Dependent
genes	try	fingers	Independent
genes	try	cancer	Independent
genes	smoke	susceptible	Independent
genes	smoke	try	Dependent
genes	smoke	fingers	Dependent
genes	smoke	cancer	Dependent
genes	fingers	susceptible	Independent
genes	fingers	try	Dependent
genes	fingers	smoke	Independent
genes	fingers	cancer	Dependent

DV1	DV2	Conditional on	Relation
genes	cancer	susceptible	Independent
genes	cancer	try	Independent
genes	cancer	smoke	Independent
genes	cancer	fingers	Independent
susceptible	try	genes	Independent
susceptible	try	smoke	Dependent
susceptible	try	fingers	Independent
susceptible	try	cancer	Independent
susceptible	smoke	genes	Dependent
susceptible	smoke	try	Dependent
susceptible	smoke	fingers	Dependent
susceptible	smoke	cancer	Dependent
susceptible	fingers	genes	Dependent
susceptible	fingers	try	Dependent
susceptible	fingers	smoke	Independent
susceptible	fingers	cancer	Dependent
susceptible	cancer	genes	Independent
susceptible	cancer	try	Independent
susceptible	cancer	smoke	Independent
susceptible	cancer	fingers	Independent
try	smoke	genes	Dependent
try	smoke	susceptible	Dependent
try	smoke	fingers	Dependent
try	smoke	cancer	Dependent
try	fingers	genes	Dependent
try	fingers	susceptible	Dependent
try	fingers	smoke	Independent
try	fingers	cancer	Dependent
try	cancer	genes	Dependent
try	cancer	susceptible	Dependent
try	cancer	smoke	Independent
try	cancer	fingers	Dependent
smoke	fingers	genes	Dependent
smoke	fingers	susceptible	Dependent
smoke	fingers	try	Dependent
smoke	fingers	cancer	Dependent
smoke	cancer	genes	Dependent
smoke	cancer	susceptible	Dependent
smoke	cancer	try	Dependent
smoke	cancer	fingers	Dependent
fingers	cancer	genes	Dependent
fingers	cancer	susceptible	Dependent
fingers	cancer	try	Dependent
fingers	cancer	smoke	Independent