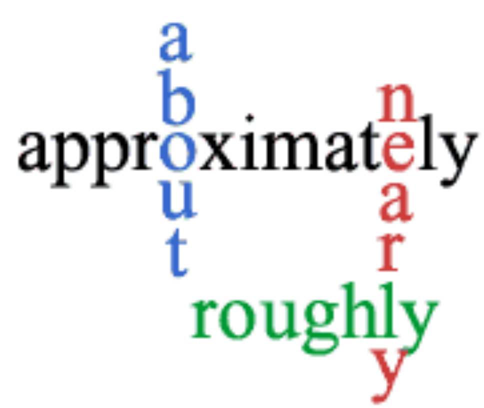
Psych Systems

Psychological Networks
Amsterdam Winter School

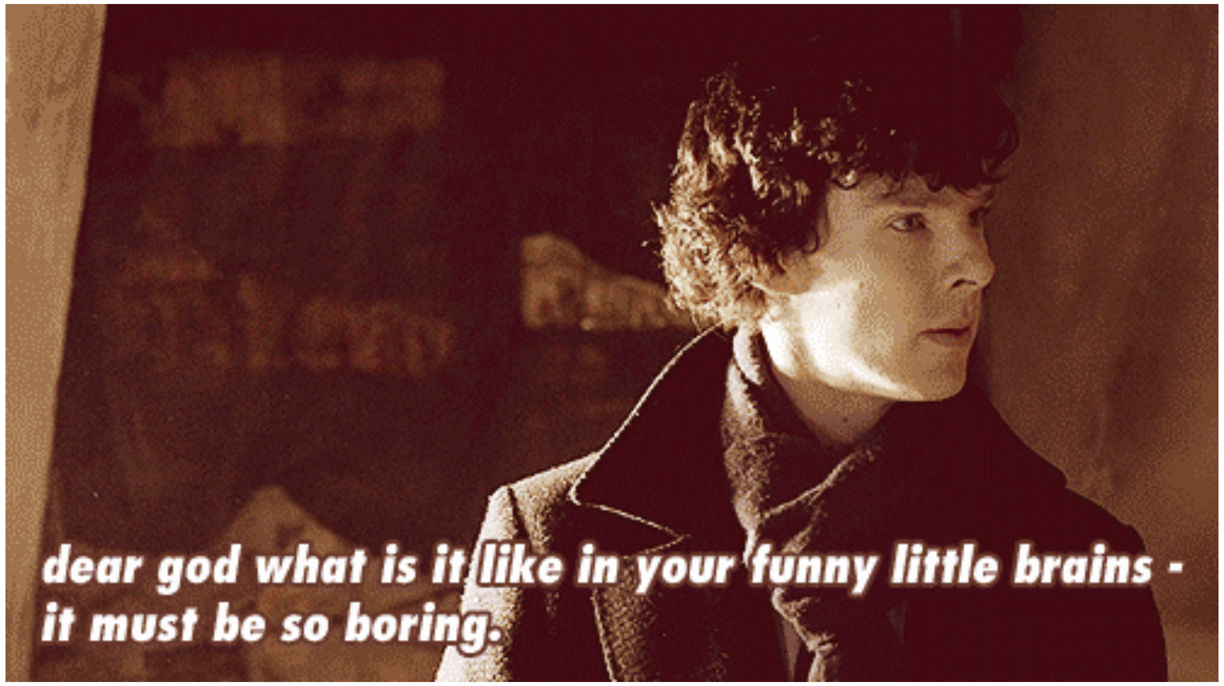
Day 3
From empirical data to a graphical model



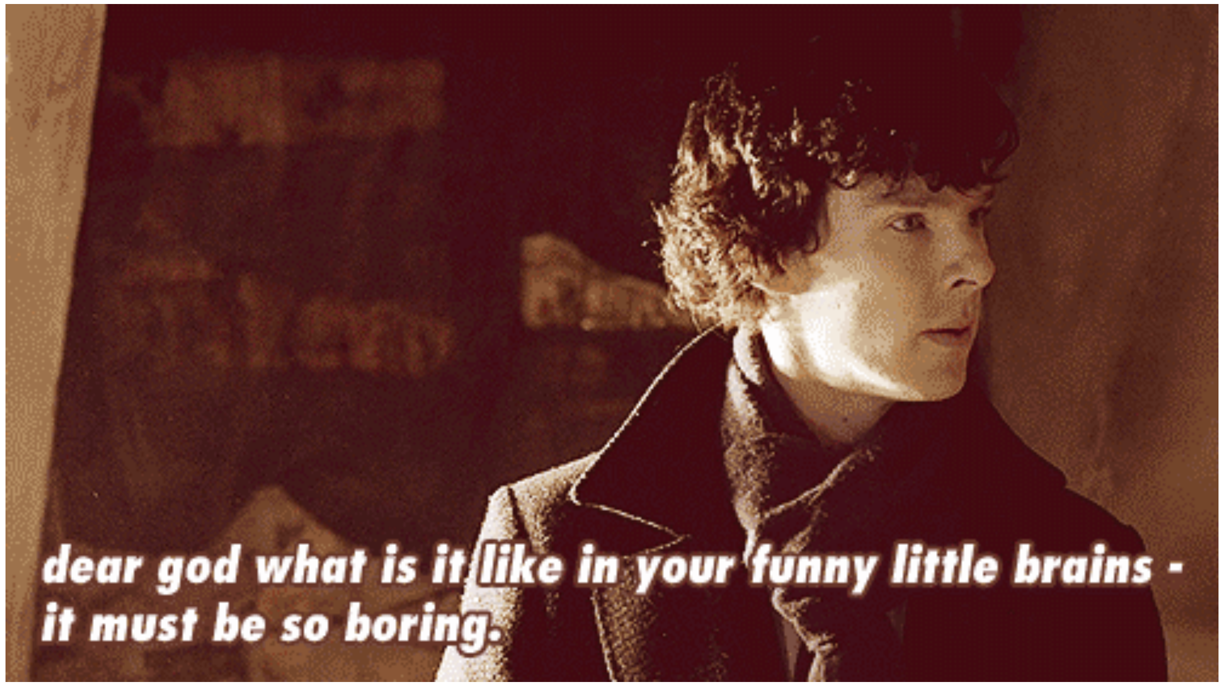
Claudia van Borkulo

Februari 15th, 2017

Sherlock - the pill game



Sherlock - the pill game



What is the network structure of psychological phenomena?

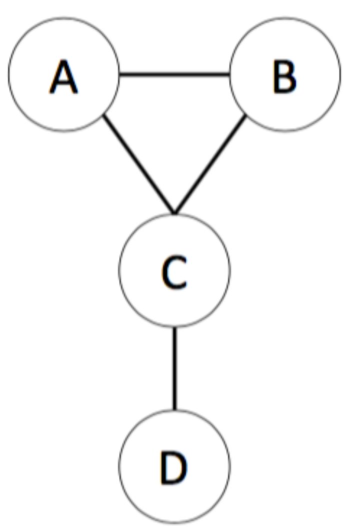
…a big challenge!

The bigger picture

SNEAK PREVIEW

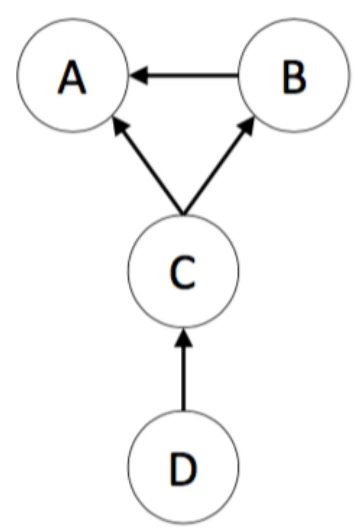
Graphical models

Markov Random Field (MRF)



Undirected graph

Bayesian Network (BN)



Directed acyclic graph

The bigger picture

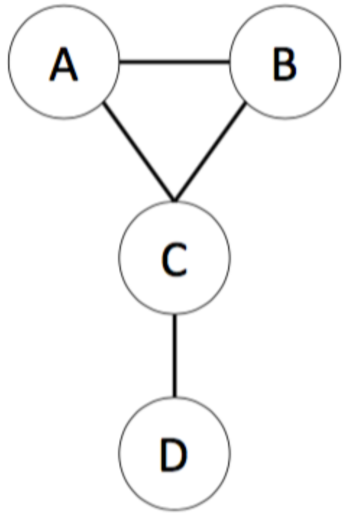
SNEAK PREVIEW

Graphical models

Wikipedia:

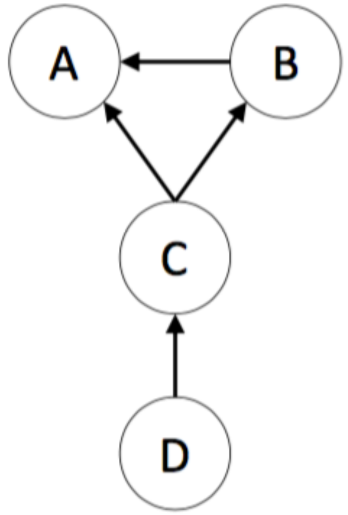
A **graphical model** or **probabilistic graphical model** (**PGM**) is a probabilistic model for which a graph expresses the conditional dependence structure between random variables.

Markov Random Field (MRF)



Undirected graph

Bayesian Network (BN)



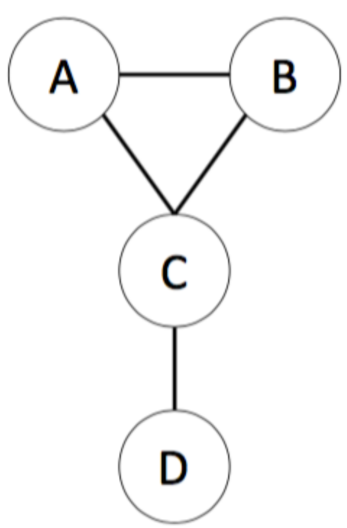
Directed acyclic graph

The bigger picture

SNEAK PREVIEW

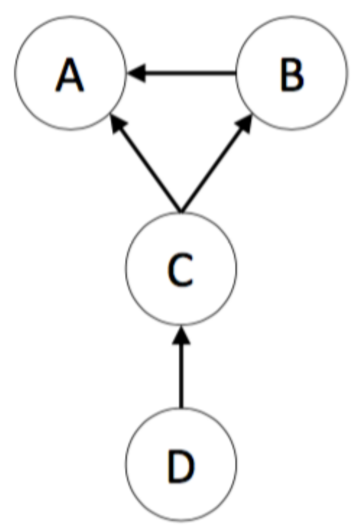
Graphical models

Markov Random Field (MRF)



Undirected graph

Bayesian Network (BN)



Directed acyclic graph

Wikipedia:

A **graphical model** or **probabilistic graphical model** (**PGM**) is a probabilistic model for which a graph expresses the conditional dependence structure between random variables.

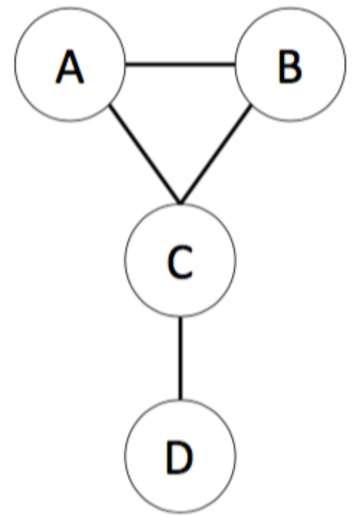
- The variables (e.g., symptoms) have pairwise/direct (causal) relationships
- An undirected edge in a MRF can be viewed as a potential causal pathway
- A missing edge means that variables are conditionally independent, i.e., independent given all other variables:

$$X_i \perp\!\!\!\perp X_j | X_{V\setminus\{i,j\}}$$

Graphical models

SNEAK PREVIEW

Markov Random Field (MRF)



Undirected graph

Ising Model

- for binary data
- cross-sectional
- estimation based on multiple logistic regression models

Gaussian Graphical Model (GGM)

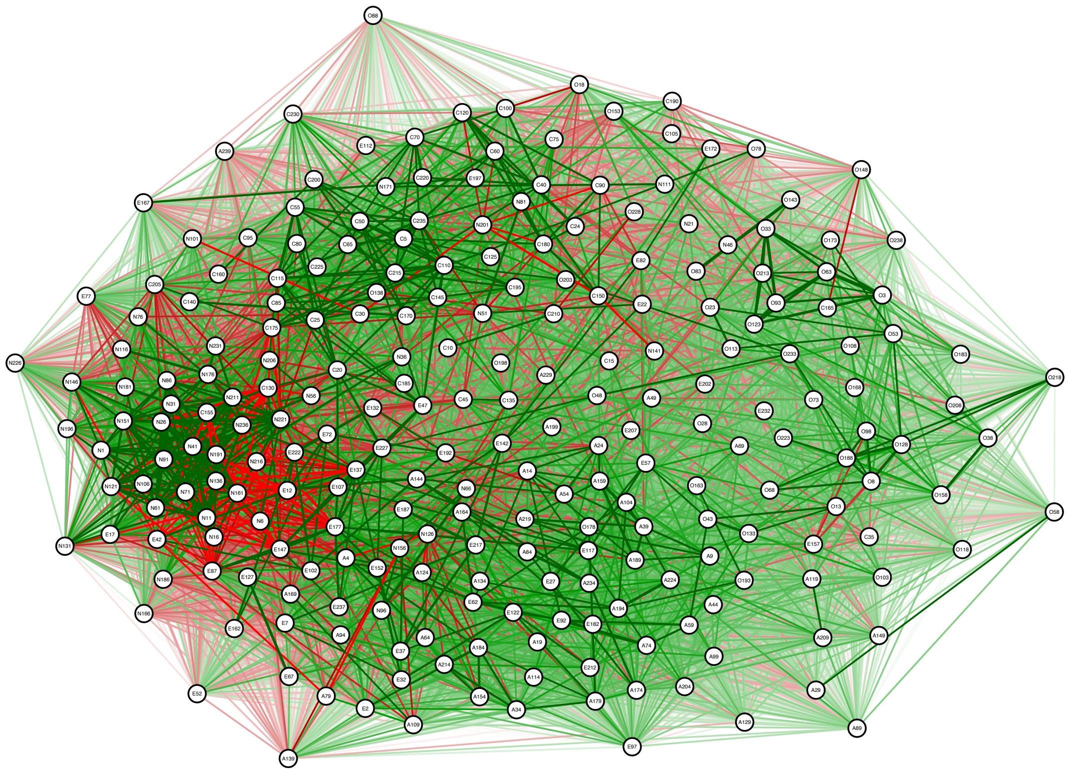
- for normally distributed data
- cross-sectional
- estimation based on multiple linear regression models (or inverse covariance matrix)

Today's program

1. **The basics: Conditional independence**
    - What does it mean to say that two symptoms ar (not) connected?
2. Estimating graphical models with L_1 regularization
3. Recent advancements
    - Network stability
    - Network comparison

What is a network?

A set of nodes and edges

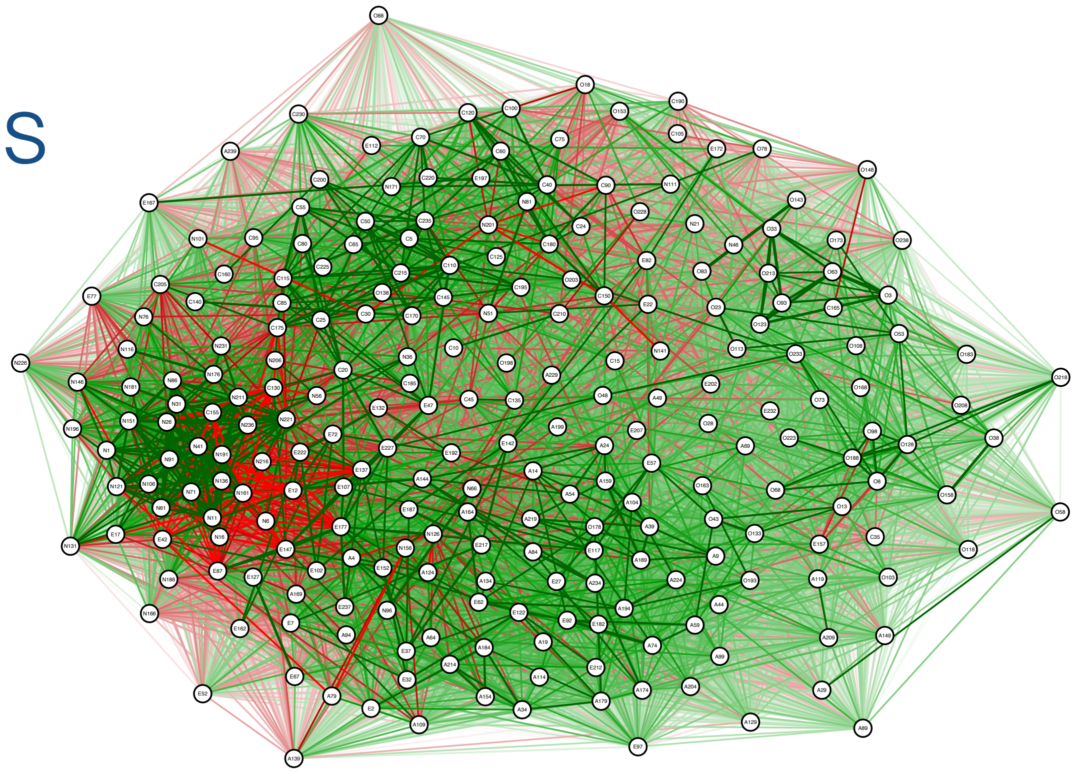


- A node is an *entity*
  - train stations
  - people
  - symptoms

- An edge is a connection between nodes
  - railways
  - friendships
  - social interaction

How to estimate a network?

A connection between symptoms can be based on:



- correlation (direct or indirect relationship)
- partial correlation (direct relationship)
- regularized relationship (afternoon)
- causal relationship (tomorrow)

How to estimate a network?

How to estimate a network?

- Another familiar analysis to establish relationships among variables

How to estimate a network?

- Another familiar analysis to establish relationships among variables
  - regression

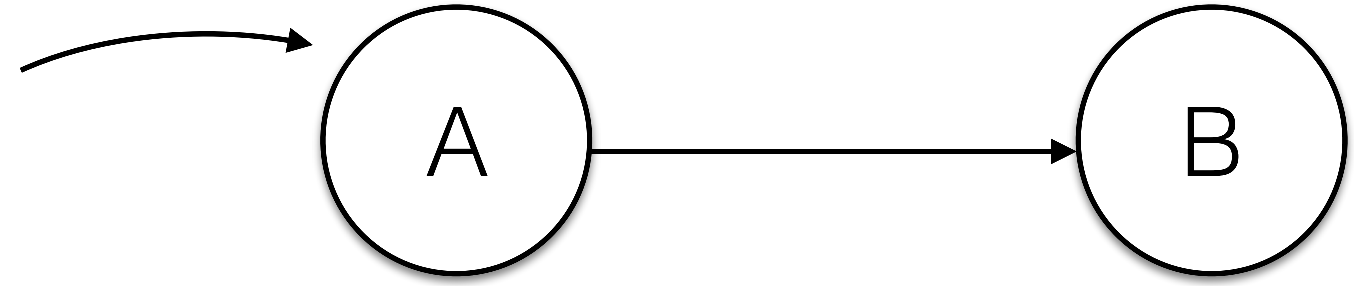
How to estimate a network?

- Another familiar analysis to establish relationships among variables
  - regression
  - $y = \tau_1 + \beta_{12}X_2 + \beta_{13}X_3 + \ldots + \varepsilon$

How to estimate a network?

- Another familiar analysis to establish relationships among variables
  - regression
  - $y = \tau_1 + \beta_{12}X_2 + \beta_{13}X_3 + \ldots + \varepsilon$

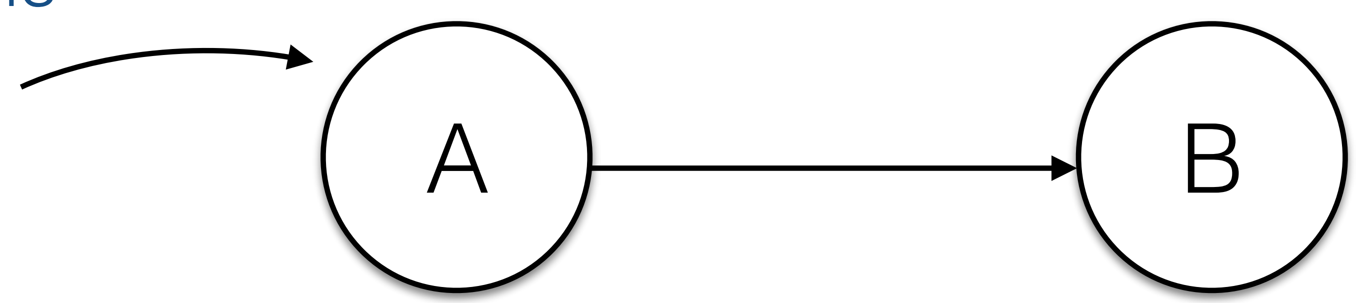
- If true generating mechanism is



How to estimate a network?

- Another familiar analysis to establish relationships among variables
  - regression
  - $y = \tau_1 + \beta_{12}X_2 + \beta_{13}X_3 + \ldots + \varepsilon$

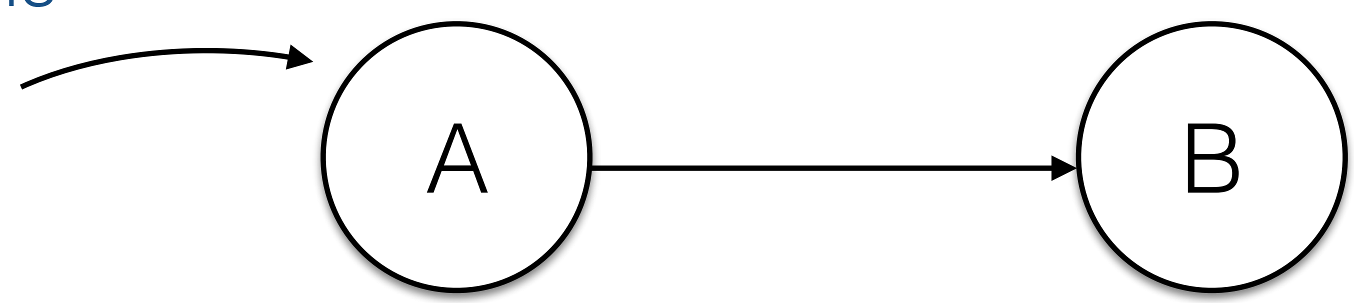
- If true generating mechanism is
- Will A predict B? $B = \tau + \beta A + \varepsilon$



How to estimate a network?

- Another familiar analysis to establish relationships among variables
  - regression
  - $y = \tau_1 + \beta_{12}X_2 + \beta_{13}X_3 + \ldots + \varepsilon$

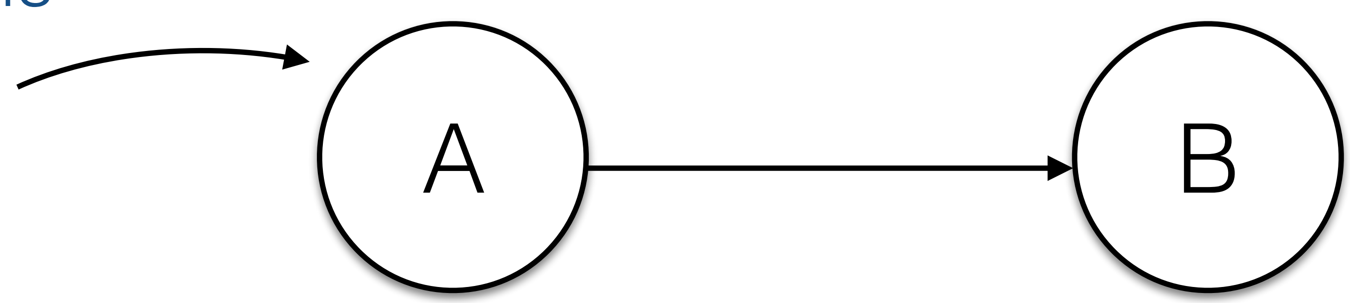
- If true generating mechanism is
- Will A predict B? $B = \tau + \beta A + \varepsilon$
  - Yes



How to estimate a network?

- Another familiar analysis to establish relationships among variables
  - regression
  - $y = \tau_1 + \beta_{12}X_2 + \beta_{13}X_3 + \ldots + \varepsilon$

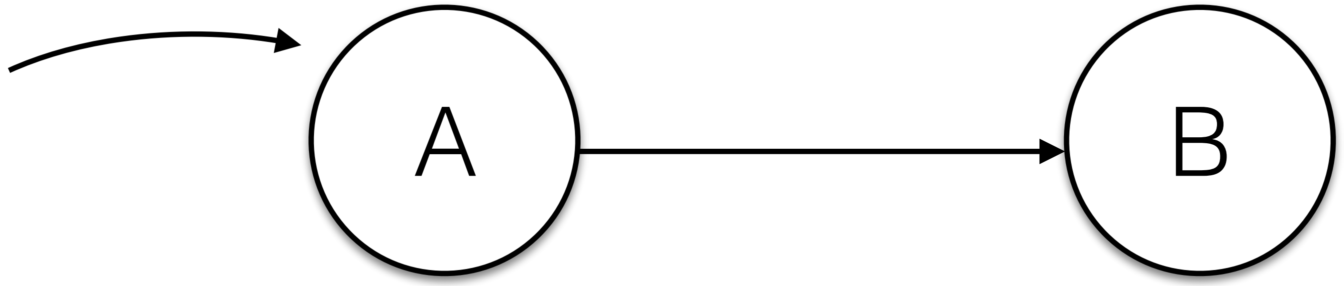
- If true generating mechanism is
- Will A predict B? $B = \tau + \beta A + \varepsilon$
  - Yes
- Will B predict A? $A = \tau + \beta B + \varepsilon$



How to estimate a network?

- Another familiar analysis to establish relationships among variables
  - regression
  - $y = \tau_1 + \beta_{12}X_2 + \beta_{13}X_3 + \ldots + \varepsilon$

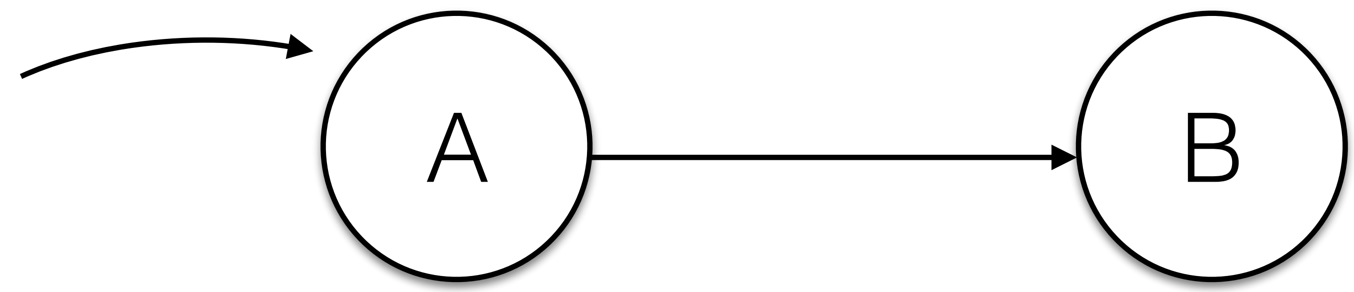
- If true generating mechanism is
- Will A predict B? $B = \tau + \beta A + \varepsilon$
  - Yes
- Will B predict A? $A = \tau + \beta B + \varepsilon$
  - Yes!

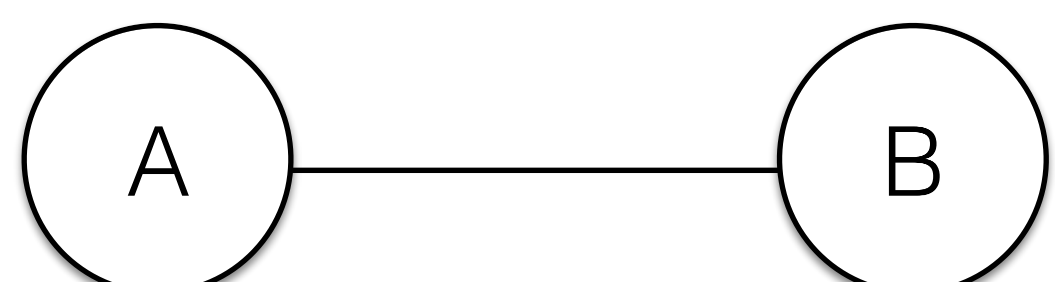


How to estimate a network?

- Another familiar analysis to establish relationships among variables
  - regression
  - $y = \tau_1 + \beta_{12}X_2 + \beta_{13}X_3 + \ldots + \varepsilon$

- If true generating mechanism is
- Will A predict B? $B = \tau + \beta A + \varepsilon$
  - Yes
- Will B predict A? $A = \tau + \beta B + \varepsilon$
  - Yes!
- Because we have cross sectional data, we don't know the direction. Therefore, often undirected networks.



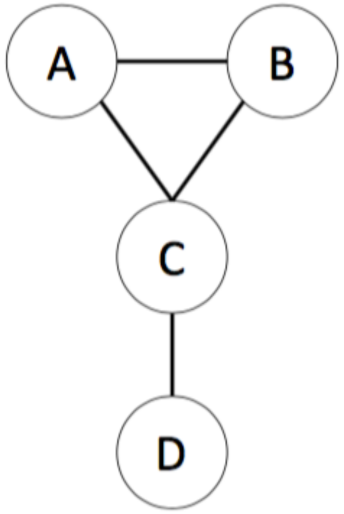


How to estimate a network?

- $y = \tau_1 + \beta_{12}X_2 + \beta_{13}X_3 + \ldots \varepsilon$
- β_{12}: slope, regression coefficient
- relates to
  - correlation
  - partial correlation
  - explained variance
  - conditional independence

Relationships among variables

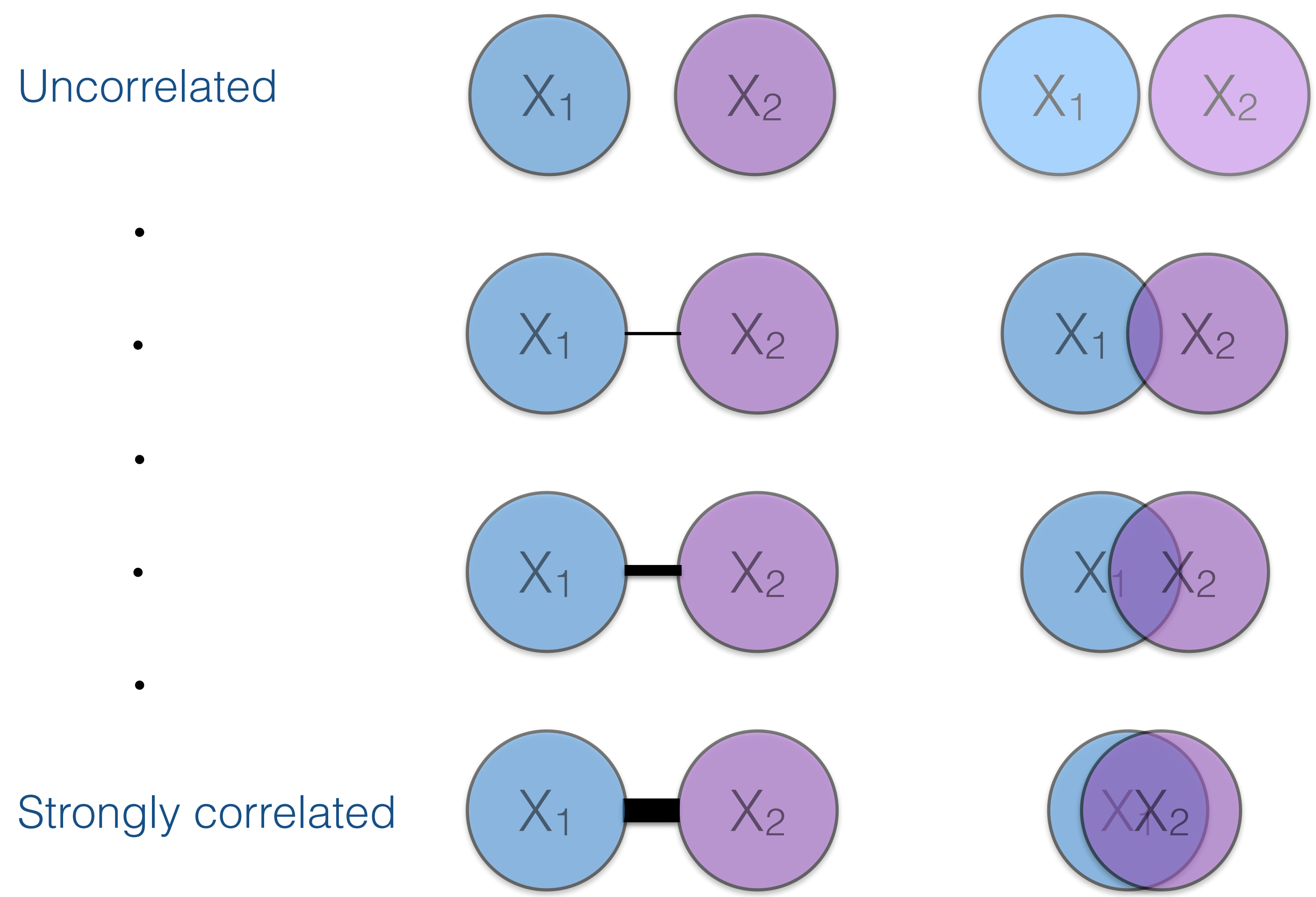
- $y = \tau_1 + \beta_{12}X_2 + \beta_{13}X_3 + \ldots \varepsilon$
- β_{12}: slope, regression coefficient
- relates to
  - correlation
  - partial correlation
  - explained variance
  - conditional independence
- These concepts are all somehow related and understanding this is essential when working with **graphical models**
- **What does it mean to say that two symptoms are connected in a network?**



Relationships among variables

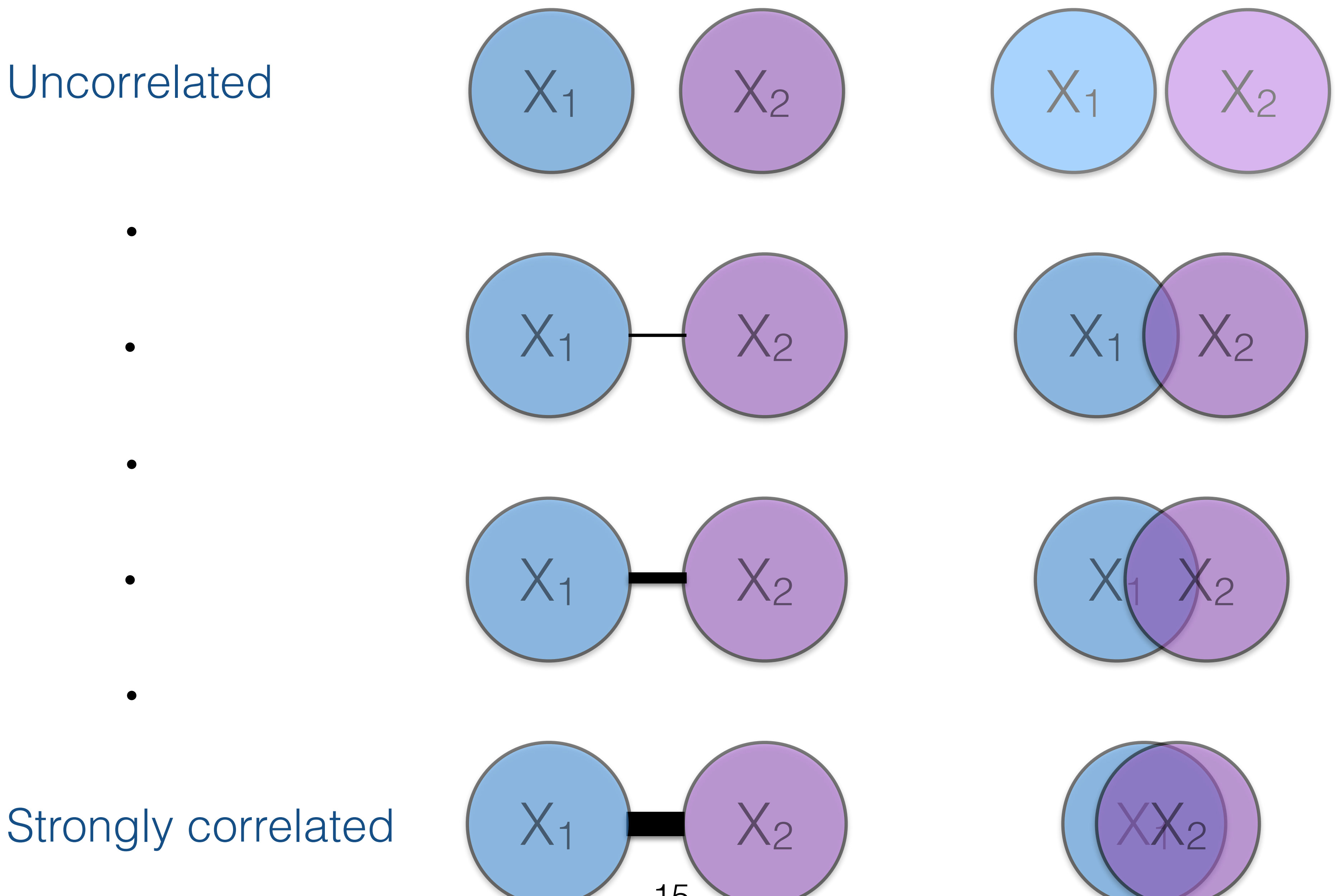
Let's start with correlations

- Are X1 and X2 correlated?



Relationships among variables

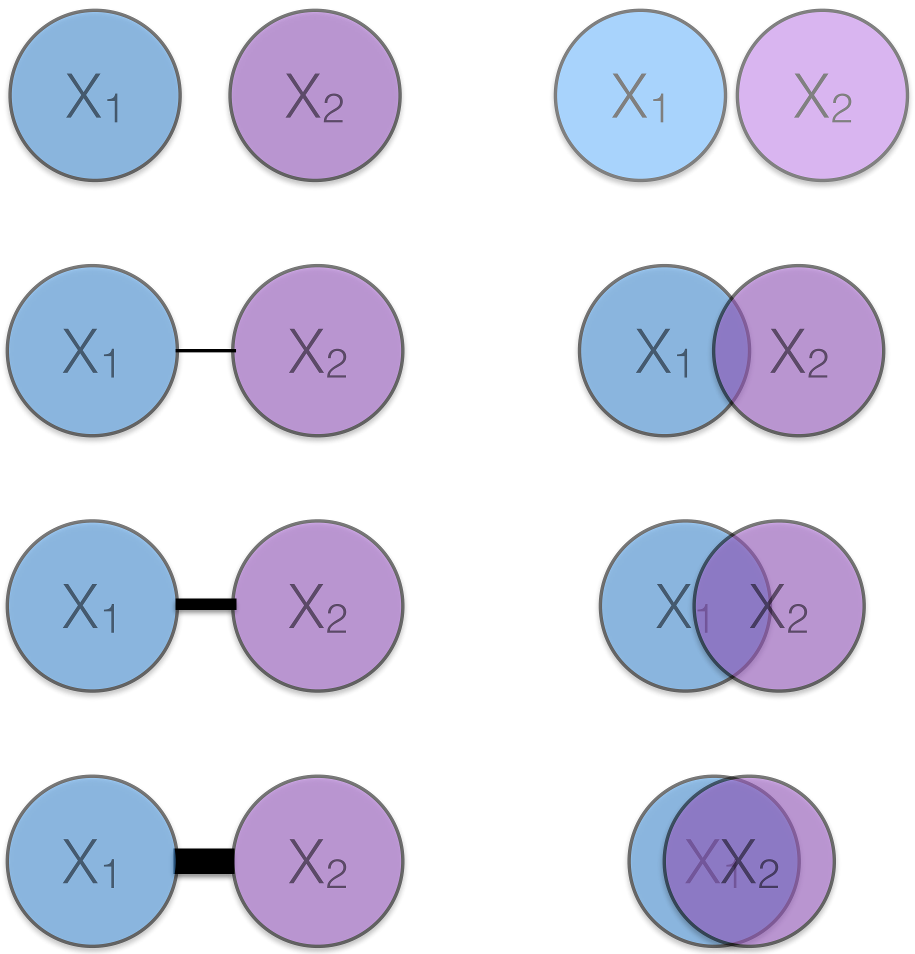
Think of correlations as **varying connection strength** or as **shared variance**



Relationships among variables

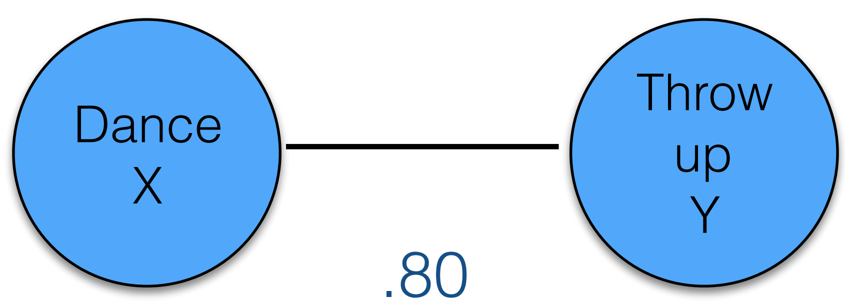
Let's start with correlations

- Are X1 and X2 correlated?
- Say, $r = .3$
- This means that 9% of the variance in X1 is explained by the variance in X2
- Note: .09 is the R^2 that you get from regression!



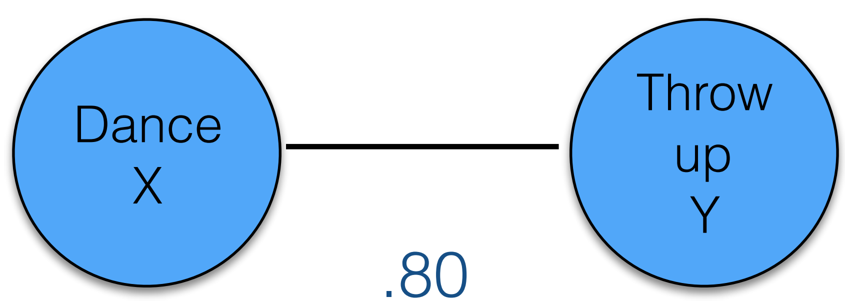
Relationships among variables

Example: two events at party



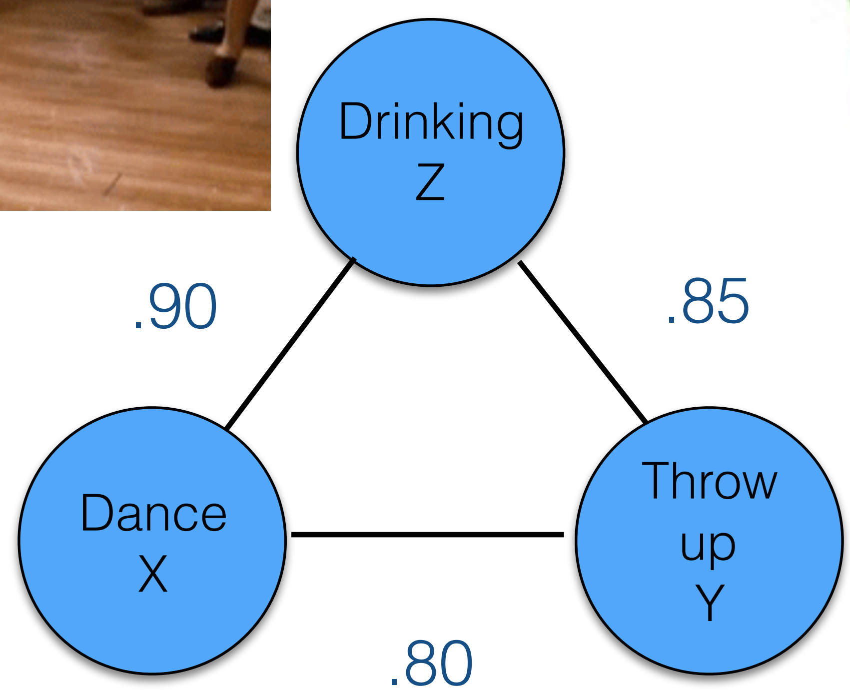
Relationships among variables

Example: two events at party



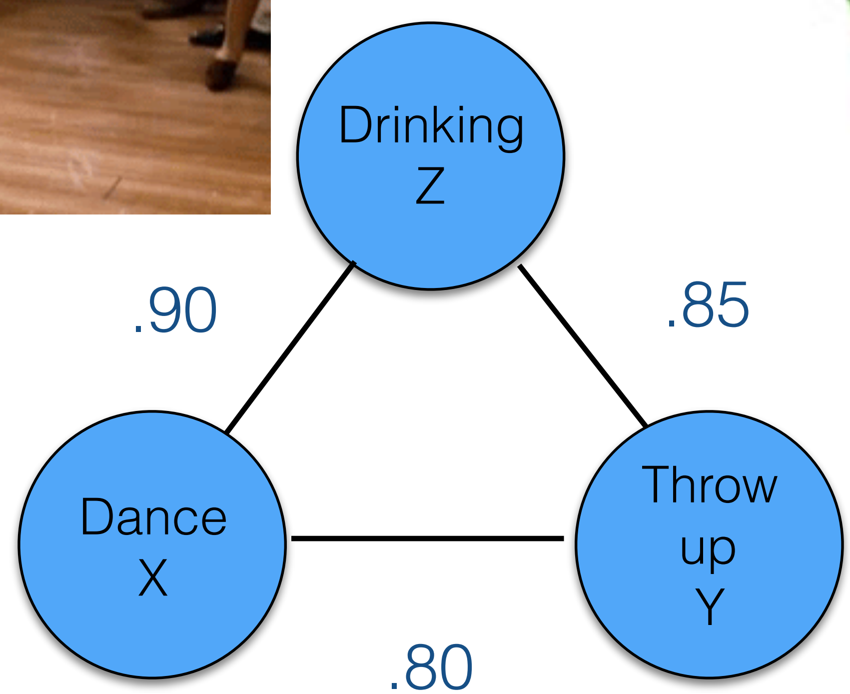
Relationships among variables

Example: two events at party

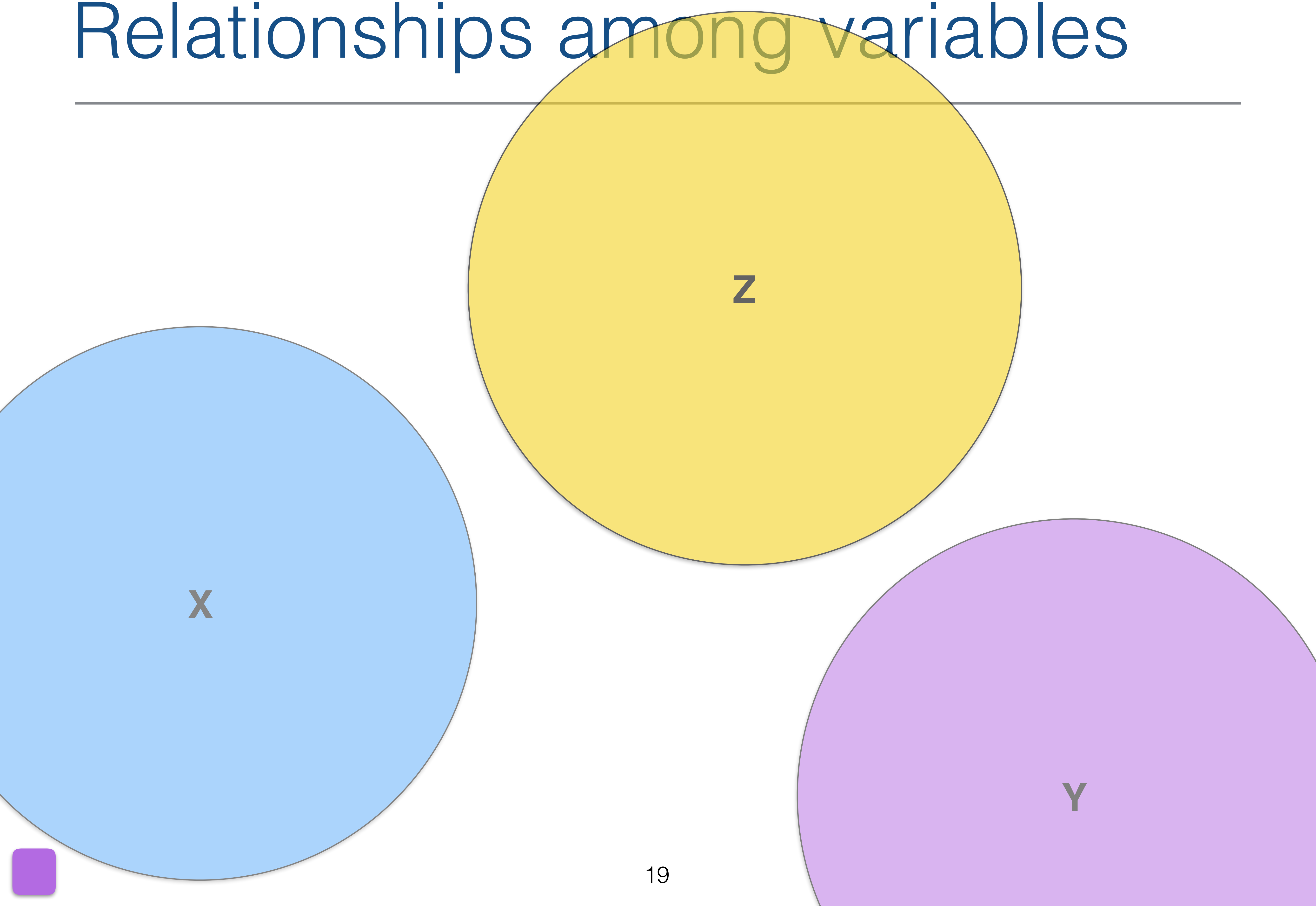


Relationships among variables

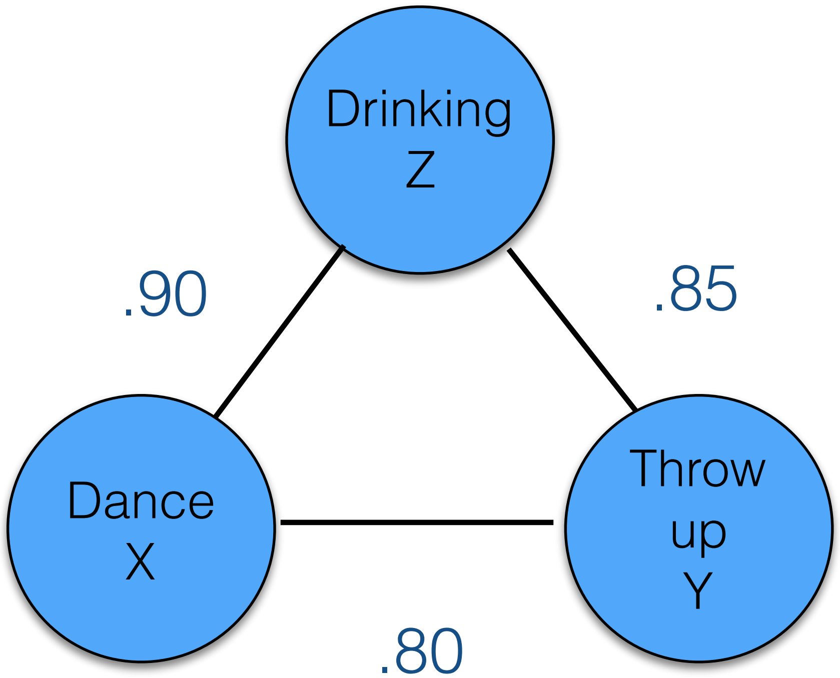
Example: two events at party



Relationships among variables

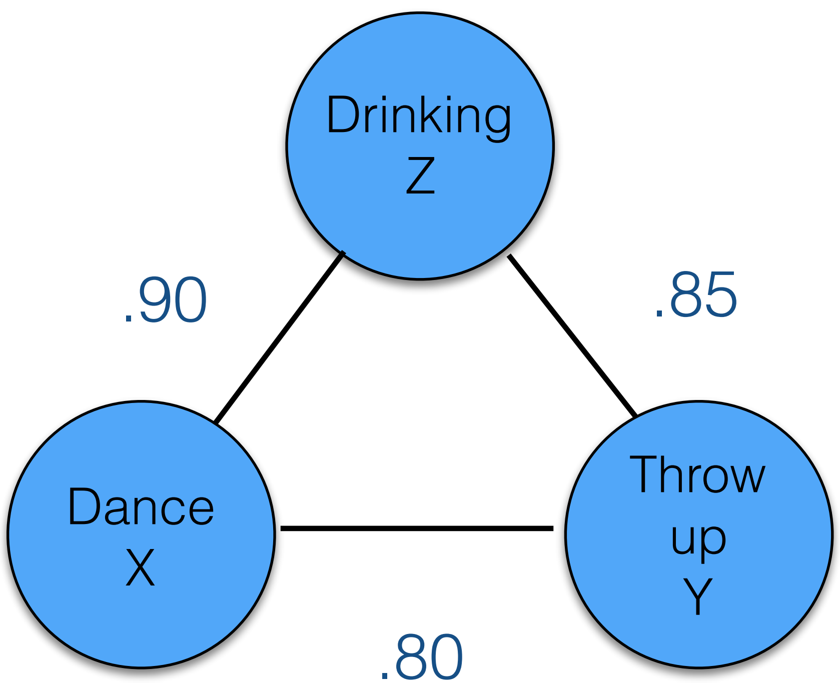


Relationships among variables



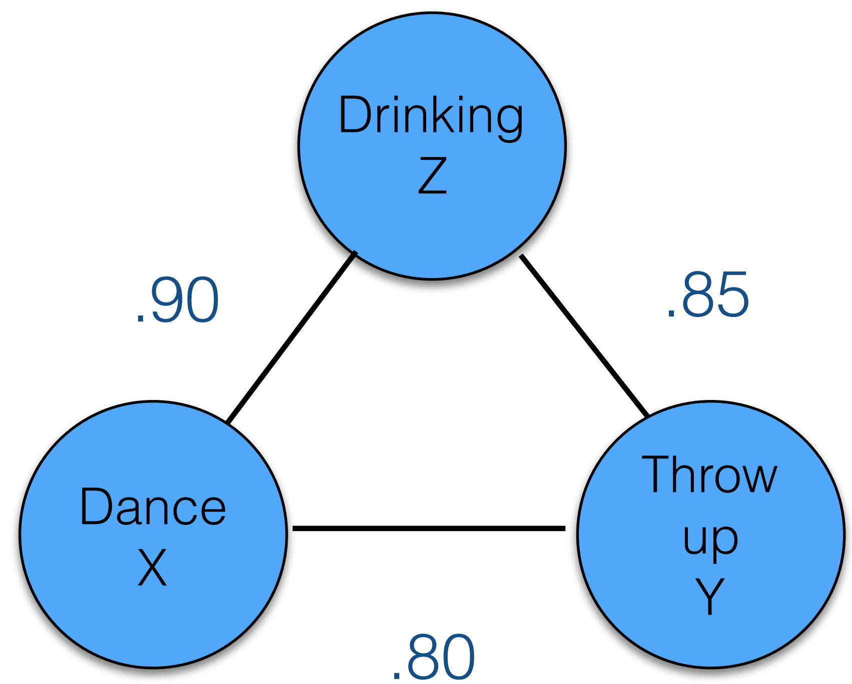
Relationships among variables

- What is the unique correlation between any pair of variables?



Relationships among variables

- What is the unique correlation between any pair of variables?
- Do the math: calculate the partial correlations

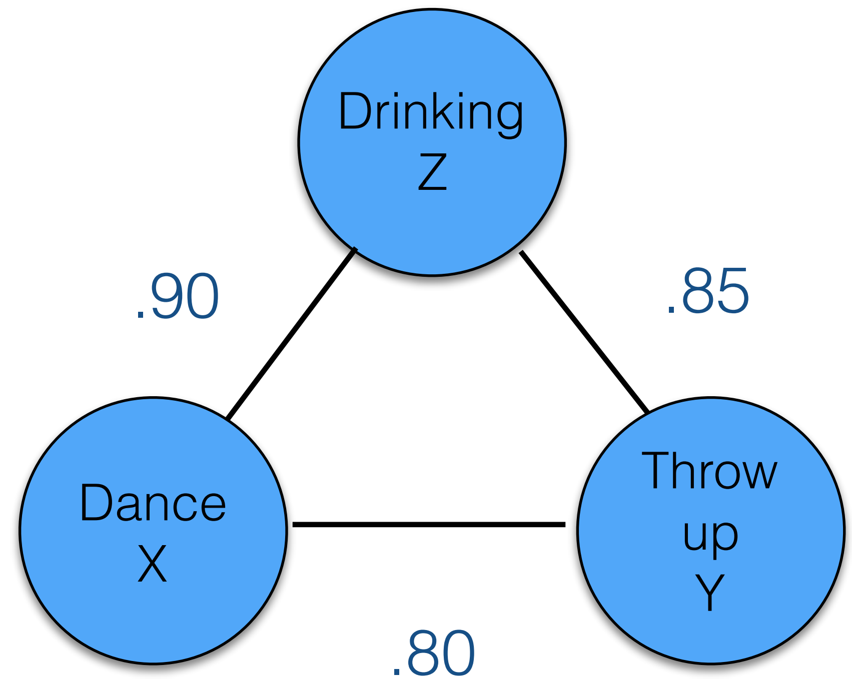


Partial correlation

$$\rho_{XY\cdot Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{ZY}}{\sqrt{1-\rho_{XZ}^2}\sqrt{1-\rho_{ZY}^2}}$$

Relationships among variables

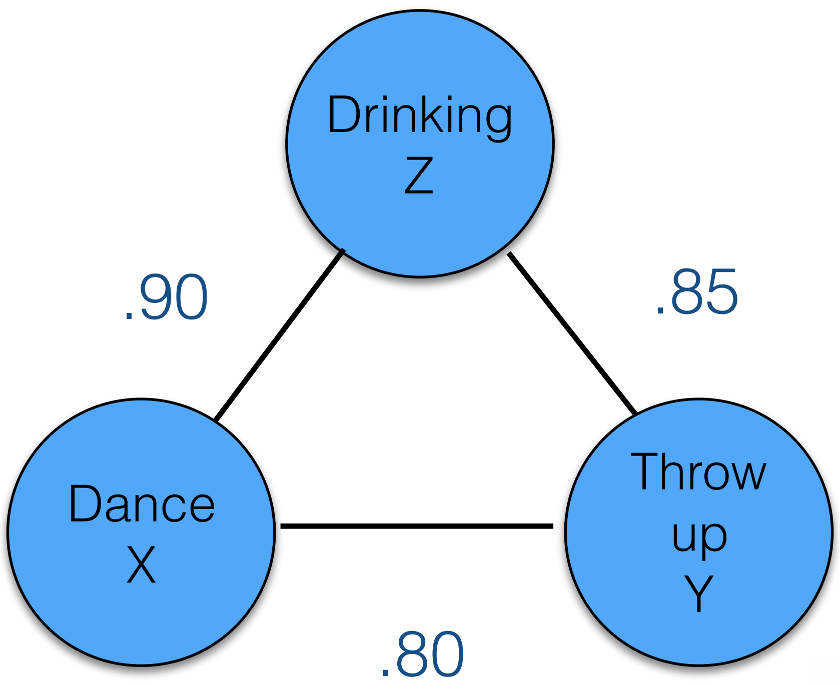
- What is the unique correlation between any pair of variables?
- Do the math: calculate the partial correlations
- Draw a schematic Venn diagram of the three variables and mark the found partial correlation area



Partial correlation

$$\rho_{XY \cdot Z} = \frac{\rho_{XY} - \rho_{XZ}\rho_{ZY}}{\sqrt{1-\rho_{XZ}^2}\sqrt{1-\rho_{ZY}^2}}$$

Relationships among variables



$r_{xy} = .80$

$r_{xz} = .90$

$r_{yz} = .85$

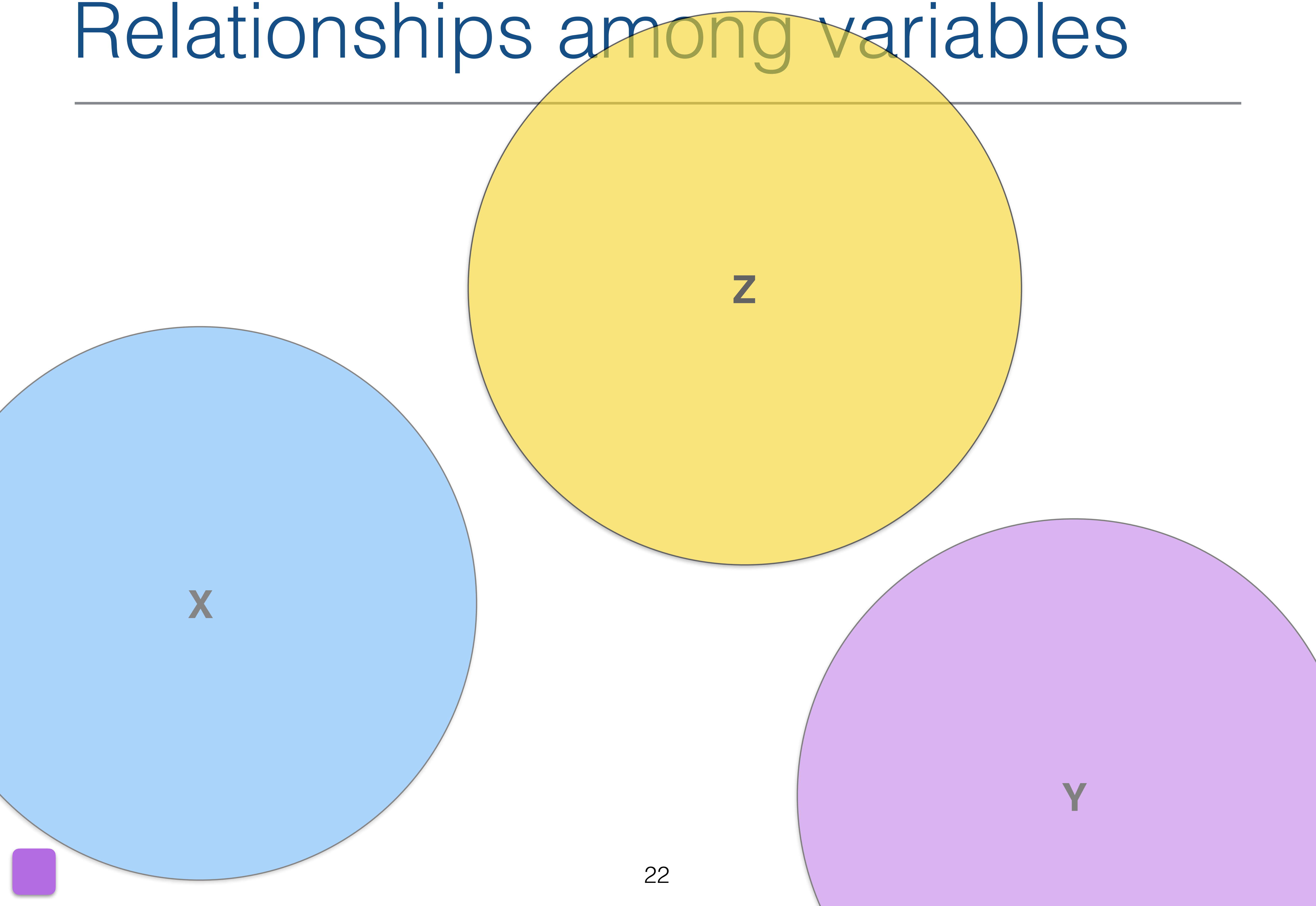
$r^2_{xz} = .81$

$r^2_{yz} = .723$

$$r_{xy.z} = \frac{r_{xy} - (r_{xz}\, r_{yz})}{\sqrt{(1 - r^2_{xz})(1 - r^2_{yz})}} = \frac{.80 - (.90 \cdot .85)}{\sqrt{(1 - .81)(1 - .723)}} =$$

$$\frac{.80 - .765}{\sqrt{(.19)\,(.277)}} = \frac{.80 - .765}{\sqrt{.053}} = \frac{.035}{.230} = .152 = r_{xy.z}$$

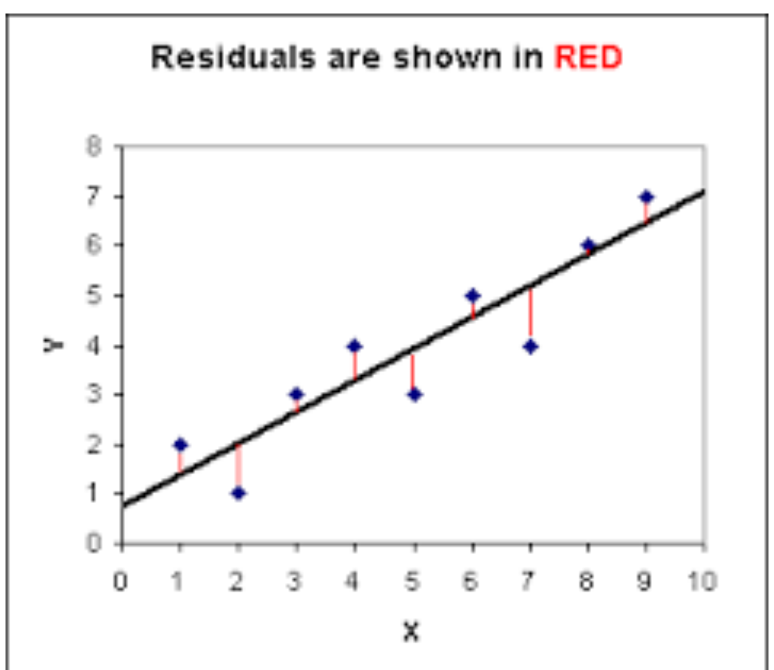
Relationships among variables

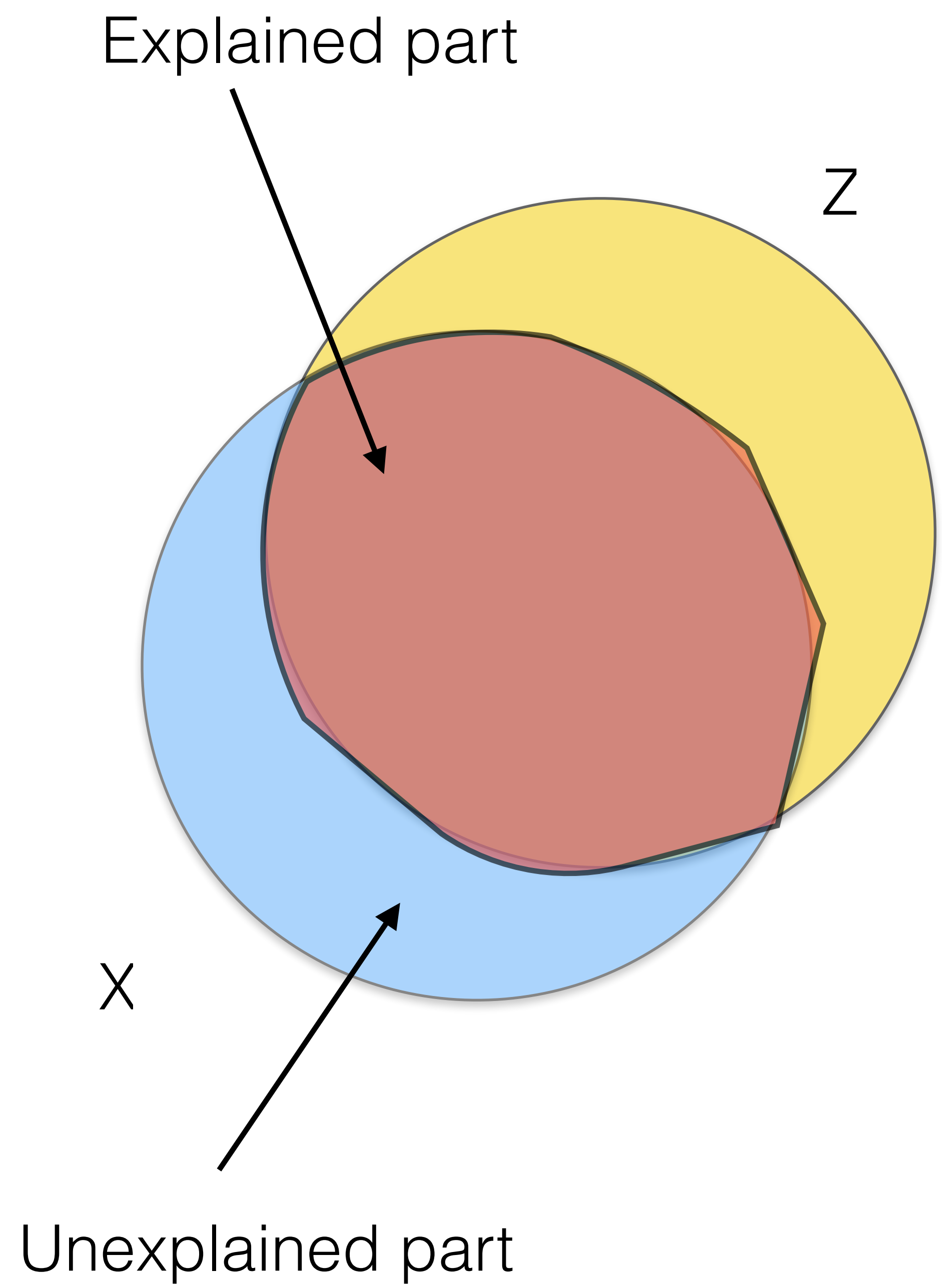


Relationships among variables

Another way of looking at this (partialling out):

- Regress Z on X: $Z = \beta_{0x} + \beta_x X + \varepsilon_x$
- Regress Z on Y: $Z = \beta_{0y} + \beta_y Y + \varepsilon_y$
- Correlate the unexplained parts (residuals; ε_x and ε_y)

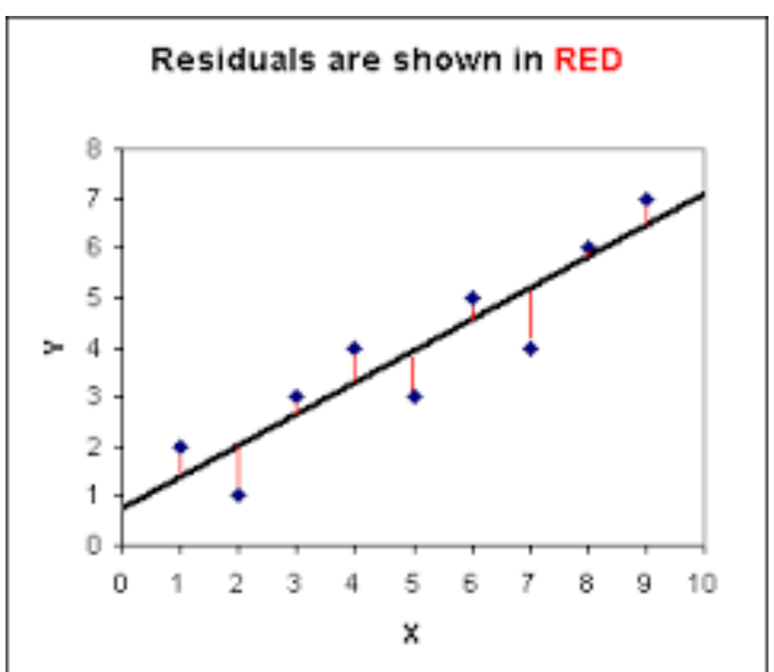


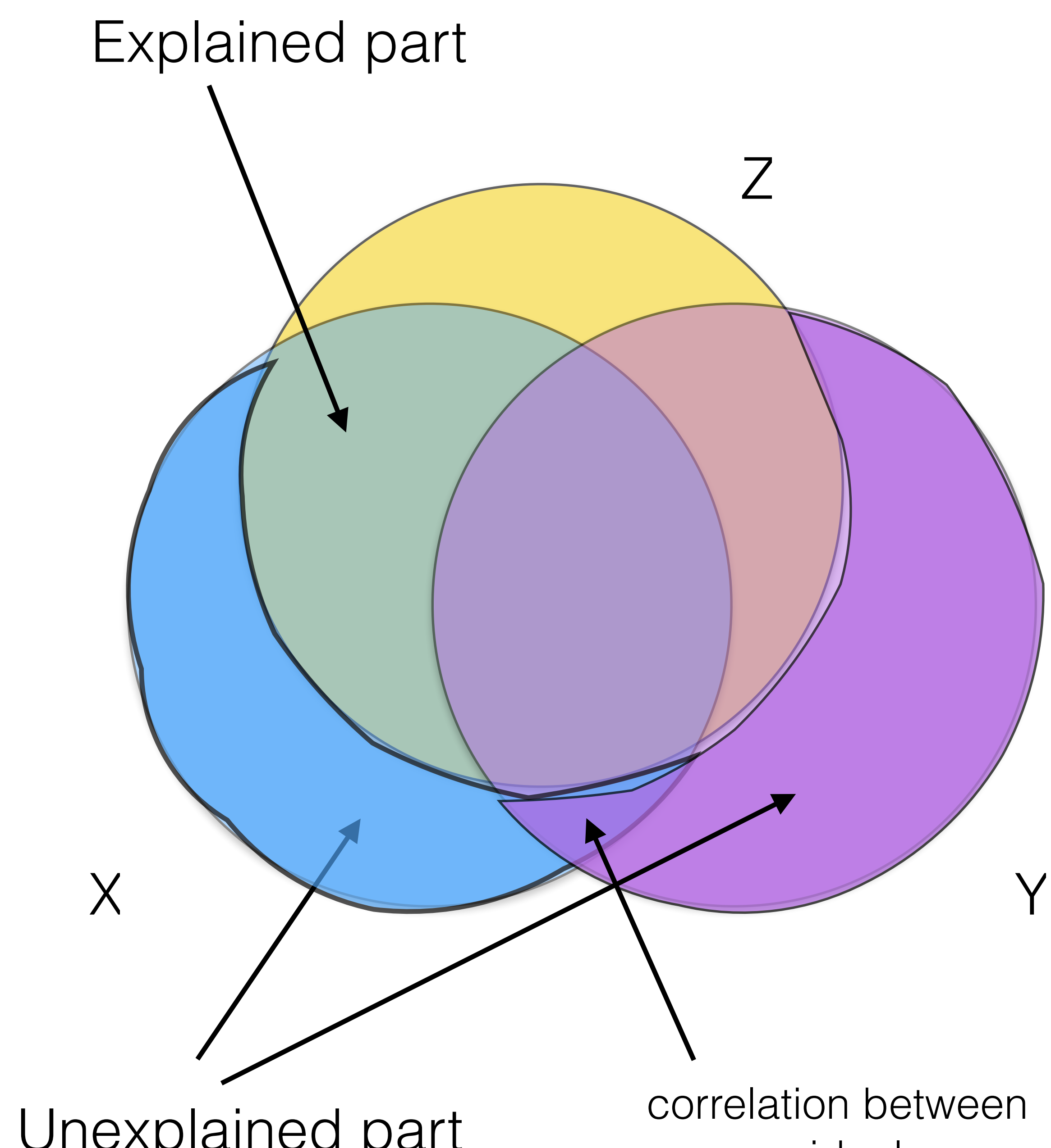


Relationships among variables

Another way of looking at this:

- Regress Z on X: $Z = \beta_{0x} + \beta_x X + \varepsilon_x$
- Regress Z on Y: $Z = \beta_{0y} + \beta_y Y + \varepsilon_y$
- Correlate the unexplained parts (residuals; ε_x and ε_y)

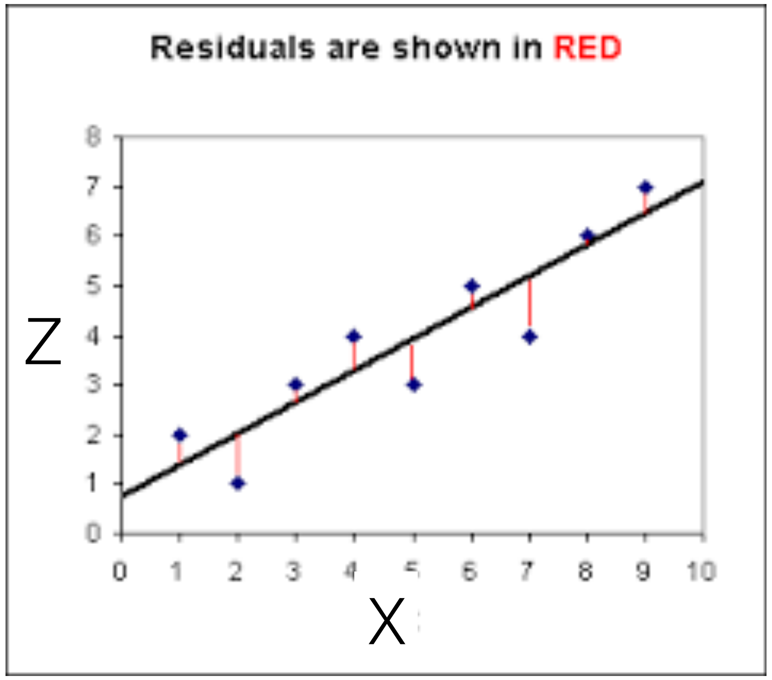


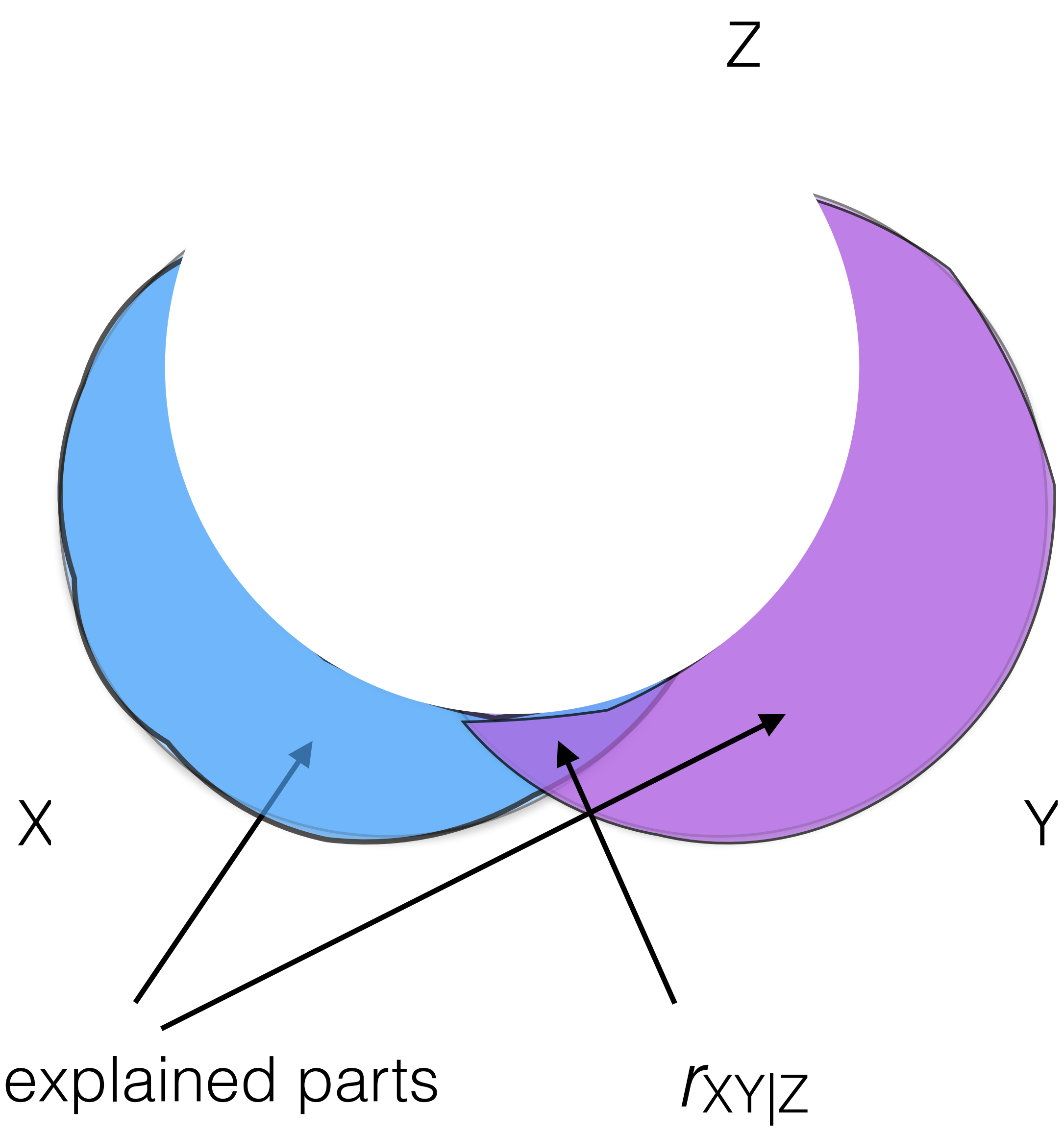


Relationships among variables

Another way of looking at this:

- Regress Z on X: $Z = \beta_{0x} + \beta_x X + \varepsilon_x$
- Regress Z on Y: $Z = \beta_{0y} + \beta_y Y + \varepsilon_y$
- Correlate the unexplained parts (residuals; ε_x and ε_y)





Quiz!

1. Which statement is correct?

The partial correlation between X and Y

is the correlation between X and Y…

A. given Z

B. when controlled for Z

C. with Z partialled out

D. that cannot be explained by Z

Quiz!

2. How many statements are correct?

A partial correlation of zero between A and B means

A. *A and B are not correlated*

B. *A and B are independent* given *C*

C. *A and B are independent* when conditioned *on C*

D. $A \perp\!\!\!\perp B \mid C$

E. *A and B might be correlated, but knowing C makes them independent*

Conditional independence

Example

C: fair (50%) or biased (90 head%) coin

T_1: first toss (head or tail)

T_2: second toss (head or tail)

Conditional independence

Example

C: fair (50%) or biased (90 head%) coin

T_1: first toss (head or tail)

T_2: second toss (head or tail)

First scenario

- I have two coins in my pocket and give you one (without telling which)
- You toss it: head comes up
- What is the probability of head for the second toss (same coin)?

Conditional independence

Example

C: fair (50%) or biased (90 head%) coin

T_1: first toss (head or tail)

T_2: second toss (head or tail)

Second scenario

- I have two coins in my pocket and give you one and I tell you it's the fair coin
- You toss it: head comes up
- What is the probability of head for the second toss (same coin) now?

Conditional independence

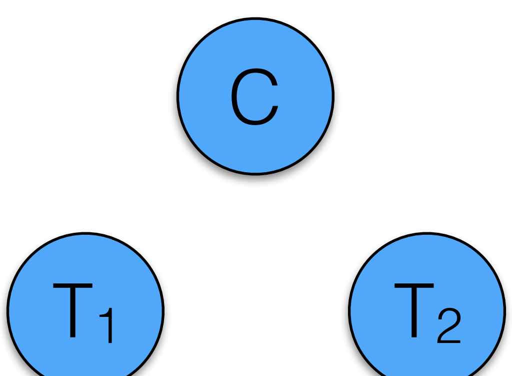
Example

C: fair (50%) or biased (90 head%) coin

T_1: first toss (head or tail)

T_2: second toss (head or tail)

T_1 and T_2 are correlated, but
knowing C makes them independent

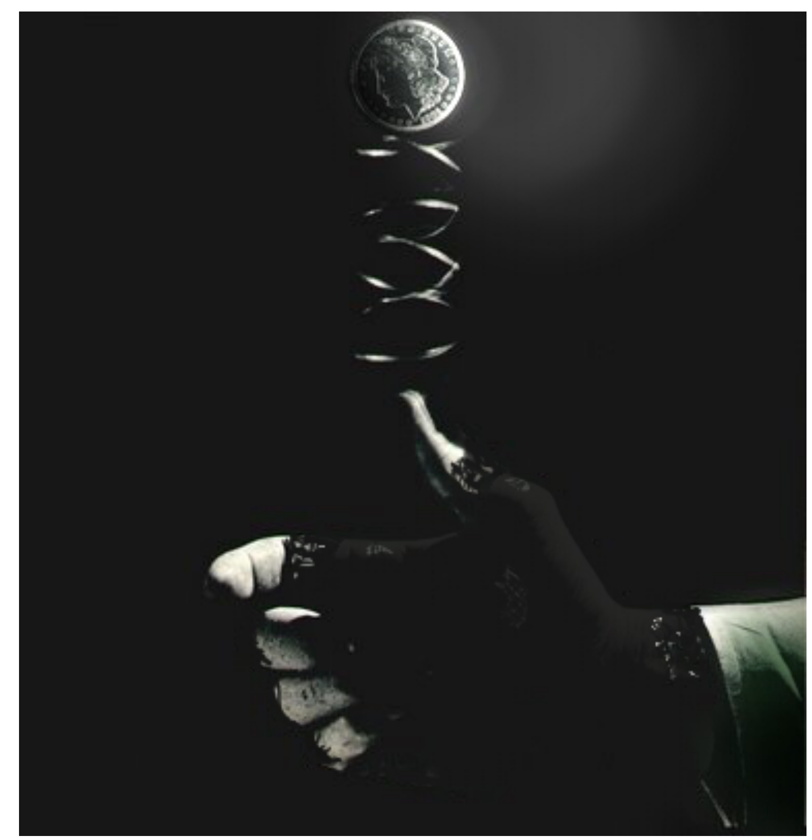


$T_1 \not\perp\!\!\!\perp T_2$

$T_1 \perp\!\!\!\perp T_2 \mid C$

'is independent of'

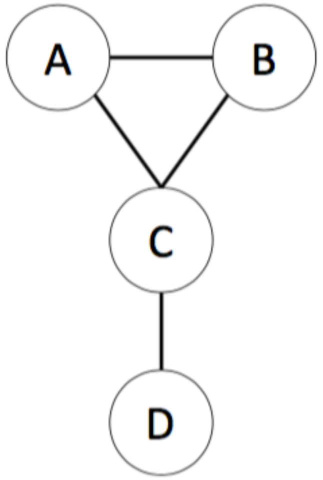
'given'



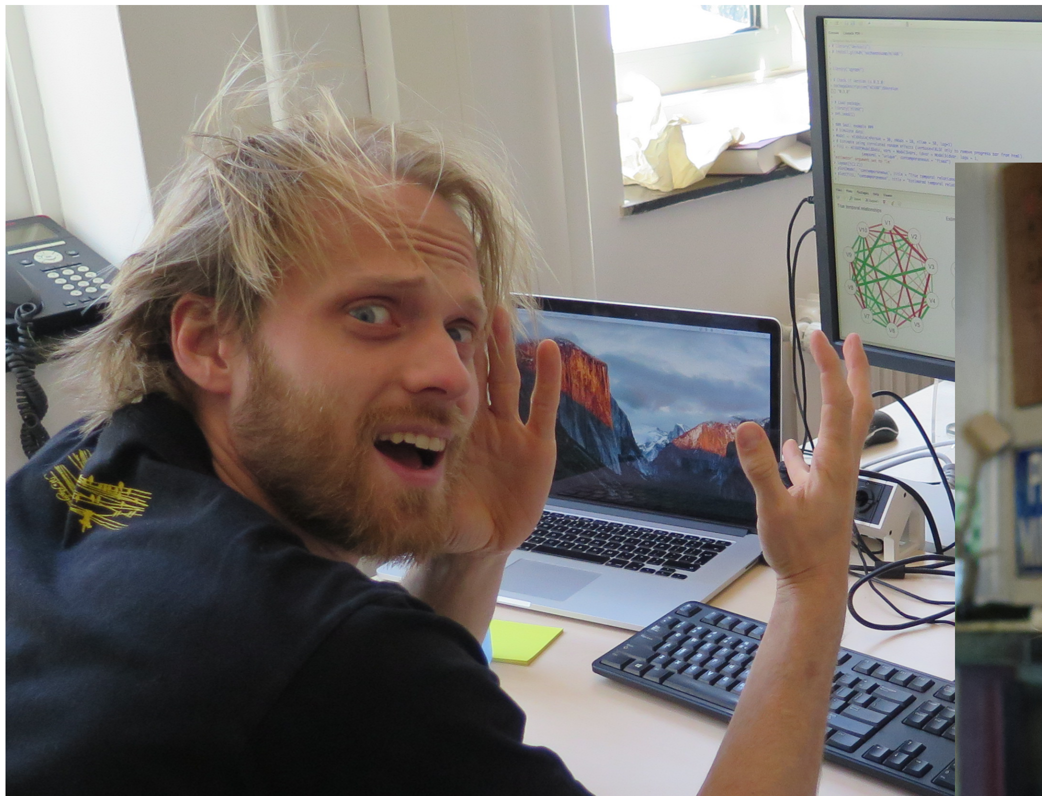
Recap

To conclude

- We are interested in conditional independence (CI) relationships in networks
    - we want to know if (and to what extent) symptoms A and B are related after controlling for all other symptoms
- Partial correlations provide information about CI's
- And that is what *graphical models* display
- You learned how the concepts of 'correlations', 'partial correlations', 'regression coefficients', 'shared variance', 'controlling for', 'conditioning on' are related.



Working with real data



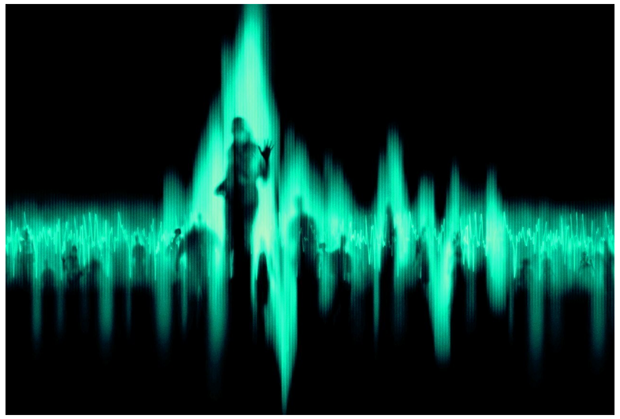
Working with real data

Things to take into account

- Types of variables (continuous, ordinal)

- Non-Normality (non-normal continuous data transformed using non paranormal transformation (Liu et al., 2009)

- Too many variables with too few participants

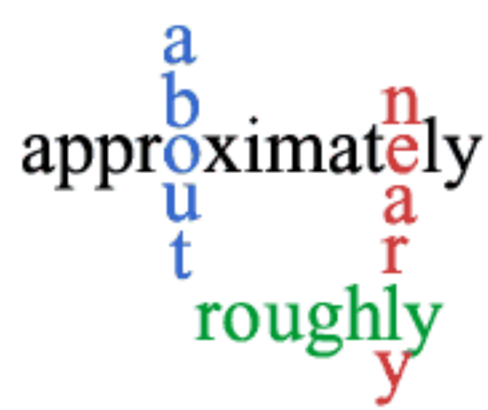


Practical

- Open Assignment_Day3_Part1.pdf
- Just follow the steps!
- If you go through the exercises quickly, you can try the exercises on your own data (if your data is cross-sectional)

Psychological Networks Amsterdam Winter School

Day 3, part 2
Estimating graphical models with L_1 regularization



Claudia van Borkulo

Februari 15th, 2017

Outline

1. The basics: Conditional independence

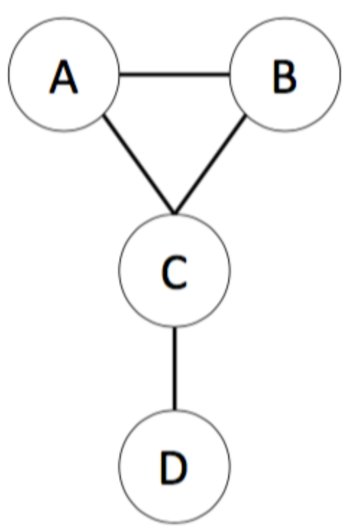
2. **Estimating graphical models with L1 regularization**

3. Recent advancements
    - Network stability
    - Network comparison

Recap

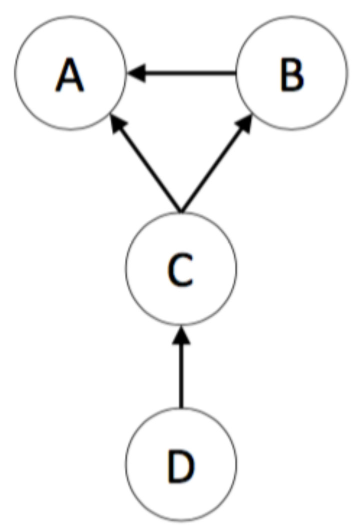
Graphical models

Markov Random Field (MRF)



Undirected graph

Bayesian Network (BN)



Directed acyclic graph

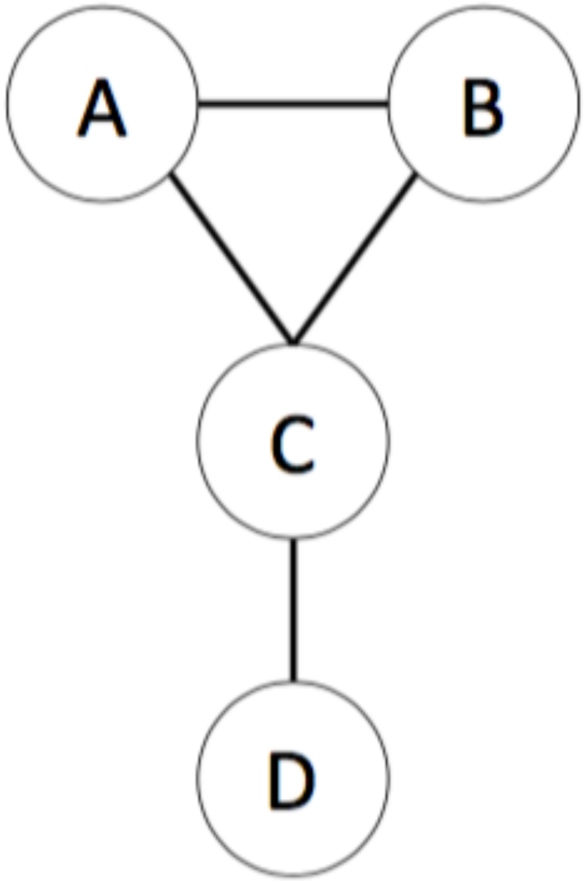
- A graphical model represents the conditional dependence structure among random variables
- The variables (e.g., symptoms) have pairwise/direct (causal) relationships
- An edge in the undirected MRF version can be viewed as a potential causal pathway
- A missing edge means that variables are conditionally independent, i.e., independent given all other variables:

$$X_i \perp\!\!\!\perp X_j | X_{V \setminus \{i,j\}}$$

Markov random fields

Two types

- Ising Model
  - for binary data
  - cross-sectional
  - estimation based on multiple logistic regression models
- Gaussian Graphical Model (GGM)
  - for normally distributed data
  - cross-sectional
  - estimation based on multiple linear regression models (or inverse of covariance matrix)



Markov Random Field (MRF)
Undirected graph

Conditional independence

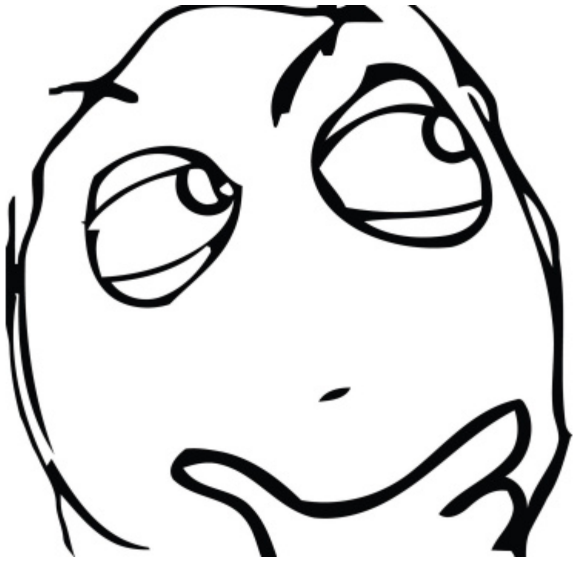
Conditional independence

CI when partial correlation is zero…. but partial correlations are often not exactly zero

Conditional independence

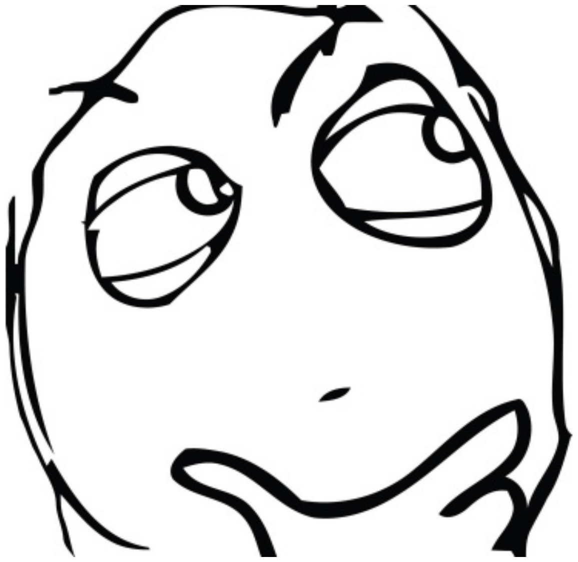
CI when partial correlation is zero…. but partial correlations are often not exactly zero

How to deal with that?



Conditional independence

CI when partial correlation is zero…. but partial correlations are often not exactly zero



How to deal with that?

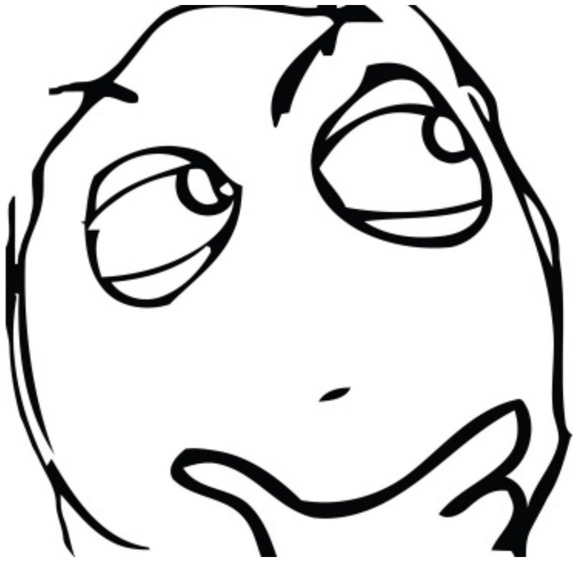
- use a threshold

arbitrary

Conditional independence

CI when partial correlation is zero…. but partial correlations are often not exactly zero

How to deal with that?



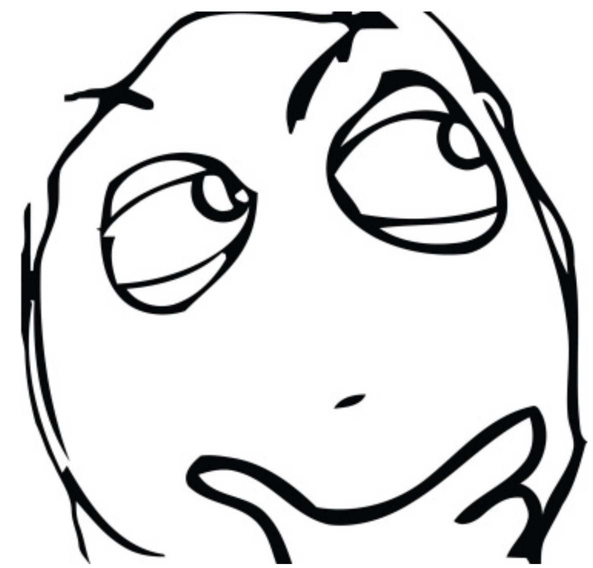
- use a threshold

  arbitrary

- significance tests

Conditional independence

CI when partial correlation is zero…. but partial correlations are often not exactly zero



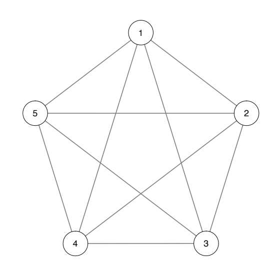
How to deal with that?

- use a threshold

  arbitrary

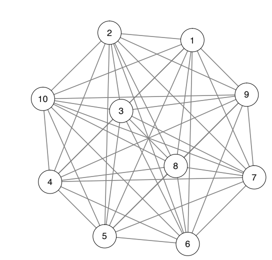
- significance tests

  Multiple testing problem: n(n-1)/2



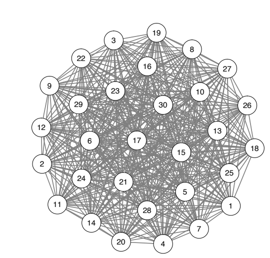
|V| = 5

|E| = 10



|V| = 10

|E| = 45

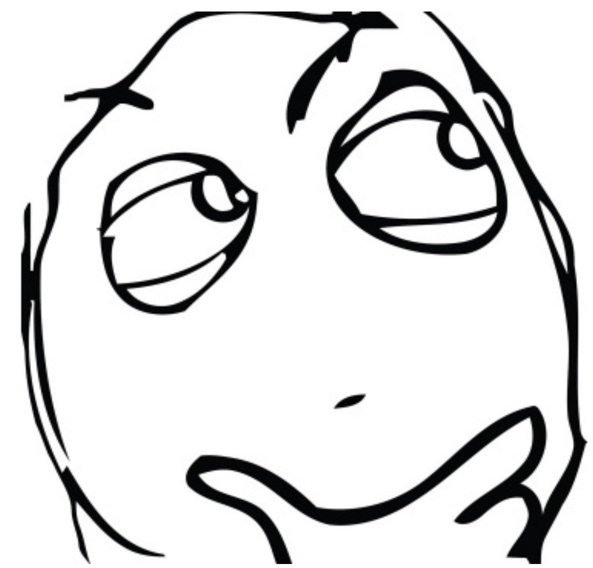


|V| = 30

|E| = 435

Conditional independence

CI when partial correlation is zero…. but partial correlations are often not exactly zero



How to deal with that?

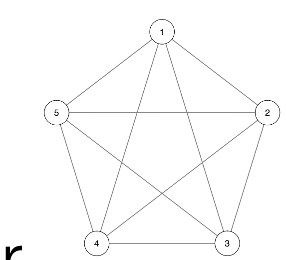
- use a threshold

  arbitrary

- significance tests

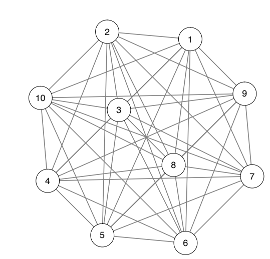
  Multiple testing problem: n(n-1)/2

  Bonferroni correction leads to loss of power



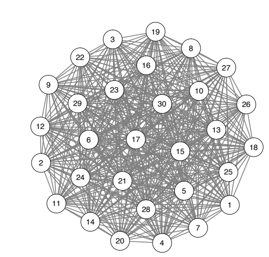
|V| = 5

|E| = 10



|V| = 10

|E| = 45



|V| = 30

|E| = 435

Conditional independence

CI when partial correlation is zero…. but partial correlations are often not exactly zero

How to deal with that?

- Use model selection to find the simplest model (sparse network) that fits best
- Impose an L_1 *penalty* on the coefficients

  (called ‘regularization')

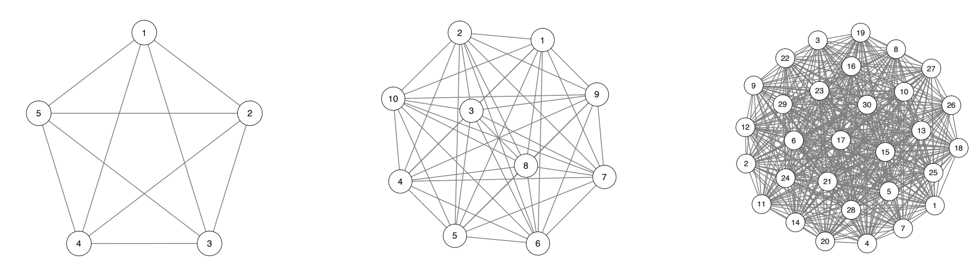
Conditional independence

CI when partial correlation is zero…. but partial correlations are often not exactly zero

How to deal with that?

- Use model selection to find the simplest model (sparse network) that fits best
- Impose an L_1 *penalty* on the coefficients

  (called 'regularization')



|V| = 5 |V| = 10 |V| = 30

|E| = 10 |E| = 45 |E| = 435

Conditional independence

Which coefficients can be regularized?

- Partial correlation coefficients
- Similar to elements of the **inverse covariance matrix** (only under multivariate normality)

} only with gaussian data

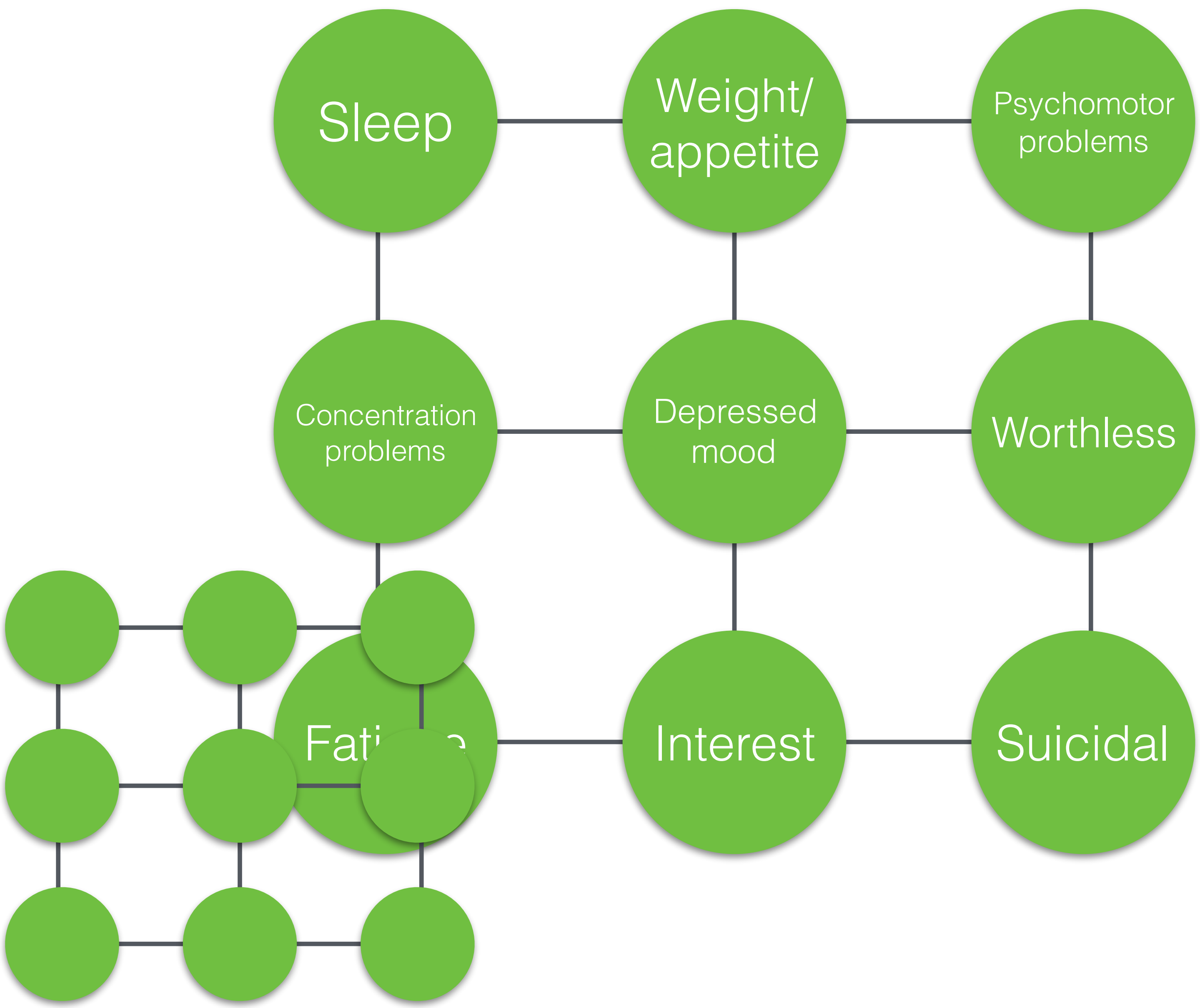
or

- regression coefficients of **node-wise regressions** (also for binary variables; logistic regressions)

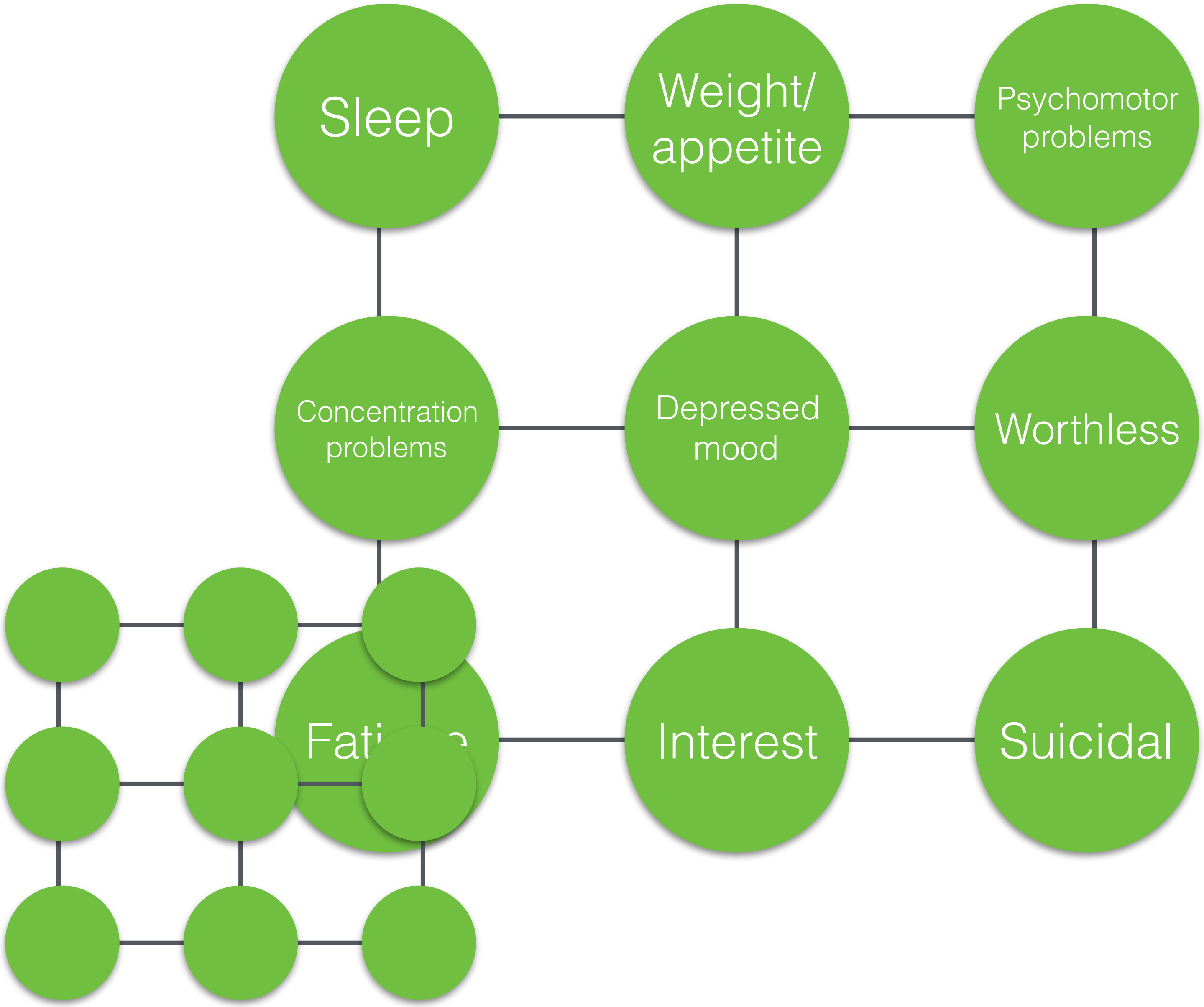
} with binary and gaussian data

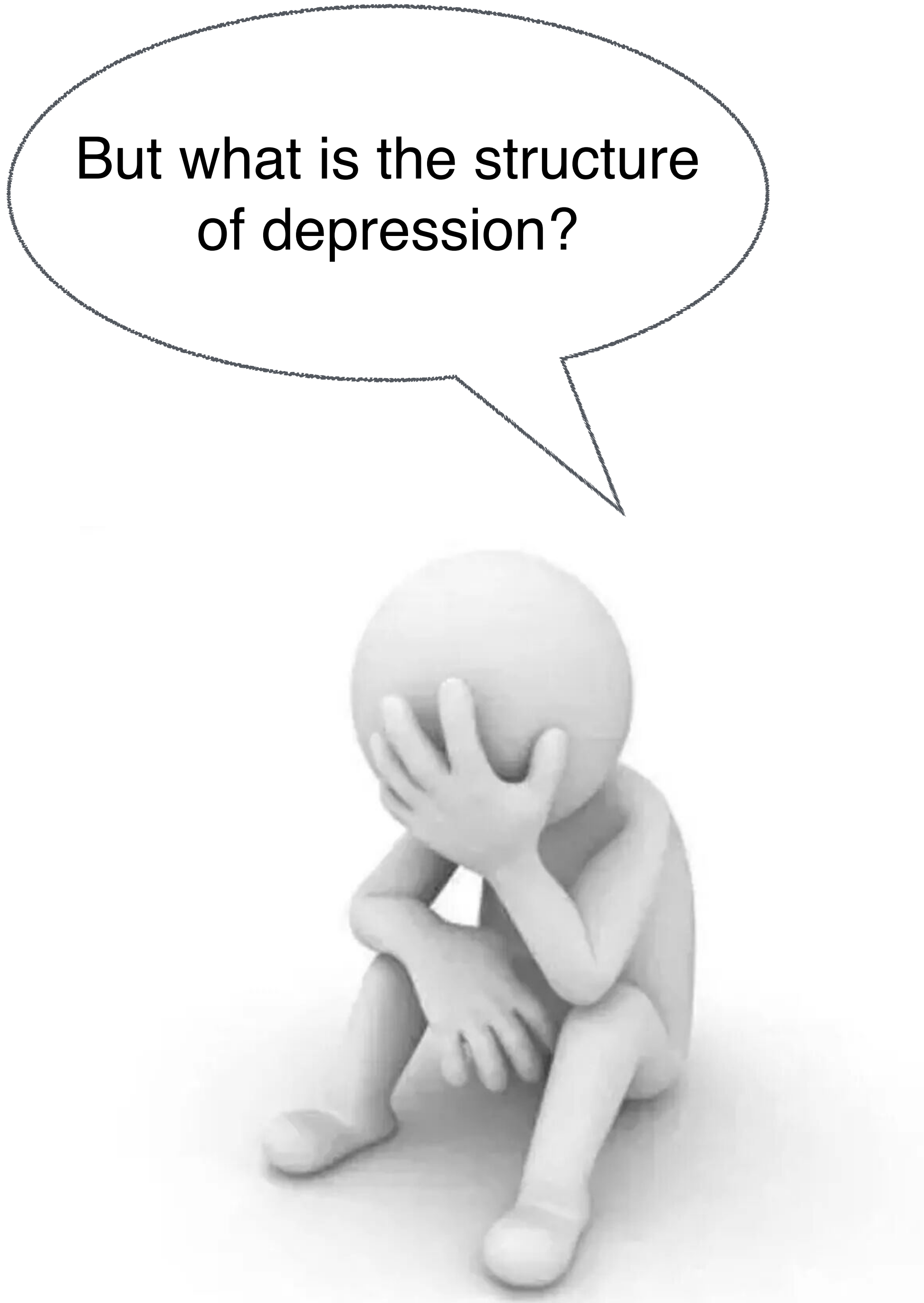
- The approximation with regression is computationally efficient and asymptotically consistent

Psychopathology

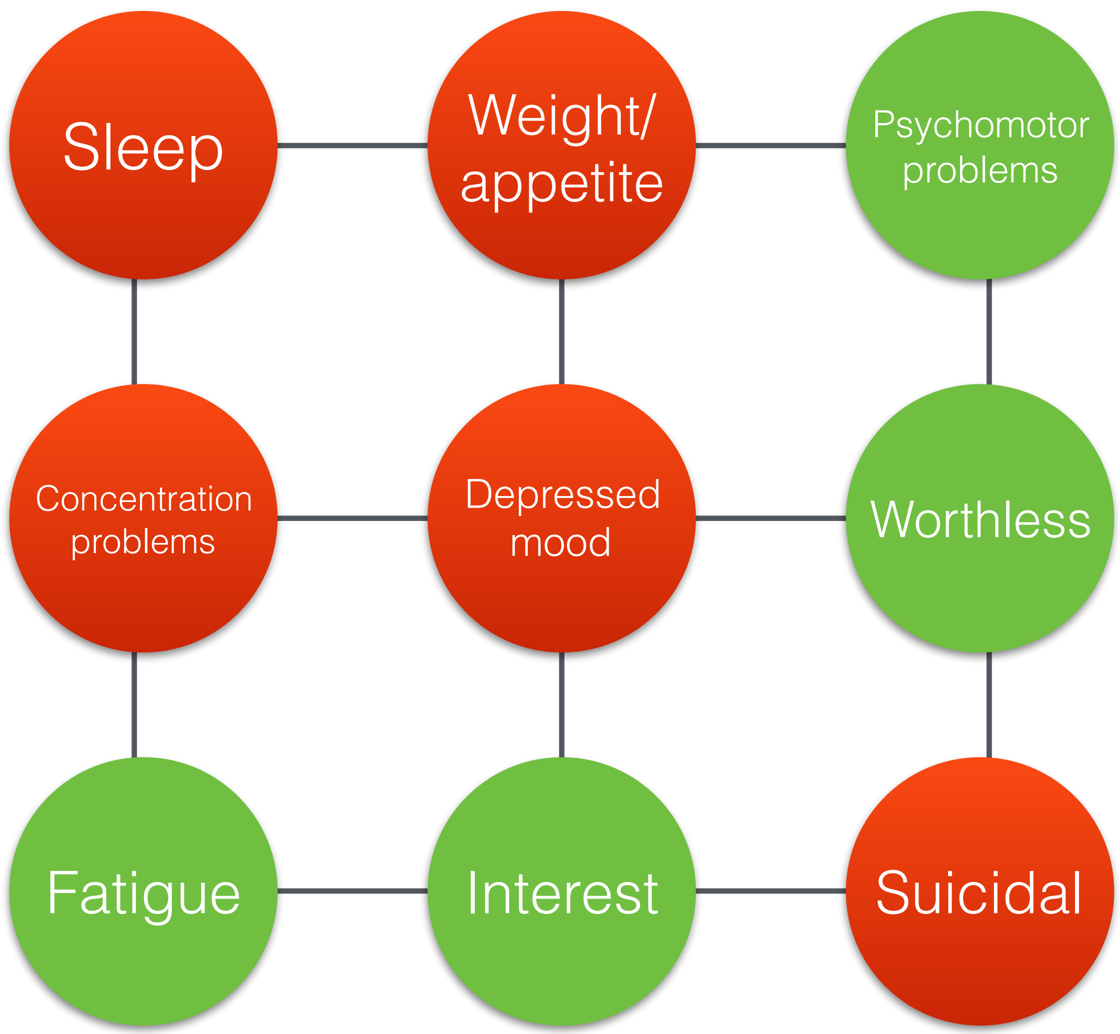


Psychopathology



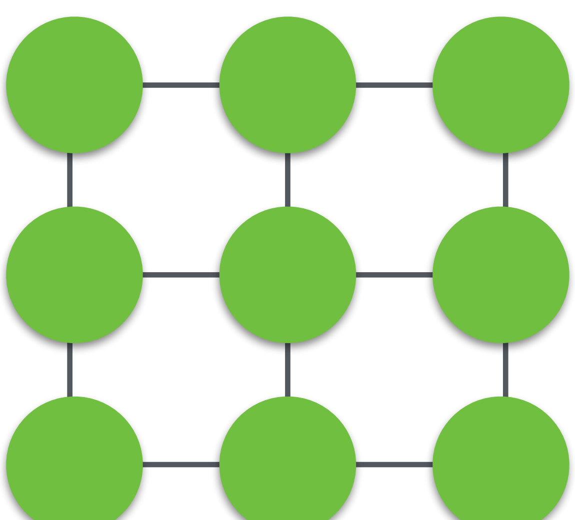


Psychopathology



Ising model

$$\mathbb{P}_{\Theta}\left(x_j \middle| x_{\setminus j}\right) = \frac{\exp\left[\tau_j x_j + x_j \sum_{k \in V_{\setminus j}} \beta_{jk} x_k\right]}{1 + \exp\left[\tau_j + \sum_{k \in V_{\setminus j}} \beta_{jk} x_k\right]}$$

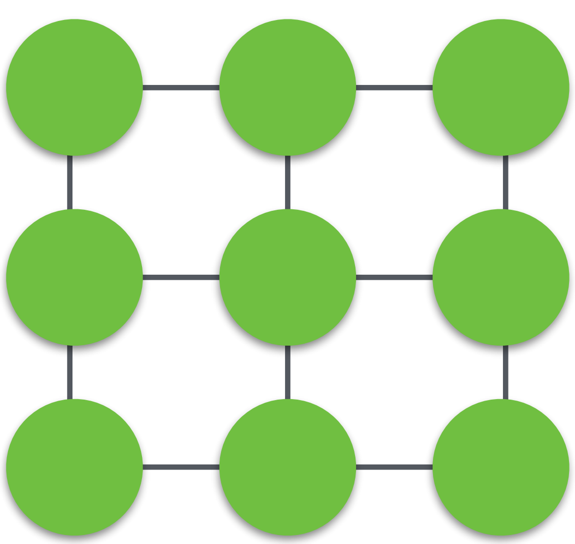


- x = (x1, x2, …,xn)
- x_j = 0 or 1
- τ_j: node parameter (threshold)
- β_{jk}: pairwise interaction parameter

Ising model

Conditional probability

$$\mathbb{P}_{\Theta}\left(x_j \middle| x_{\setminus j}\right) = \frac{\exp\left[\tau_j x_j + x_j \sum_{k \in V_{\setminus j}} \beta_{jk} x_k\right]}{1 + \exp\left[\tau_j + \sum_{k \in V_{\setminus j}} \beta_{jk} x_k\right]}$$

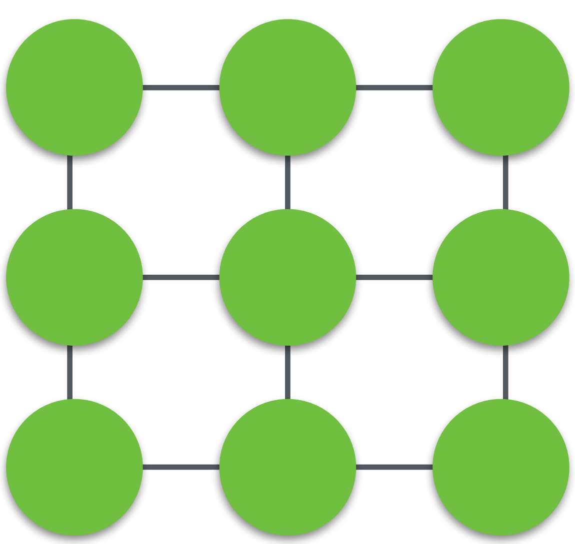


- $x = (x1, x2, \ldots, xn)$
- $x_j = 0$ or 1
- τ_j: node parameter (threshold)
- β_{jk}: pairwise interaction parameter

Ising model

Conditional probability

$$\mathbb{P}_{\Theta}\left(x_j \middle| x_{\setminus j}\right) = \frac{\exp\left[\tau_j x_j + x_j \sum_{k \in V_{\setminus j}} \beta_{jk} x_k\right]}{1 + \exp\left[\tau_j + \sum_{k \in V_{\setminus j}} \beta_{jk} x_k\right]}$$



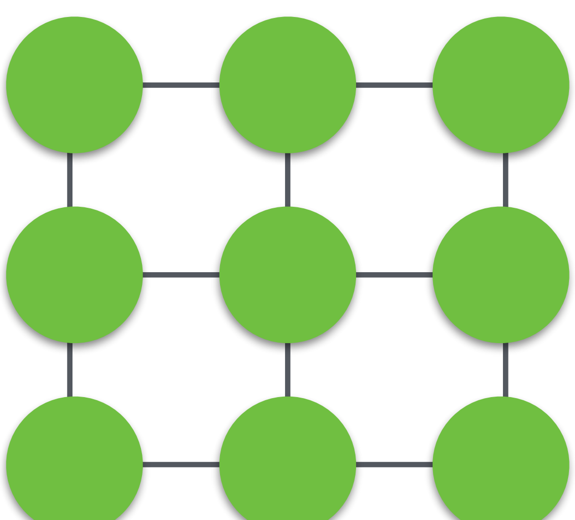
- x = (x1, x2, …,xn)
- x_j = 0 or 1
- τ_j: node parameter (threshold)
- β_{jk}: pairwise interaction parameter

Ising model

Conditional probability

$$\mathbb{P}_{\Theta}\left(x_j \middle| x_{\setminus j}\right) = \frac{\exp\left[\tau_j x_j + x_j \sum_{k \in V_{\setminus j}} \beta_{jk} x_k\right]}{1 + \exp\left[\tau_j + \sum_{k \in V_{\setminus j}} \beta_{jk} x_k\right]}$$

$\tau_1 + \beta_{12}X_2 + \beta_{13}X_3 + \ldots$



- x = (x1, x2, …,xn)
- x_j = 0 or 1
- τ_j: node parameter (threshold)
- β_{jk}: pairwise interaction parameter

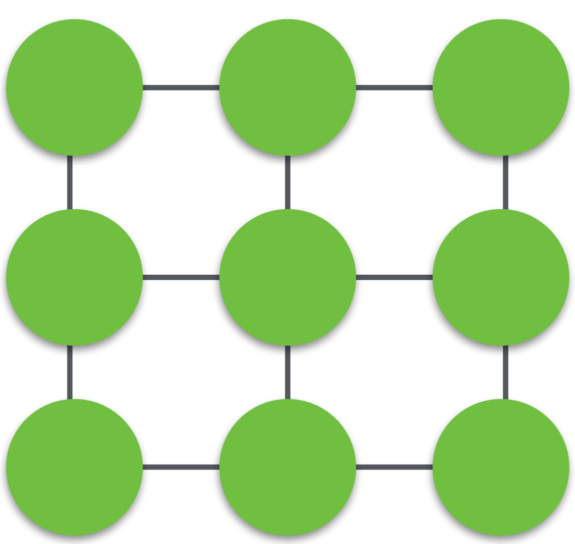
Ising model

Autonomous disposition of x_j

Conditional probability

$$\mathbb{P}_{\Theta}(x_j | x_{\setminus j}) = \frac{\exp\left[\tau_j x_j + x_j \sum_{k \in V_{\setminus j}} \beta_{jk} x_k\right]}{1 + \exp\left[\tau_j + \sum_{k \in V_{\setminus j}} \beta_{jk} x_k\right]}$$

$\tau_1 + \beta_{12}X_2 + \beta_{13}X_3 + \dots$



- $x = (x1, x2, \dots, xn)$
- $x_j = 0$ or 1
- τ_j: node parameter (threshold)
- β_{jk}: pairwise interaction parameter

Ising model

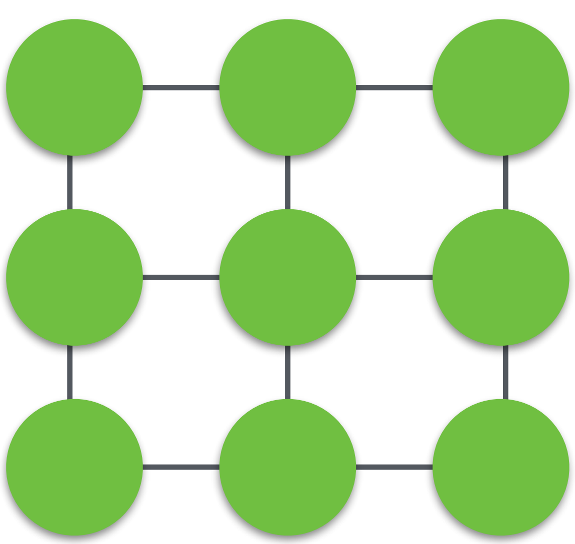
Conditional probability

Autonomous disposition of x_j

Interaction strength between x_j and x_k

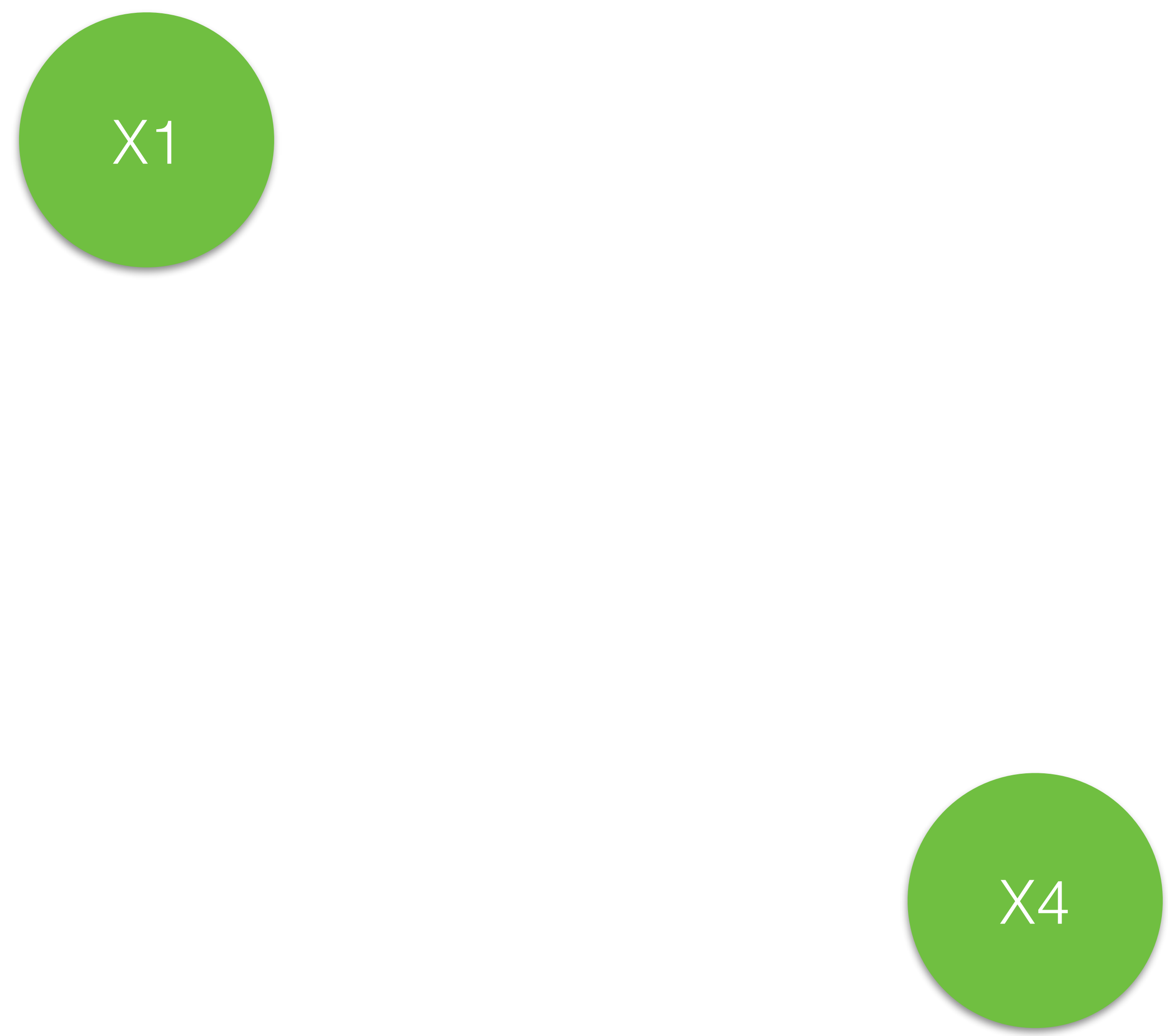
$$\mathbb{P}_\Theta\left(x_j \mid x_{\setminus j}\right) = \frac{\exp\left[\tau_j x_j + x_j \sum_{k \in V_{\setminus j}} \beta_{jk} x_k\right]}{1 + \exp\left[\tau_j + \sum_{k \in V_{\setminus j}} \beta_{jk} x_k\right]}$$

$\tau_1 + \beta_{12}X_2 + \beta_{13}X_3 + \ldots$



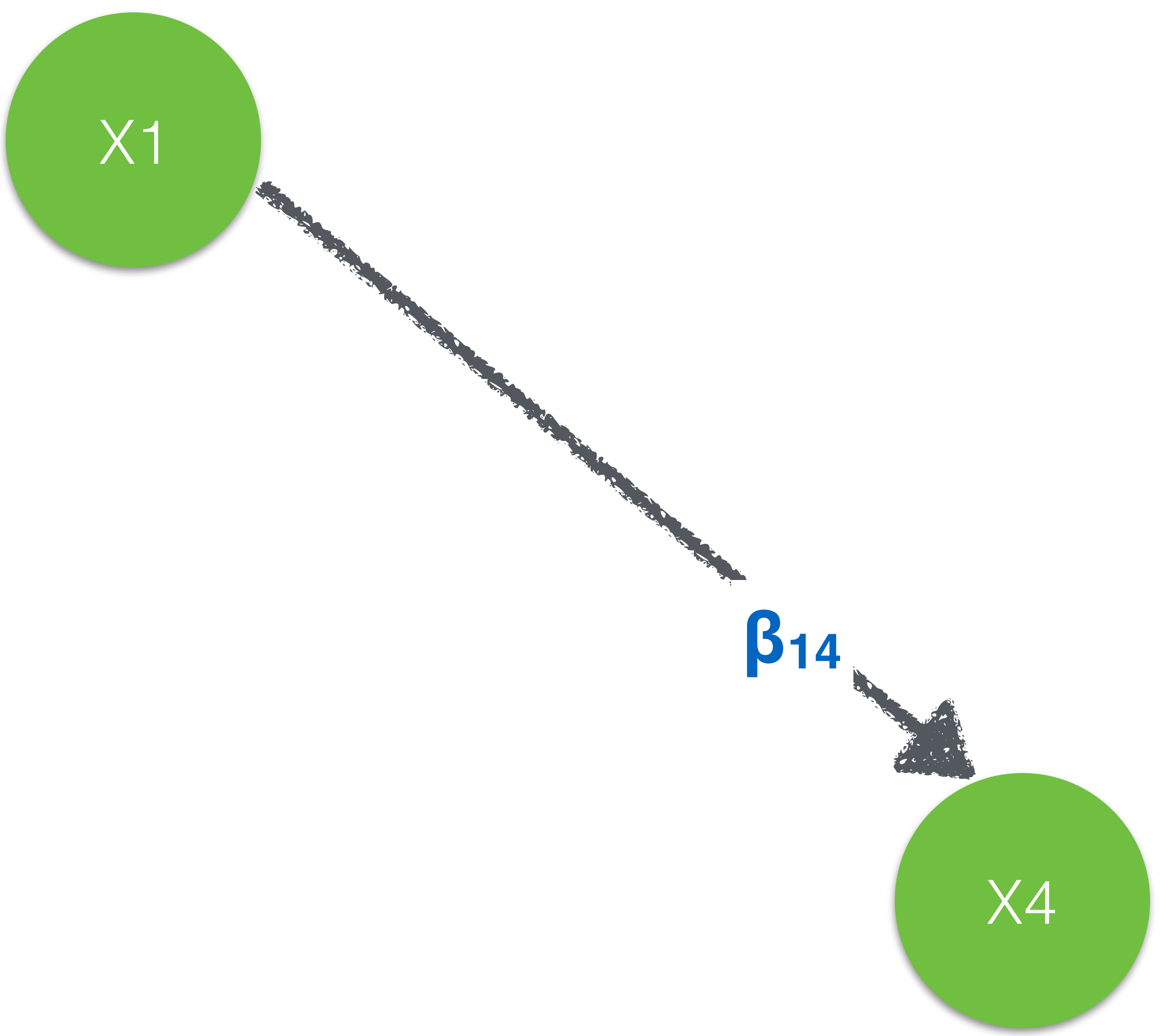
- $x = (x1, x2, \ldots, xn)$
- x_j = 0 or 1
- τ_j: node parameter (threshold)
- β_{jk}: pairwise interaction parameter

Conditional independence



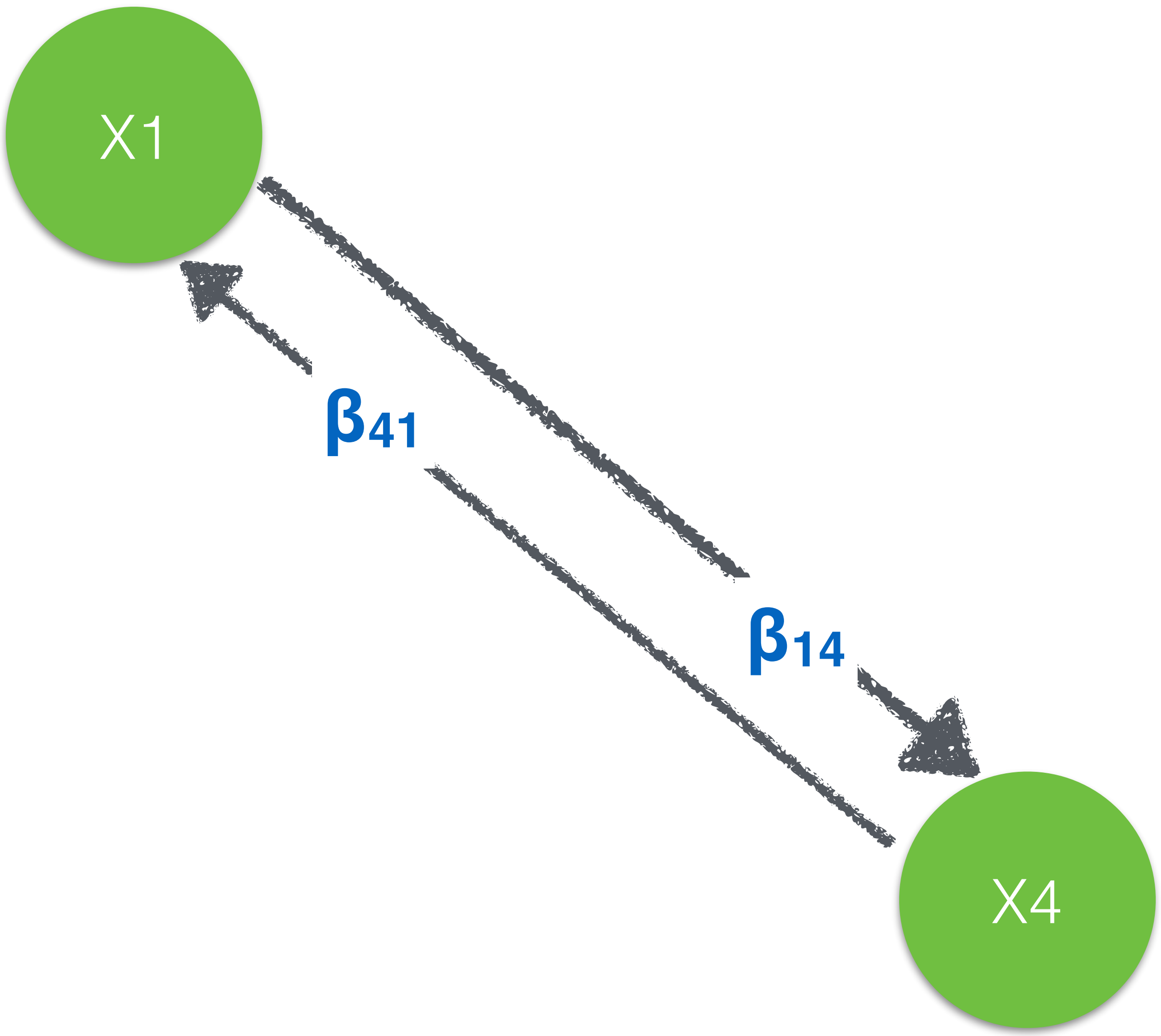
Conditional independence

- Regress X4 on X1: you get β_{14}



Conditional independence

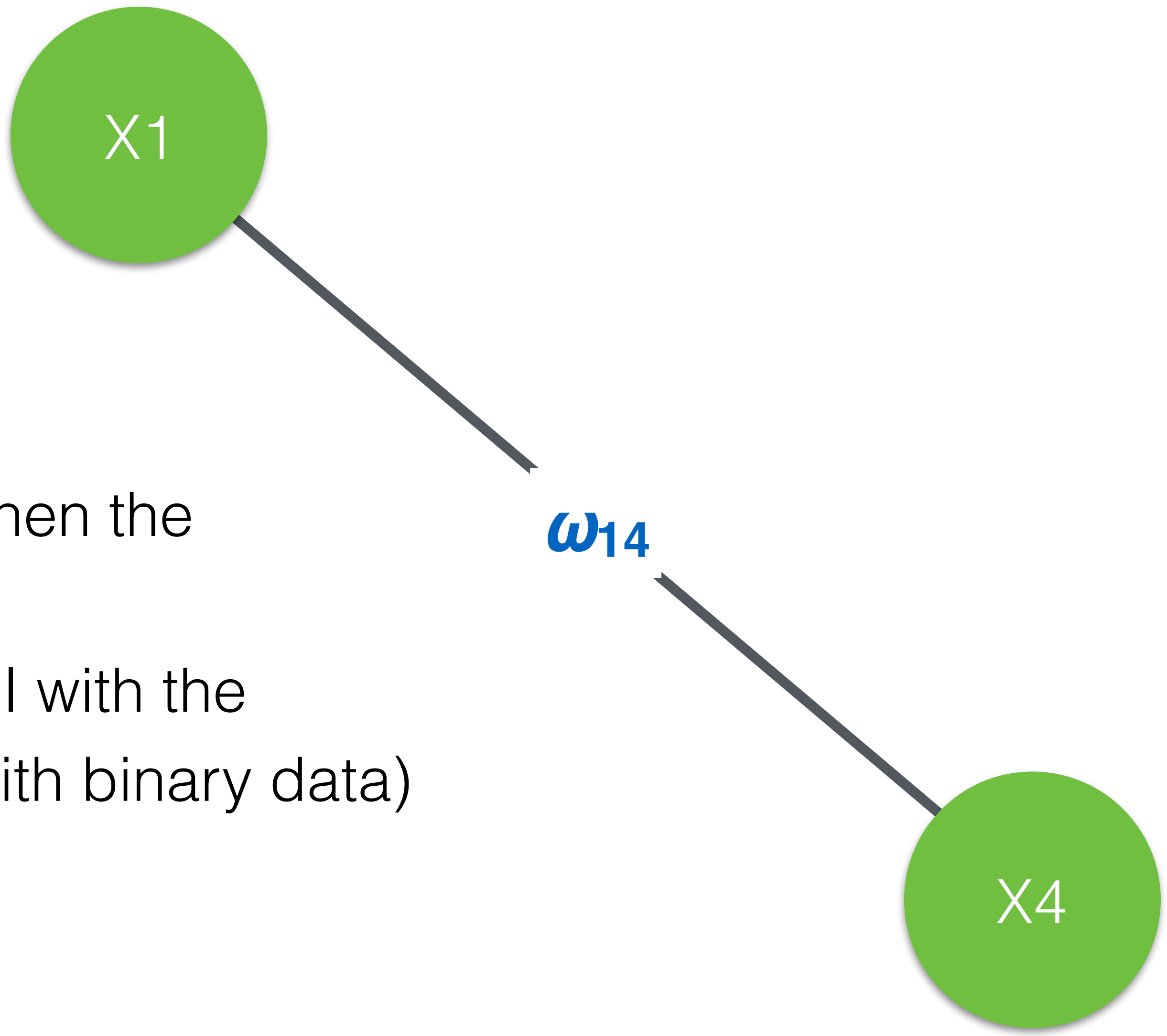
- Regress X4 on X1: you get β_{14}
- Regress X1 on X4: you get β_{41}



Conditional independence

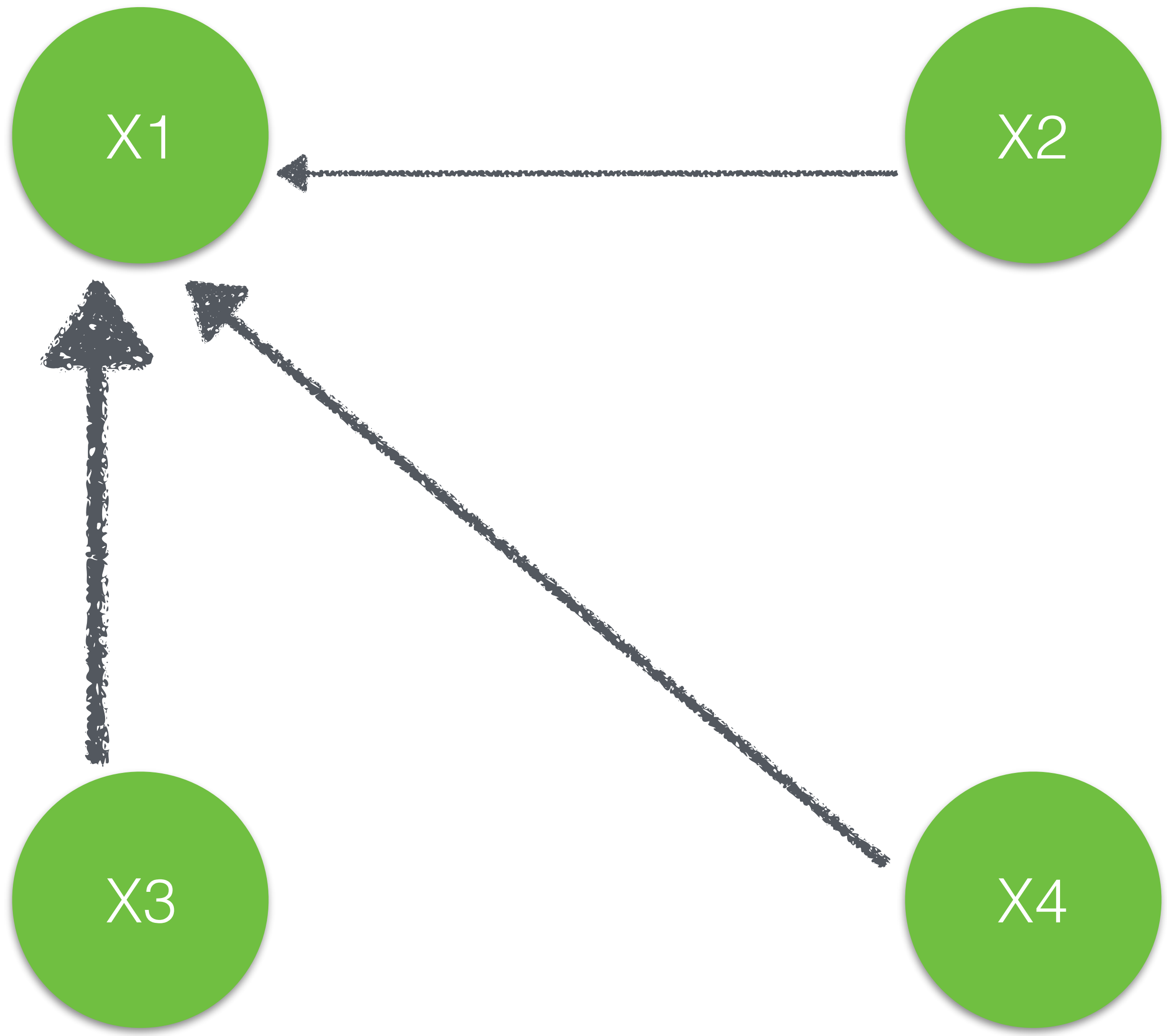
- Regress X4 on X1: you get β_{14}
- Regress X1 on X4: you get β_{41}
- Average the coefficients

  $\omega_{ij} = (\beta_{ij} + \beta_{ji})/2$



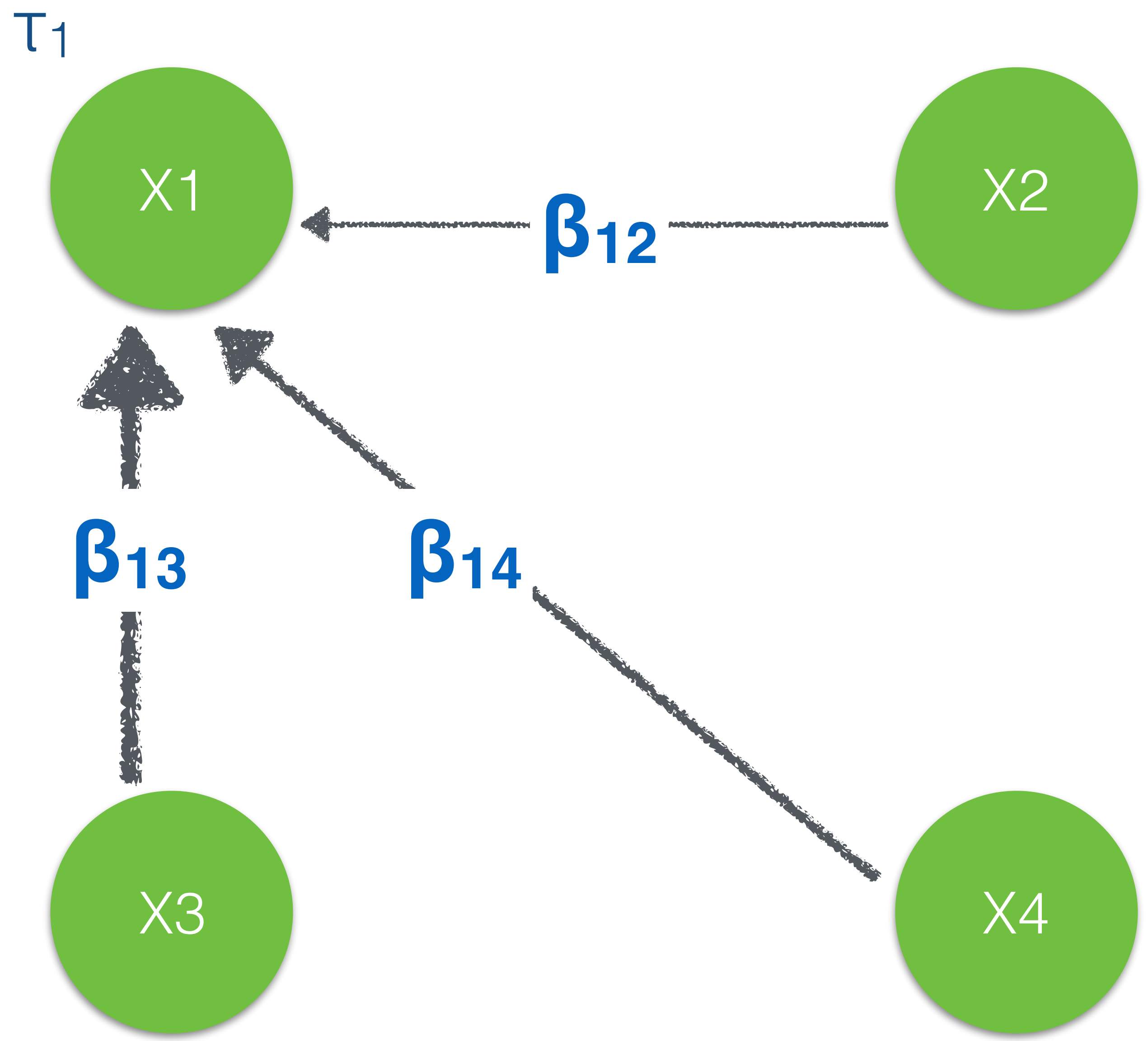
- This works for binary data (Ising model)
- For Gaussian Graphical model it works when the coefficients are scaled
- But: for GGM you can directly establish CI with the inverse covariance matrix (not possible with binary data)

With more variables



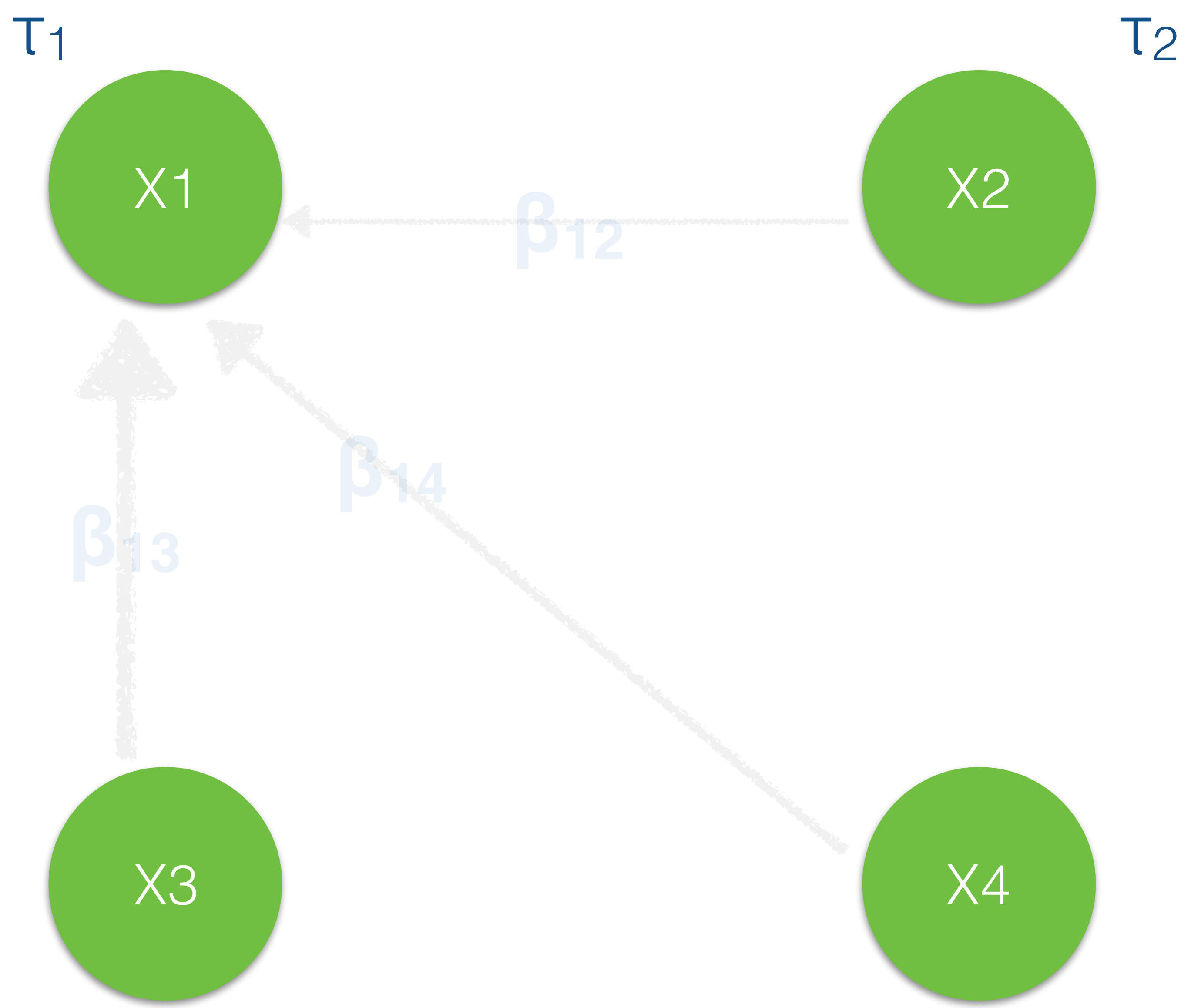
Perform regression of X_1 on all other variables

With more variables



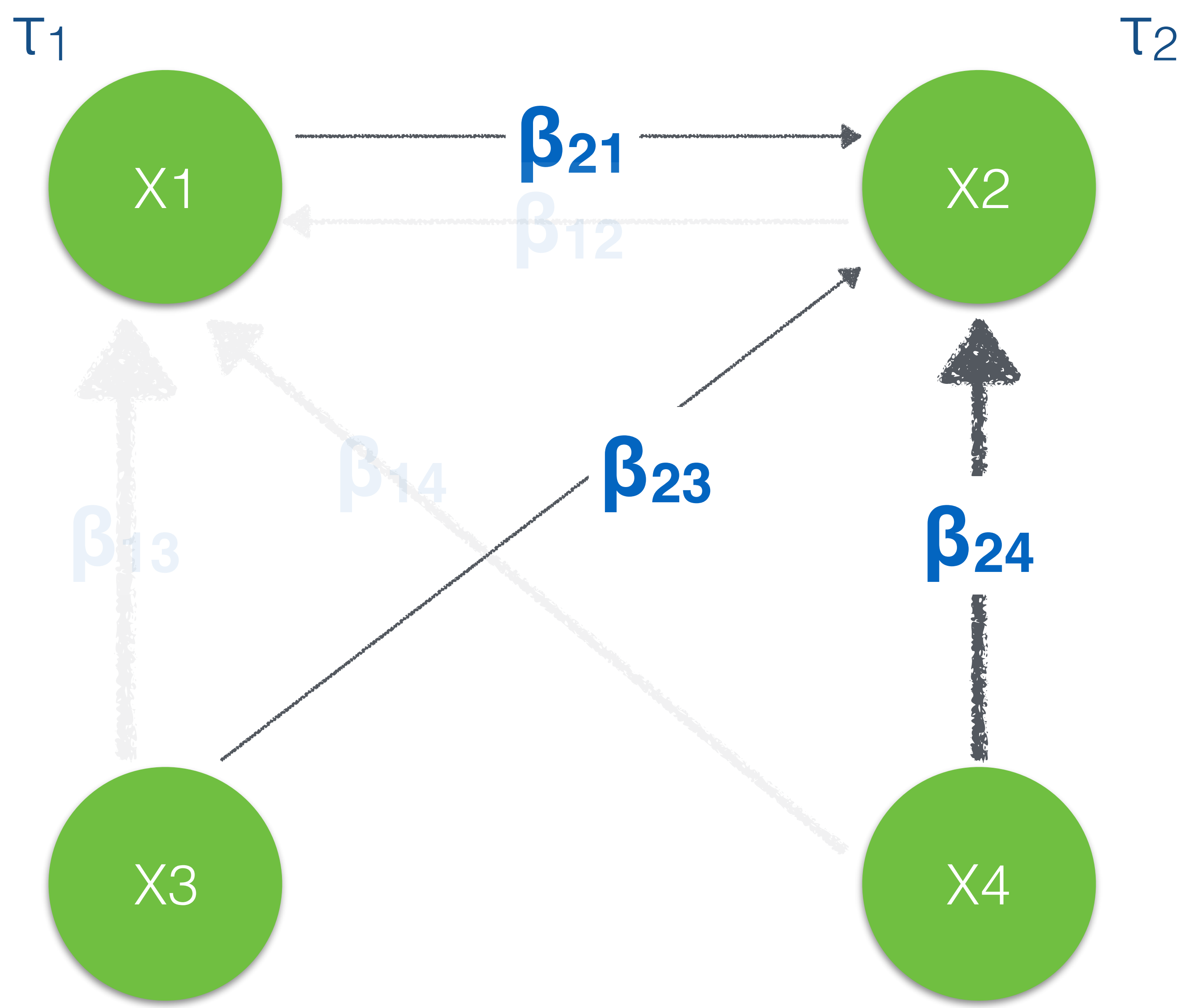
Perform regression of X_1 on all other variables

Basic idea



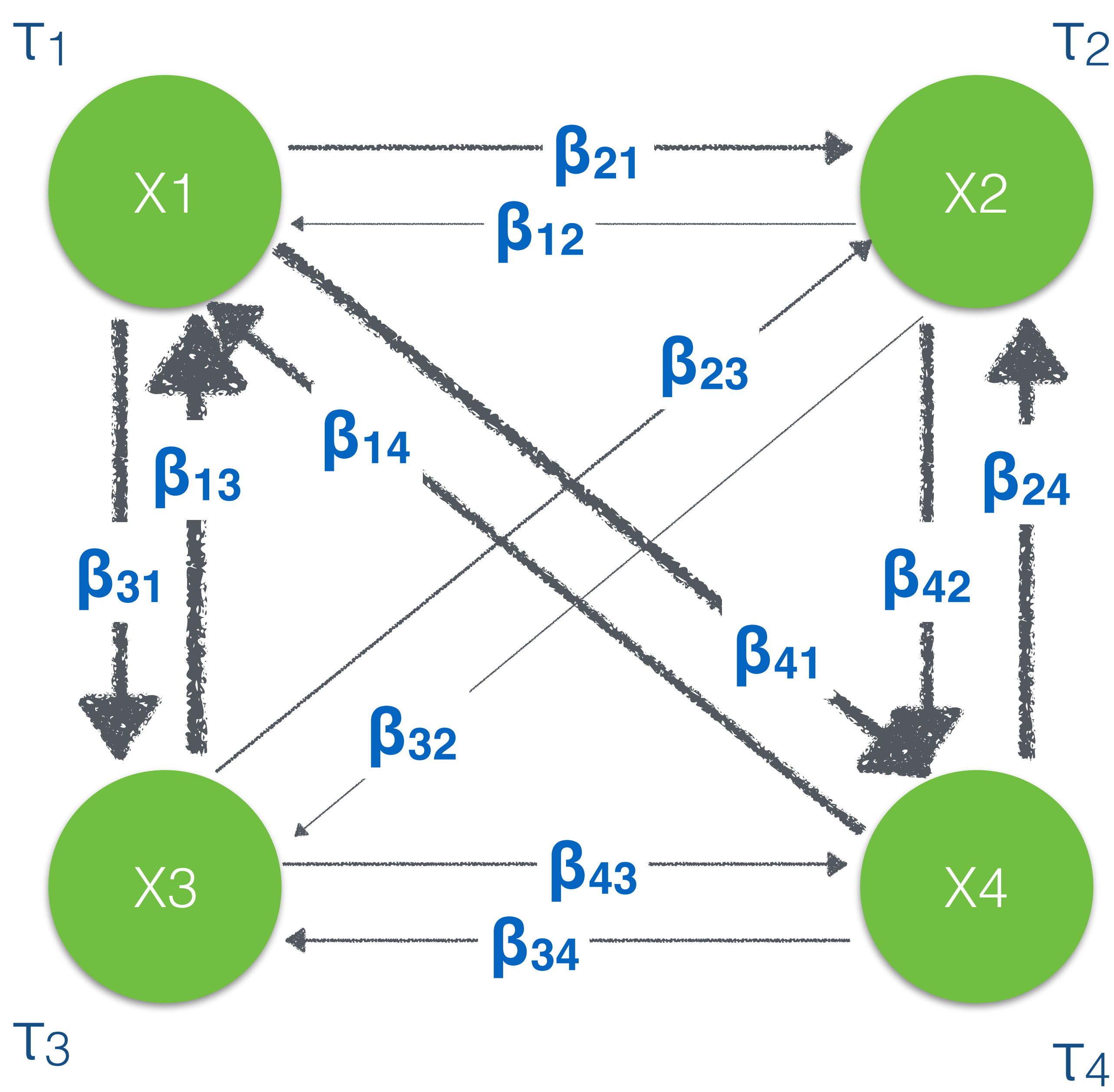
Repeat this for every variable

Basic idea



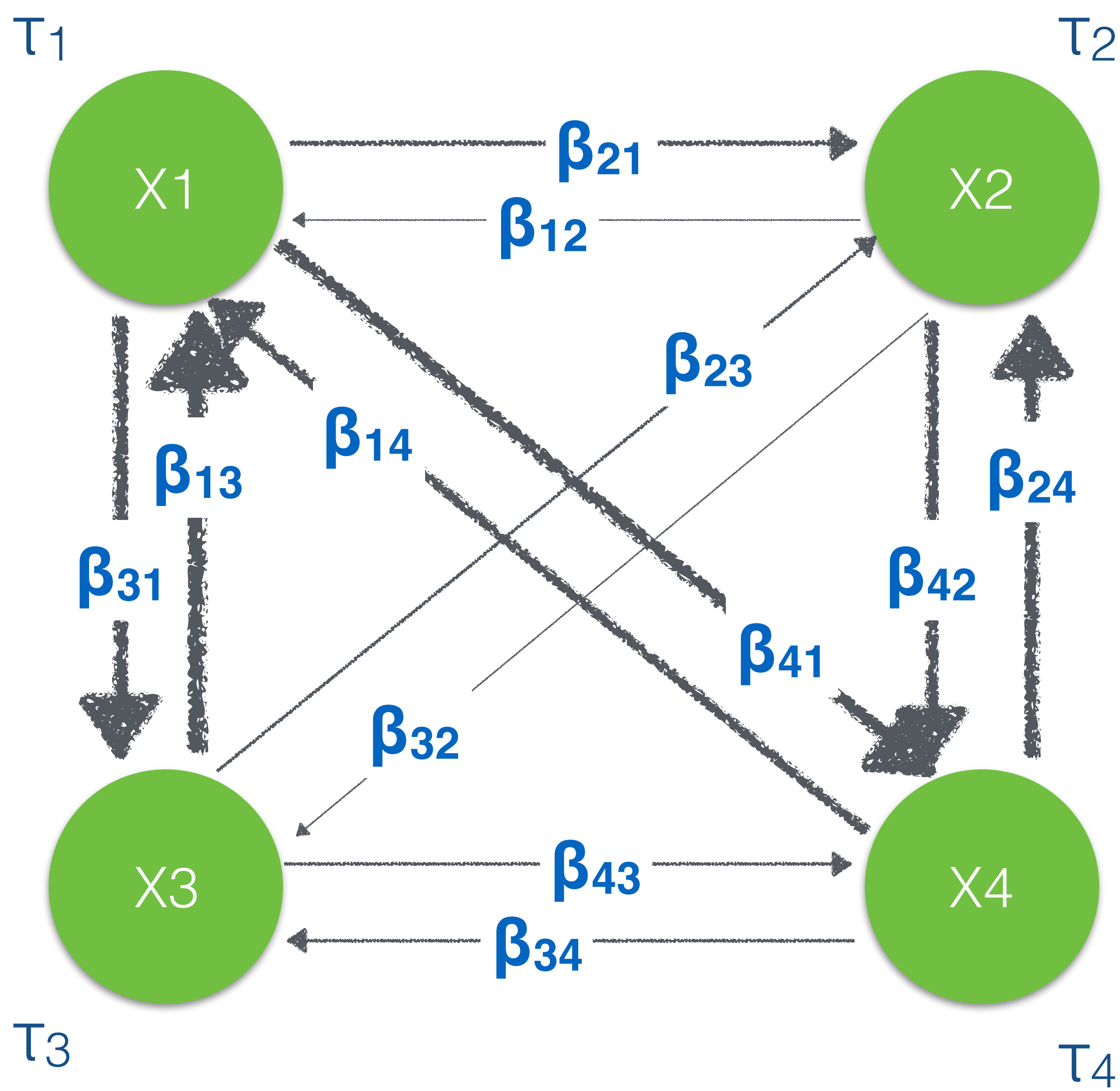
Repeat this for every variable

Basic idea



τ1
τ2
X1
X2
β21
β12
β23
β14
β13
β24
β31
β42
β41
β32
X3
X4
β43
β34
τ3
τ4

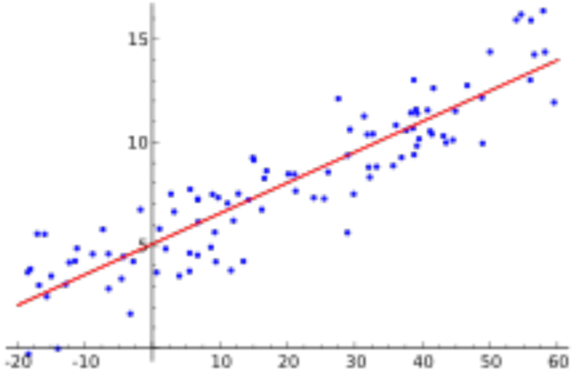
Basic idea



Control model complexity and prevent overfitting:
L_1-regularized logistic regression

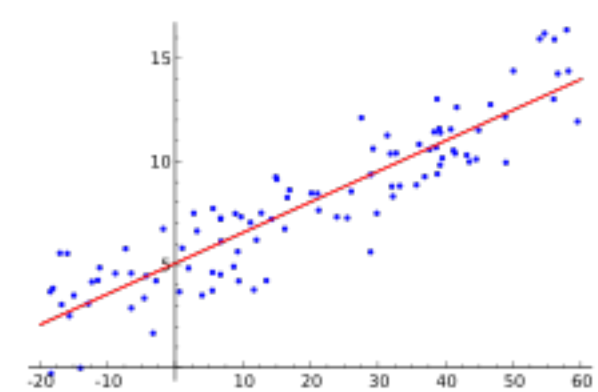
L_1 regularization

- aka Lasso
- Least Absolute Shrinkage and Selection Operator
- Involves subset selection (sparsity)
- Normal regression involves optimizing a function to find the solution that minimizes the sum of squared residuals



L_1 regularization

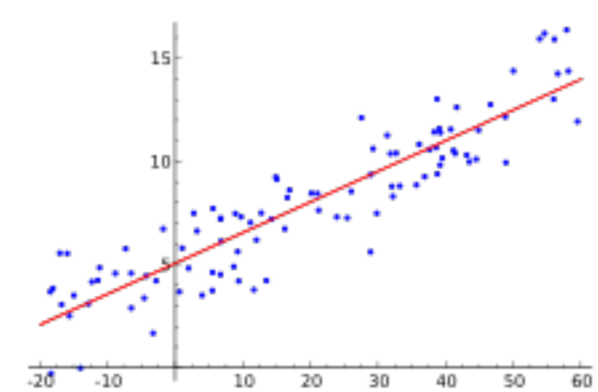
- aka Lasso
- Least Absolute Shrinkage and Selection Operator
- Involves subset selection (sparsity)
- Normal regression involves optimizing a function to find the solution that minimizes the sum of squared residuals
- With L_1 regularization the function to optimize is extended with an extra term:



L_1 regularization

- aka Lasso
- Least Absolute Shrinkage and Selection Operator
- Involves subset selection (sparsity)
- Normal regression involves optimizing a function to find the solution that minimizes the sum of squared residuals
- With L_1 regularization the function to optimize is extended with an extra term:

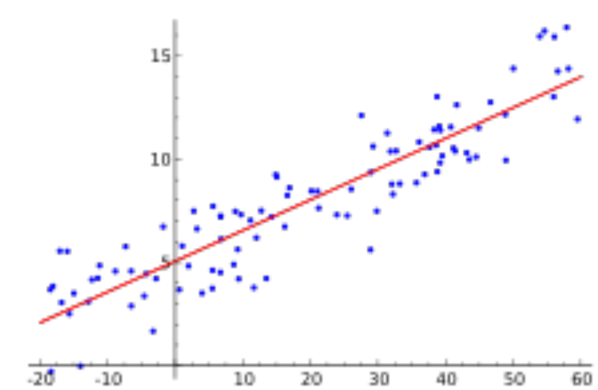
$$\min_{\beta} \frac{1}{2}\|y - X\beta\|^2$$



L_1 regularization

- aka Lasso
- Least Absolute Shrinkage and Selection Operator
- Involves subset selection (sparsity)
- Normal regression involves optimizing a function to find the solution that minimizes the sum of squared residuals
- With L_1 regularization the function to optimize is extended with an extra term:

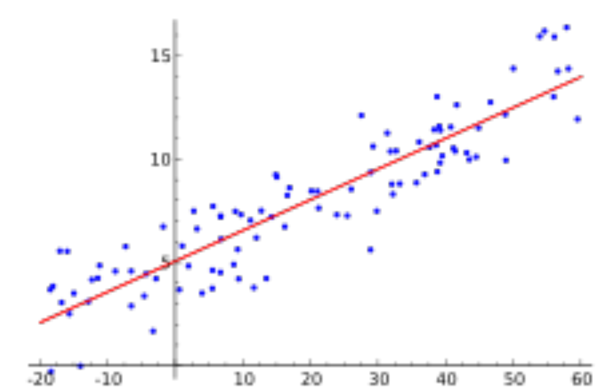
$$\min_{\beta} \frac{1}{2}\|y - X\beta\|^2 + \lambda \sum_{i=1}^{p} |\beta_i|$$



L_1 regularization

- aka Lasso
- Least Absolute Shrinkage and Selection Operator
- Involves subset selection (sparsity)
- Normal regression involves optimizing a function to find the solution that minimizes the sum of squared residuals
- With L_1 regularization the function to optimize is extended with an extra term:

$$\min_{\beta} \frac{1}{2}\|y - X\beta\|^2 \qquad \text{s.t.} \sum_{i=1}^{p} |\beta_i| \leq t$$



L_1-regularization

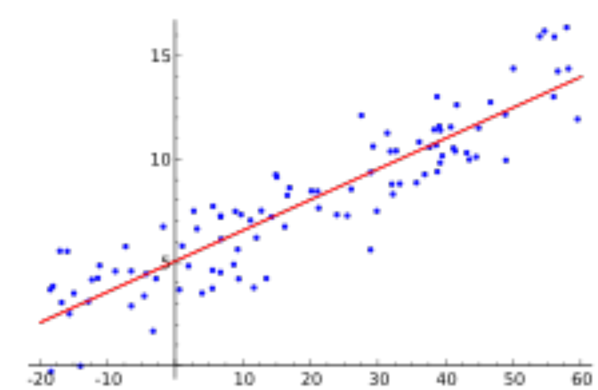
Property of L1- regularization:

- ensures that some coefficients are set to zero (exactly)
- shrinks other coefficients

Convenient property :

- use this for problem with small conditional dependencies
- instead of ignoring multiple testing problem and Bonferroni corrections

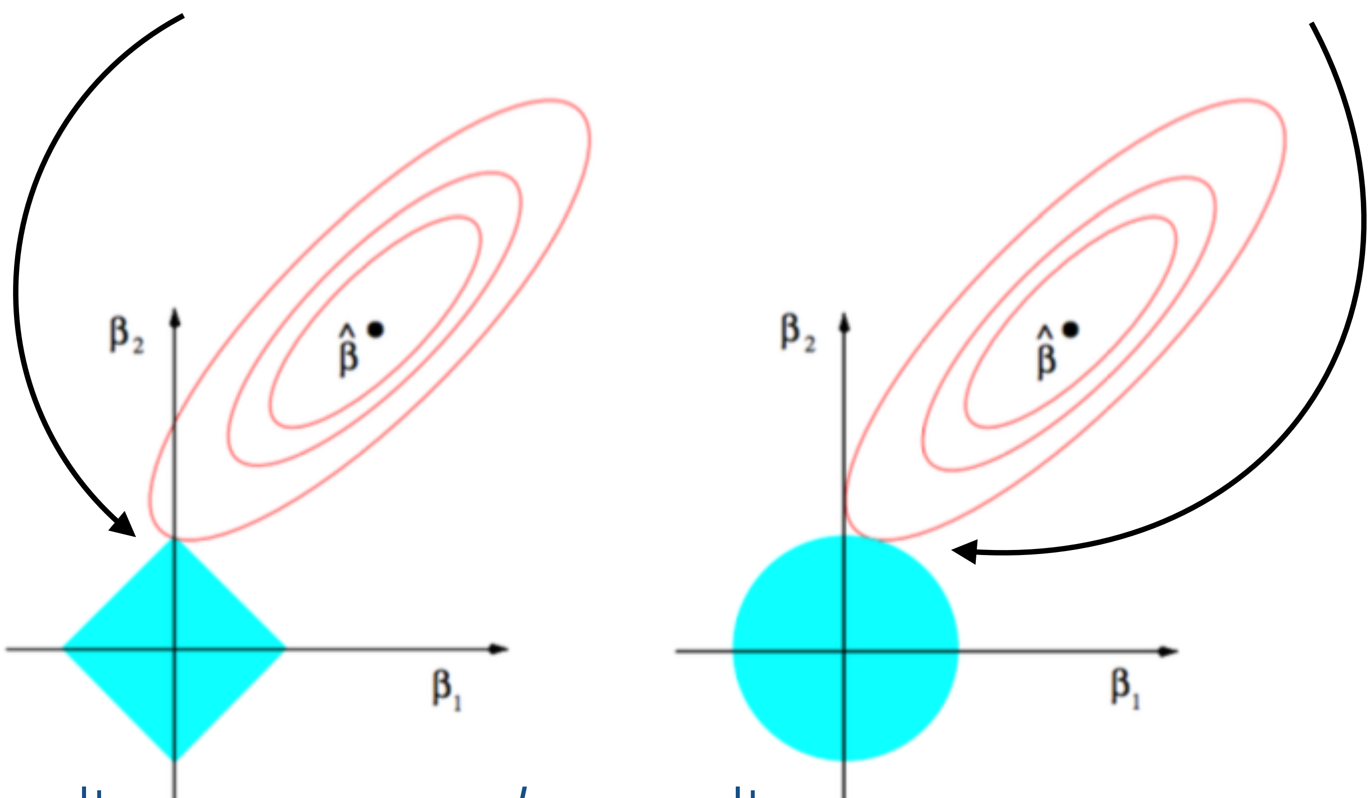
$$\min_{\beta} \frac{1}{2}\|y - X\beta\|^2 \qquad \text{s.t.} \sum_{i=1}^{p} |\beta_i| \leq t$$



L_1-regularization

$$\min_{\beta} \frac{1}{2}\|y - X\beta\|^2 \qquad \text{s.t.} \sum_{i=1}^{p} |\beta_i| \le t$$

The solution that satisfies the constraint



L_1 penalty

L_2 penalty

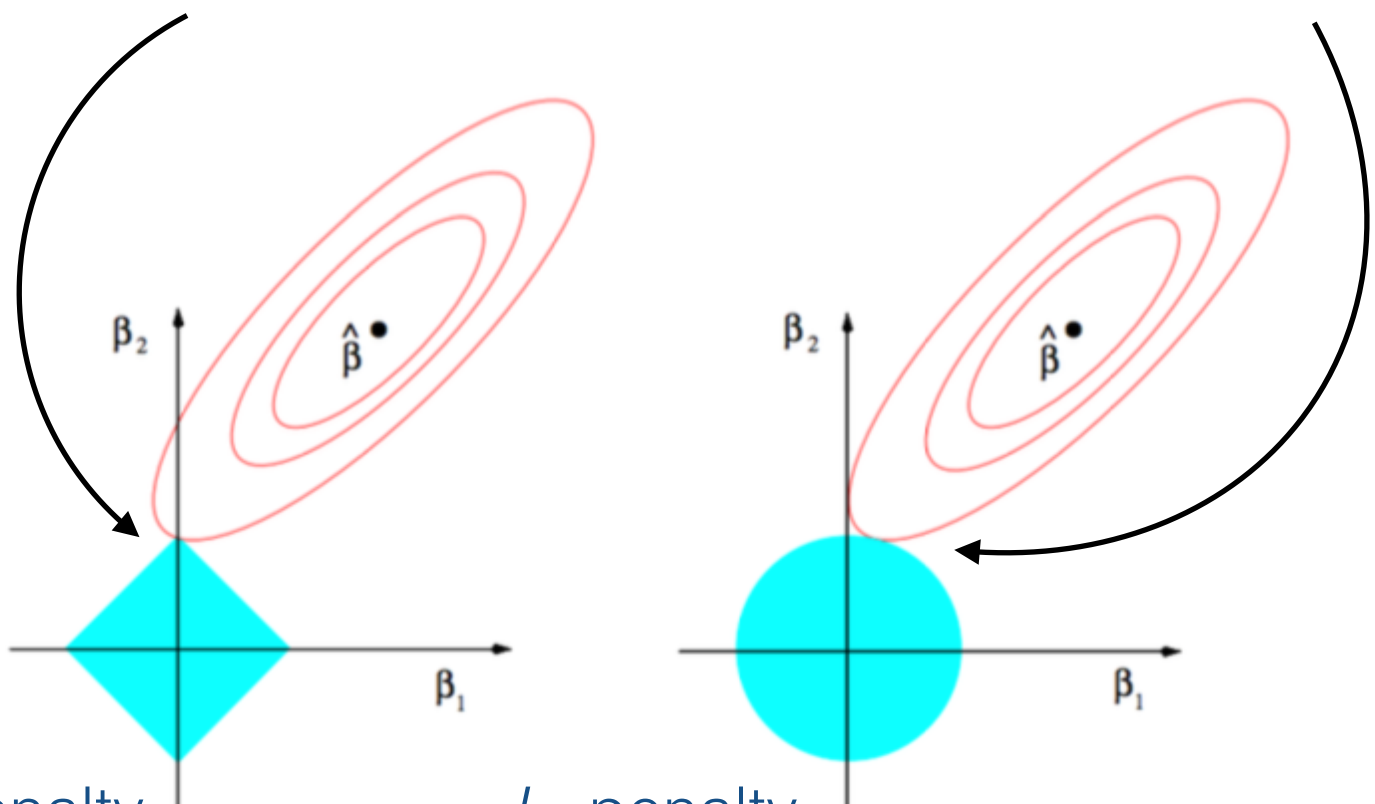
L1- regularization:

ensures that some coefficients are set to zero (exactly) and shrinks other coefficients

L_1-regularization

$$\min_{\beta} \frac{1}{2}\|y - X\beta\|^2 \qquad \text{s.t.} \sum_{i=1}^{p} |\beta_i| \leq t$$

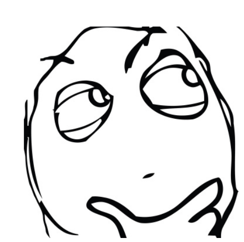
The solution that satisfies the constraint



L_1 penalty L_2 penalty

L1- regularization:

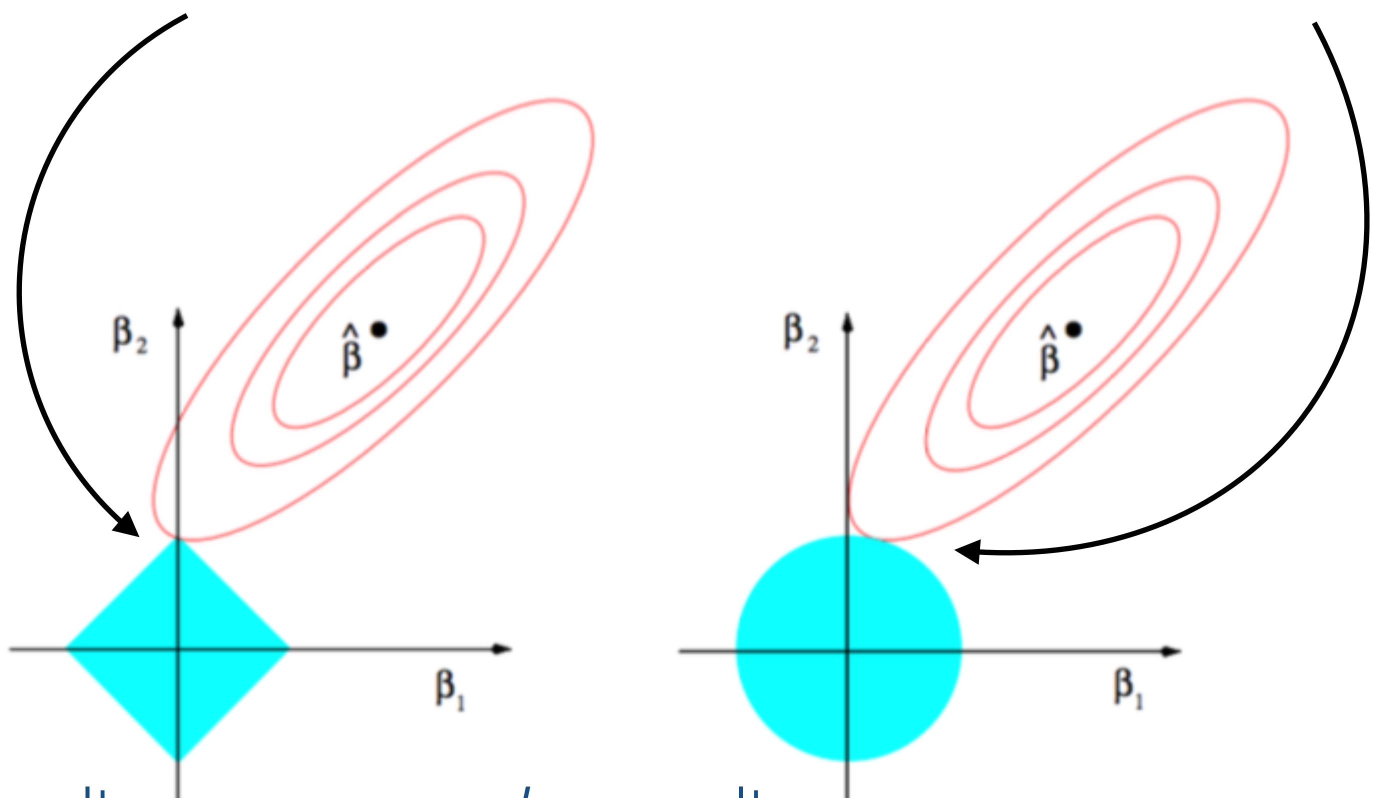
ensures that some coefficients are set to zero (exactly) and shrinks other coefficients



L_1-regularization

$$\min_{\beta} \frac{1}{2}\|y - X\beta\|^2 \qquad \text{s.t.} \sum_{i=1}^{p} |\beta_i| \leq t$$

The solution that satisfies the constraint



L_1 penalty L_2 penalty

L1- regularization:

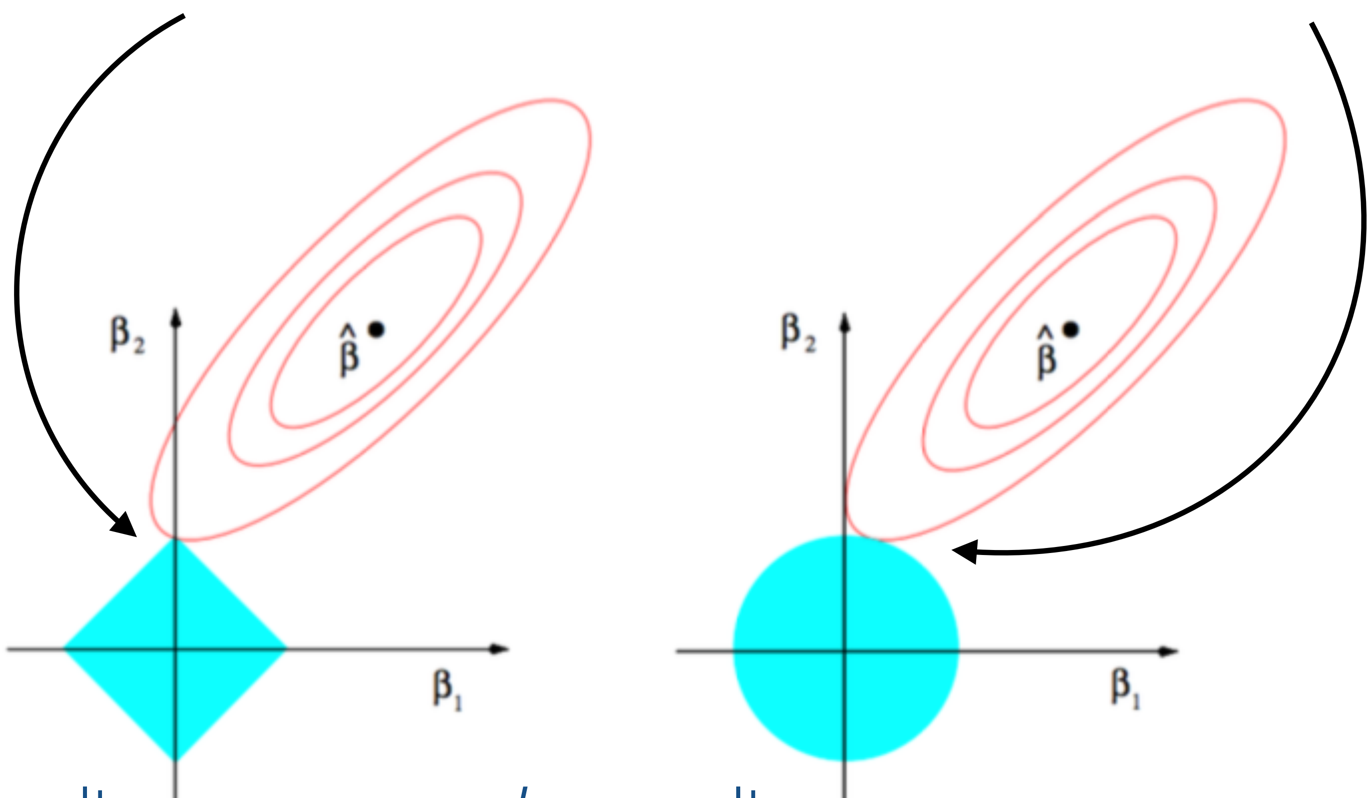
ensures that some coefficients are set to zero (exactly) and shrinks other coefficients

L_1-regularization

$$\min_{\beta} \frac{1}{2}\|y - X\beta\|^2$$

$$\text{s.t.} \sum_{i=1}^{p} |\beta_i| \leq t$$

The solution that satisfies the constraint

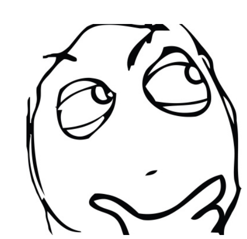


L_1 penalty

L_2 penalty

L1- regularization:

ensures that some coefficients are set to zero (exactly) and shrinks other coefficients

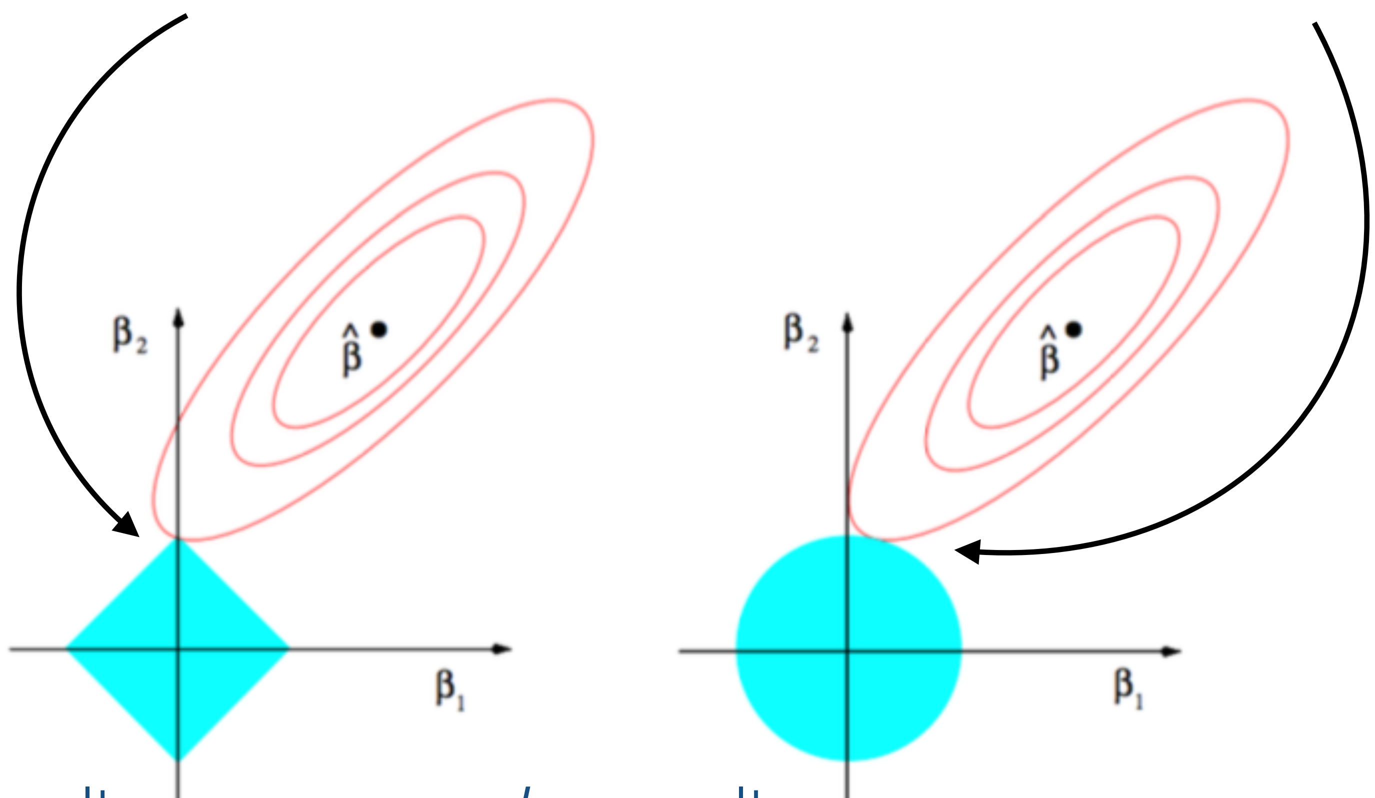


L_1-regularization

$$\min_{\beta} \frac{1}{2}\|y - X\beta\|^2$$

$$\text{s.t.} \sum_{i=1}^{p} |\beta_i| \leq t$$

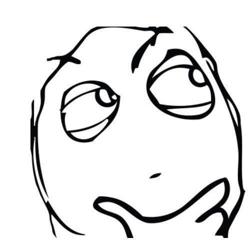
The solution that satisfies the constraint



L_1 penalty L_2 penalty

L1- regularization:

ensures that some coefficients are set to zero (exactly) and shrinks other coefficients

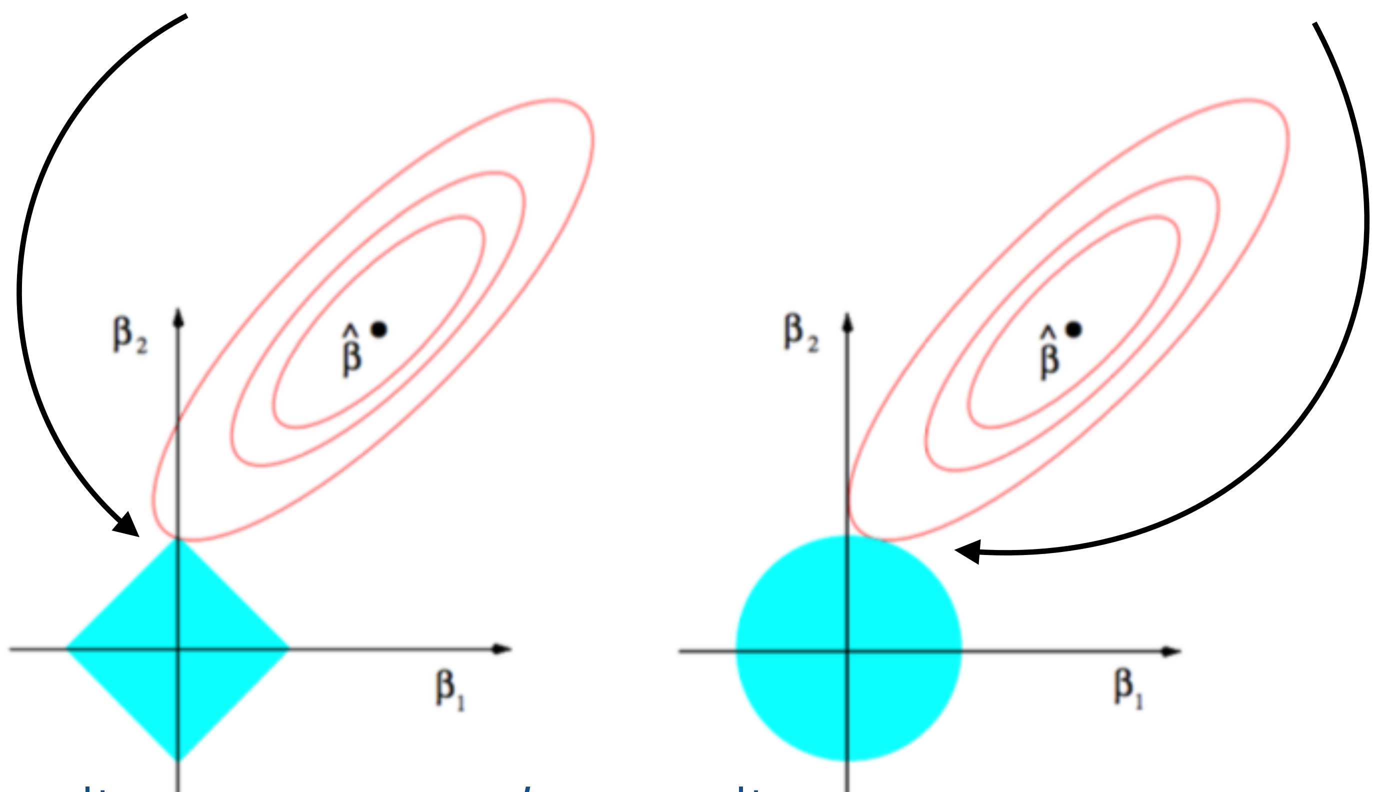


L_1-regularization

$$\min_{\beta} \frac{1}{2}\|y - X\beta\|^2$$

$$\text{s.t.} \sum_{i=1}^{p} |\beta_i| \leq t$$

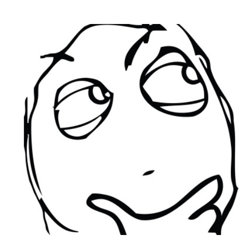
The solution that satisfies the constraint



L_1 penalty L_2 penalty

L1- regularization:

ensures that some coefficients are set to zero (exactly) and shrinks other coefficients



L_1-regularization

$$\min_{\beta} \frac{1}{2}\|y - X\beta\|^2 + \lambda \sum_{i=1}^{p} |\beta_i|$$

tuning parameter

L_1-regularization

$$\hat{\Theta}_j^{\rho} = \arg\min_{\Theta_j} \left\{ -x_{ij} \cdot \left(\tau_j + \sum_{k \in V_{\setminus j}} \beta_{jk} x_{ik} \right) + \log\left(1 + \exp\left\{ \tau_j + \sum_{k \in V_{\setminus j}} x_{ik} \beta_{jk} \right\} \right) + \rho \sum_{k \in V_{\setminus j}} \left| \beta_{jk} \right| \right\}$$

tuning parameter

- i: independent observations {1, 2, …, n}
- $\hat{\Theta}_j^{\rho}$ matrix with βjk and τj
- ρ: tuning parameter
- `R` package `glmnet`: 100 values of ρ

L_1-regularization

$$\hat{\Theta}_j^{\rho} = \arg\min_{\Theta_j}\left\{-x_{ij}\cdot\left(\tau_j+\sum_{k\in V_{\setminus j}}\beta_{jk}x_{ik}\right) + \log\left(1+\exp\left\{\tau_j+\sum_{k\in V_{\setminus j}} x_{ik}\beta_{jk}\right\}\right)+\rho\sum_{k\in V_{\setminus j}}\left|\beta_{jk}\right|\right\}$$

tuning parameter

- This function is known to be convex (it has a minimum)
- For a specific value of λ, you can find the βs by minimizing this function

- i: independent observations {1, 2, …, n}
- $\hat{\Theta}_j^{\rho}$ matrix with β_{jk} and τ_j
- ρ: tuning parameter
- R package **`glmnet`**: 100 values of ρ

L_1-regularization

$$\hat{\Theta}_j^{\rho} = \arg\min_{\Theta_j} \left\{ -x_{ij} \cdot \left(\tau_j + \sum_{k \in V_{\setminus j}} \beta_{jk} x_{ik} \right) + \log\left(1 + \exp\left\{ \tau_j + \sum_{k \in V_{\setminus j}} x_{ik} \beta_{jk} \right\} \right) + \rho \sum_{k \in V_{\setminus j}} \left| \beta_{jk} \right| \right\}$$

tuning parameter

- For different values of λ you get different sets of βs
- Now we can select the best fitting set of βs
- With a goodness-of-fit measure
    - extended BIC (Bayesian Information Criterion)
    - extension: penalty on the number of variables AND on the number of edges
- But… EBIC involves choosing a *hyperparameter γ*!

L_1-regularization

- Extended Bayesian Information Criterion
- Based on negative (log) likelihood
- $$\hat{\Theta}_j^{\rho} = \arg\min_{\Theta_j} \left\{ -x_{ij} \cdot \left(\tau_j + \sum_{k \in V_{\setminus j}} \beta_{jk} x_{ik} \right) + \log\left(1 + \exp\left\{ \tau_j + \sum_{k \in V_{\setminus j}} x_{ik} \beta_{jk} \right\} \right) + \rho \sum_{k \in V_{\setminus j}} \left| \beta_{jk} \right| \right\}$$
- $EBIC_{\gamma}(J) = -2L + |J|\log(n) + 2\gamma|J|\log(p-1)$

J: number of neighbors

N: number of observations

p: number of variables

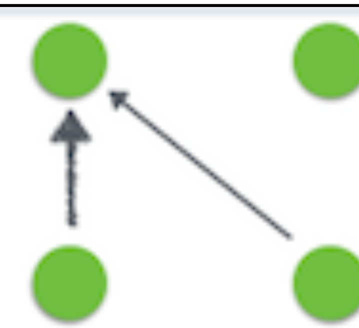
γ: hyperparameter

L_1 regularization

$\rho = .025$ β_{12} β_{13} β_{14}

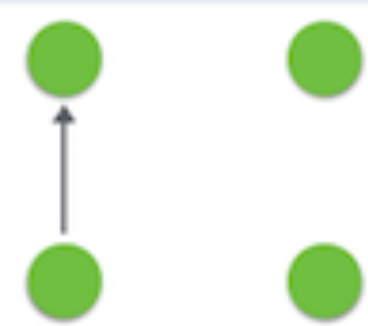
EBIC = -1274

$\rho = .185$ β_{13} β_{14}



EBIC = -1525

$\rho = .243$ β_{13}



EBIC = -1308

↑ tuning parameter

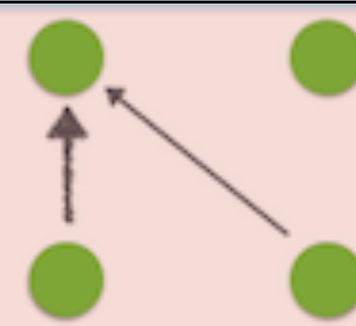
$\gamma = .25$

L_1 regularization

ρ = .025 β_{12} β_{13} β_{14}

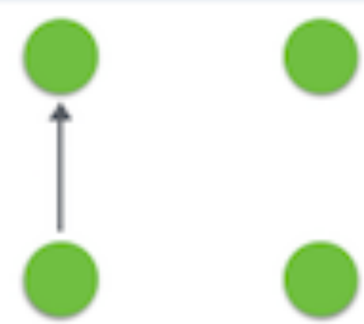
EBIC = -1274

ρ = .185 β_{13} β_{14}



EBIC = -1525

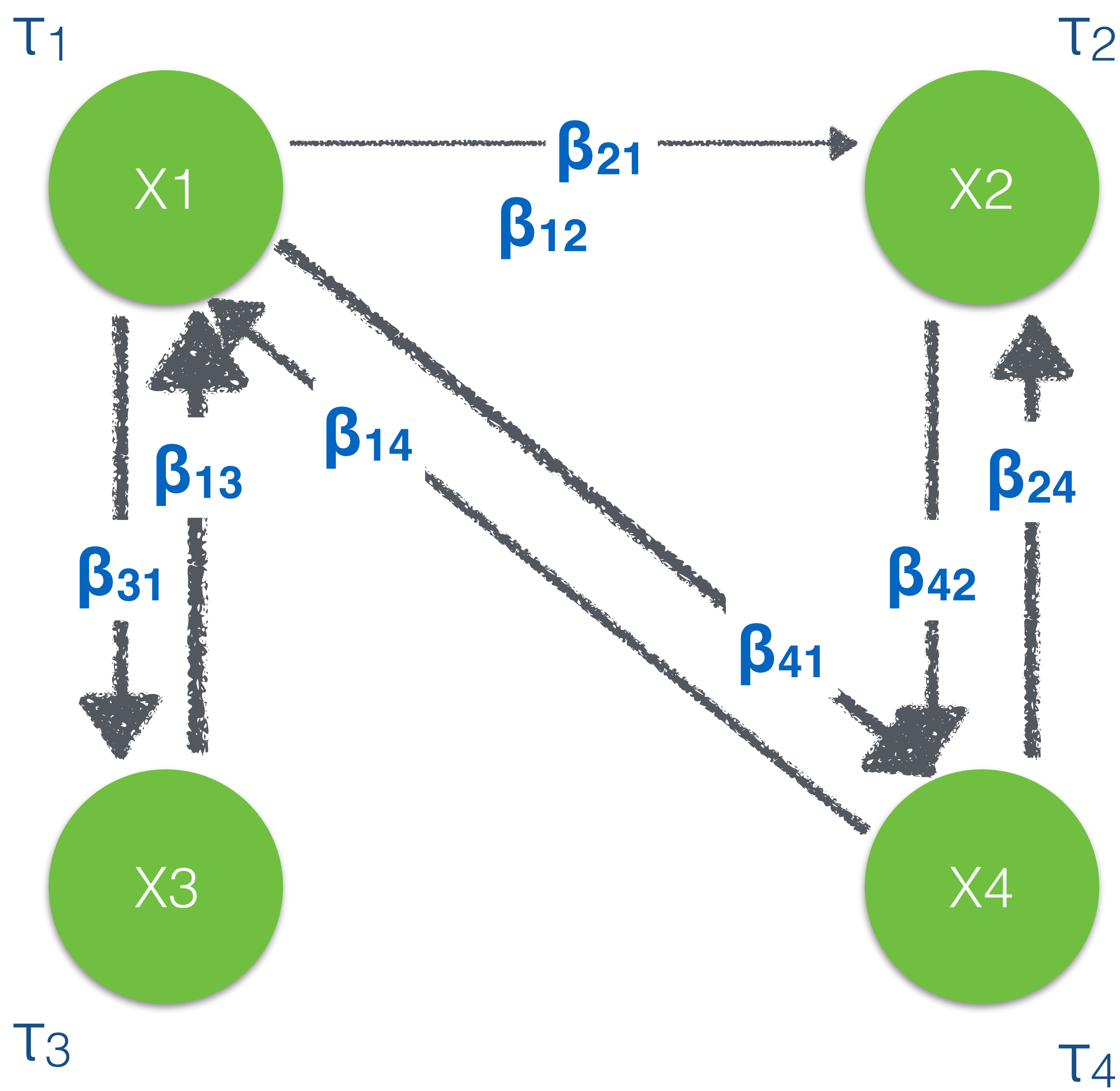
ρ = .243 β_{13}

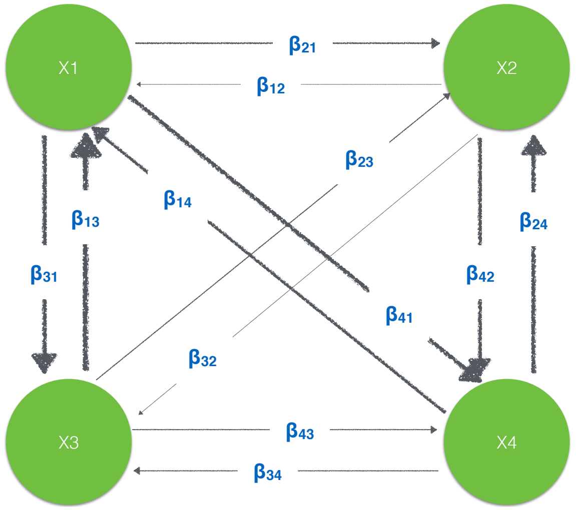


EBIC = -1308

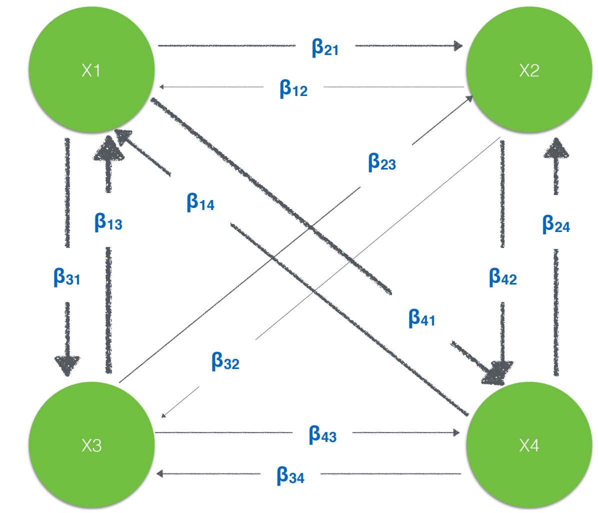
↑ tuning parameter

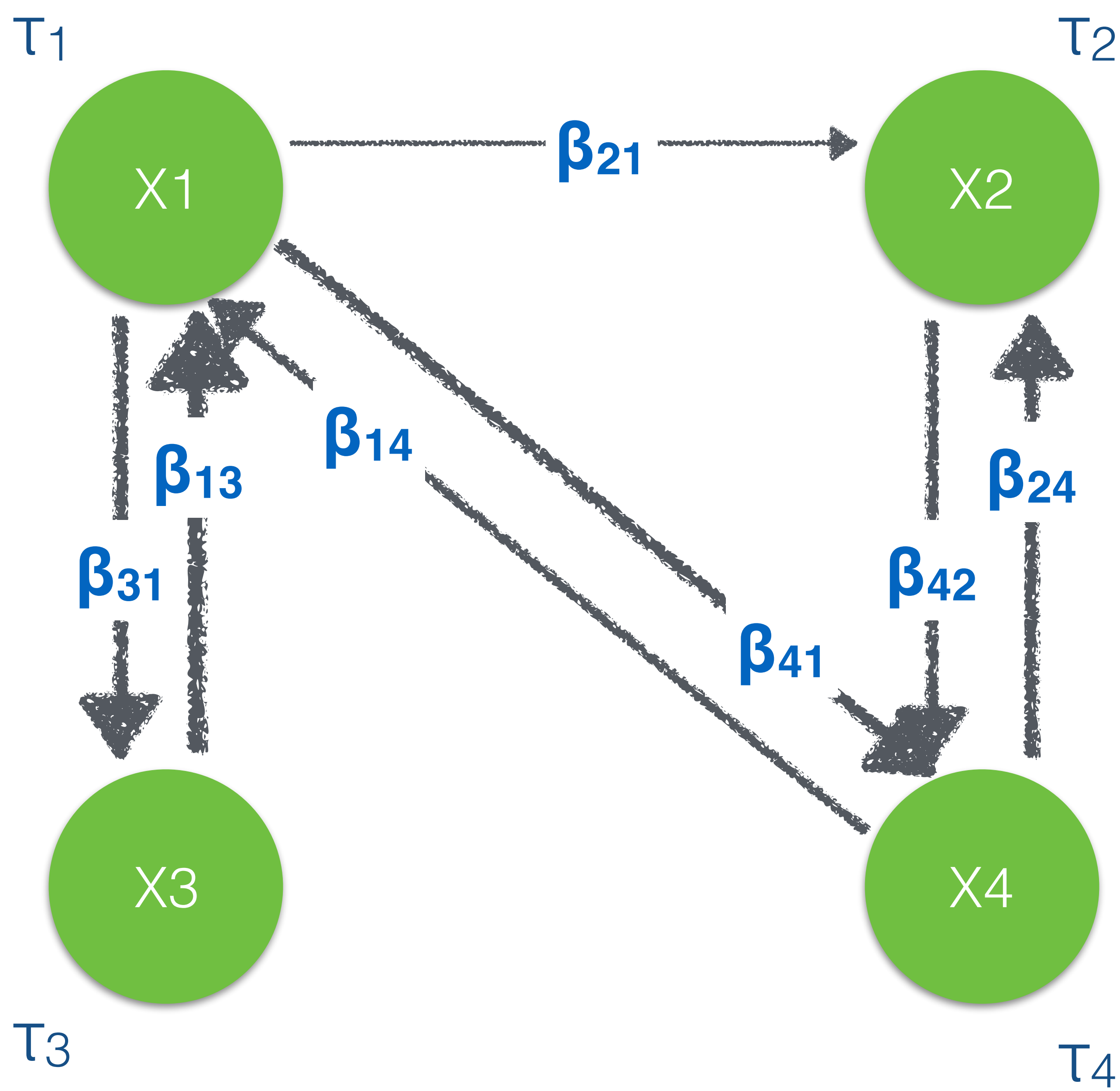
γ = .25



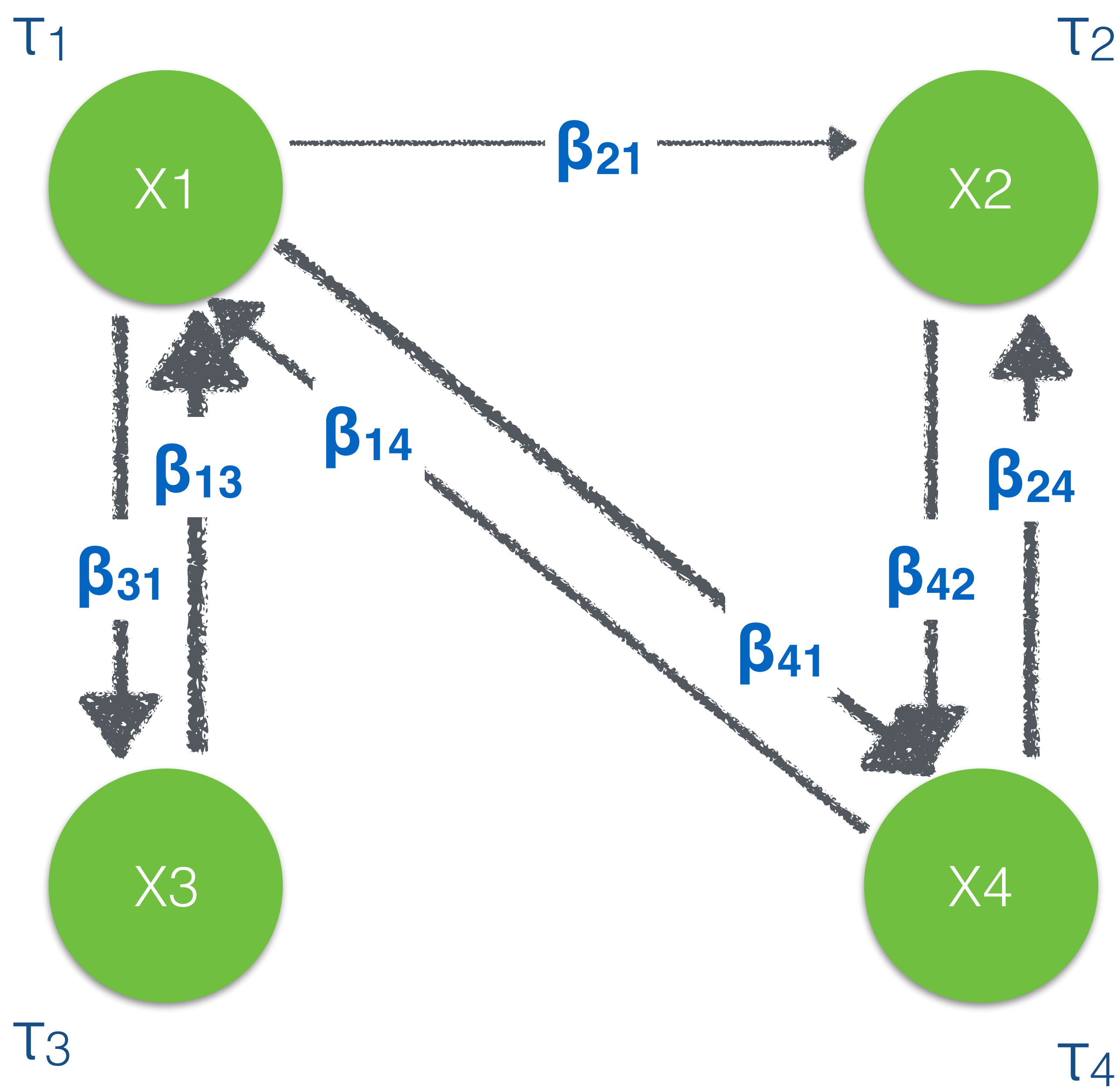


- collect regularized parameters
- but….

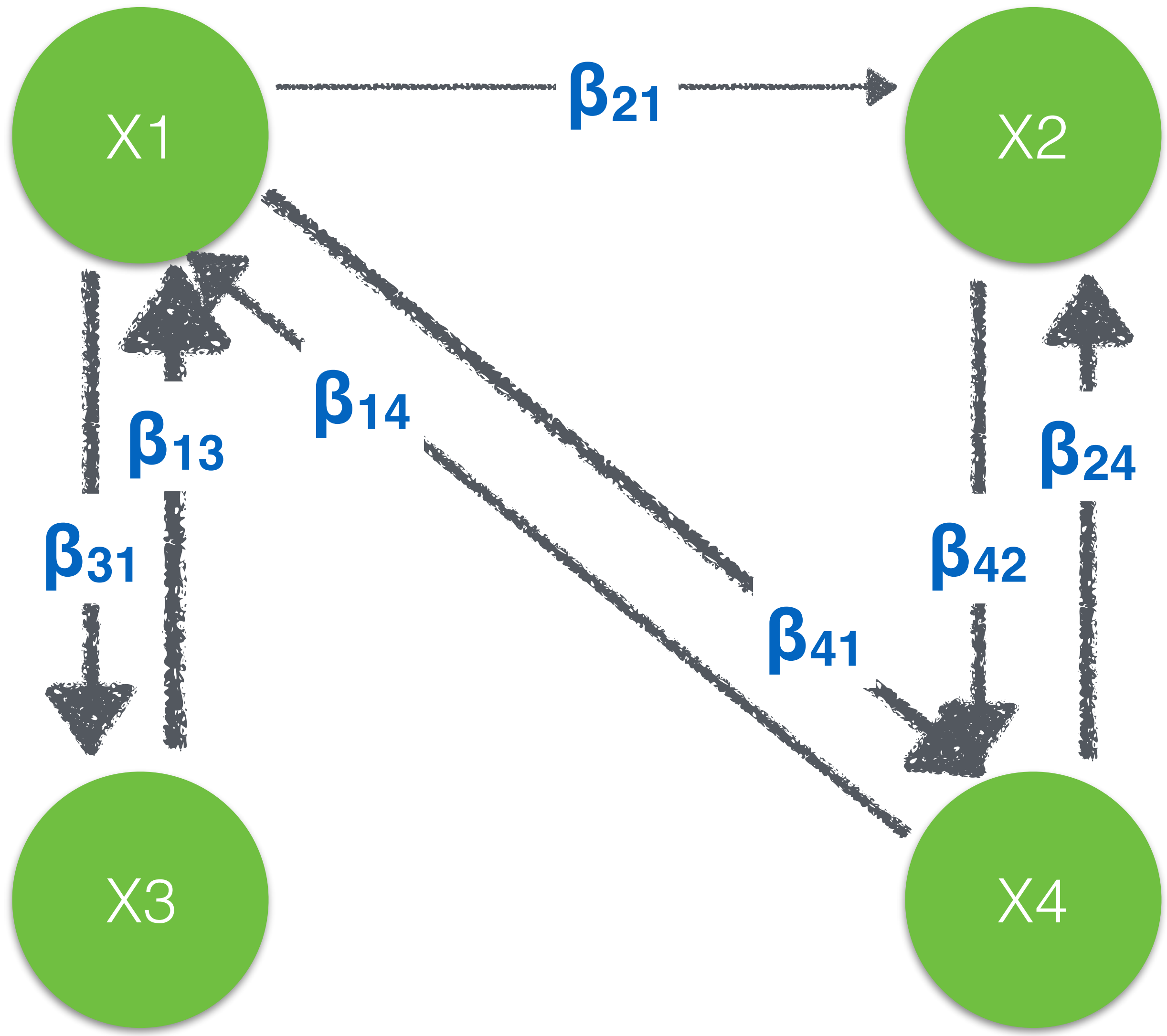




- collect regularized parameters
- but….



- collect regularized parameters
- but….

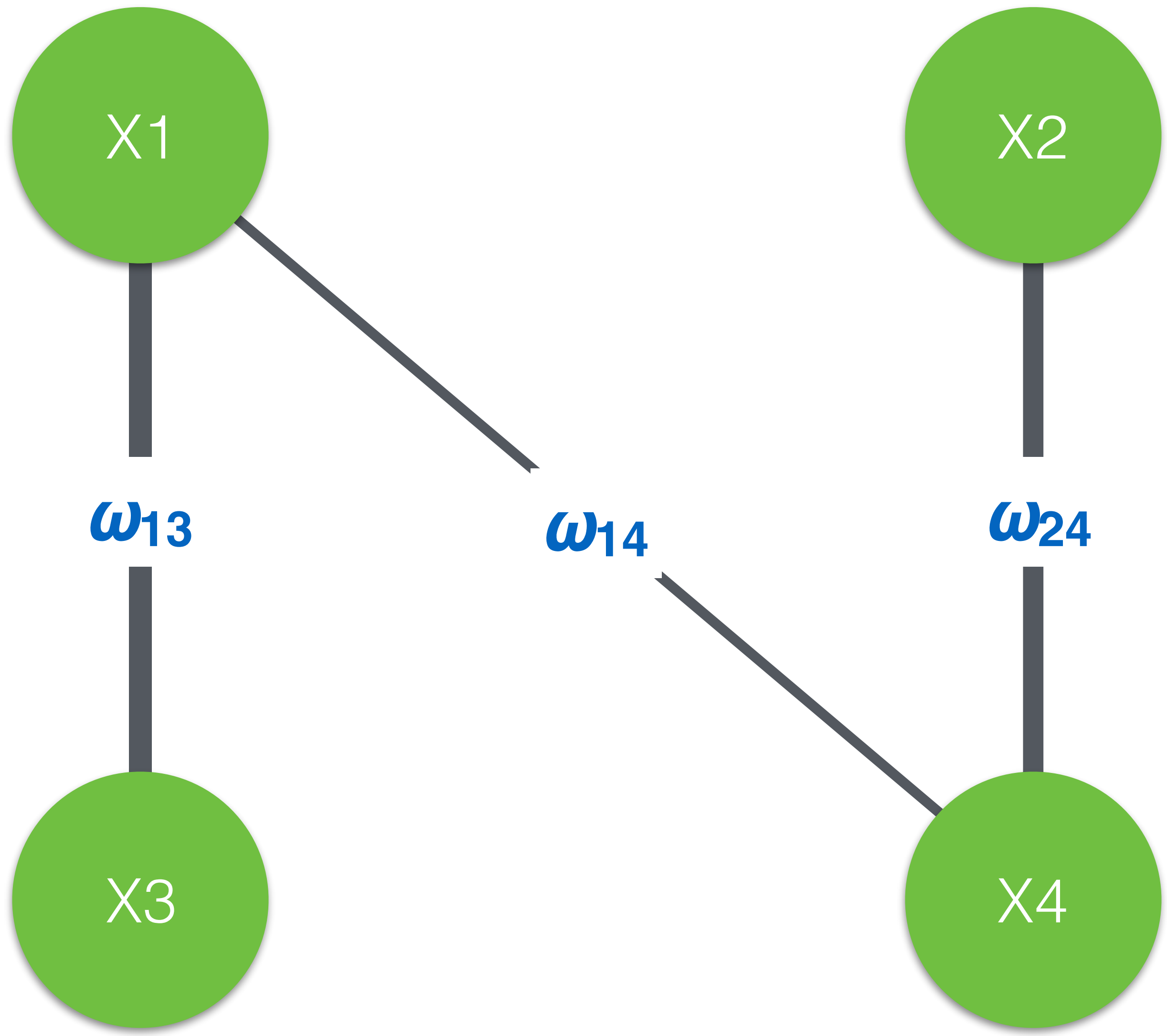


AND-rule:

if $\beta_{ij} \neq 0$ AND $\beta_{ji} \neq 0$

then $\omega_{ij} = (\beta_{ij} + \beta_{ji})/2$

else $\omega_{ij} = 0$

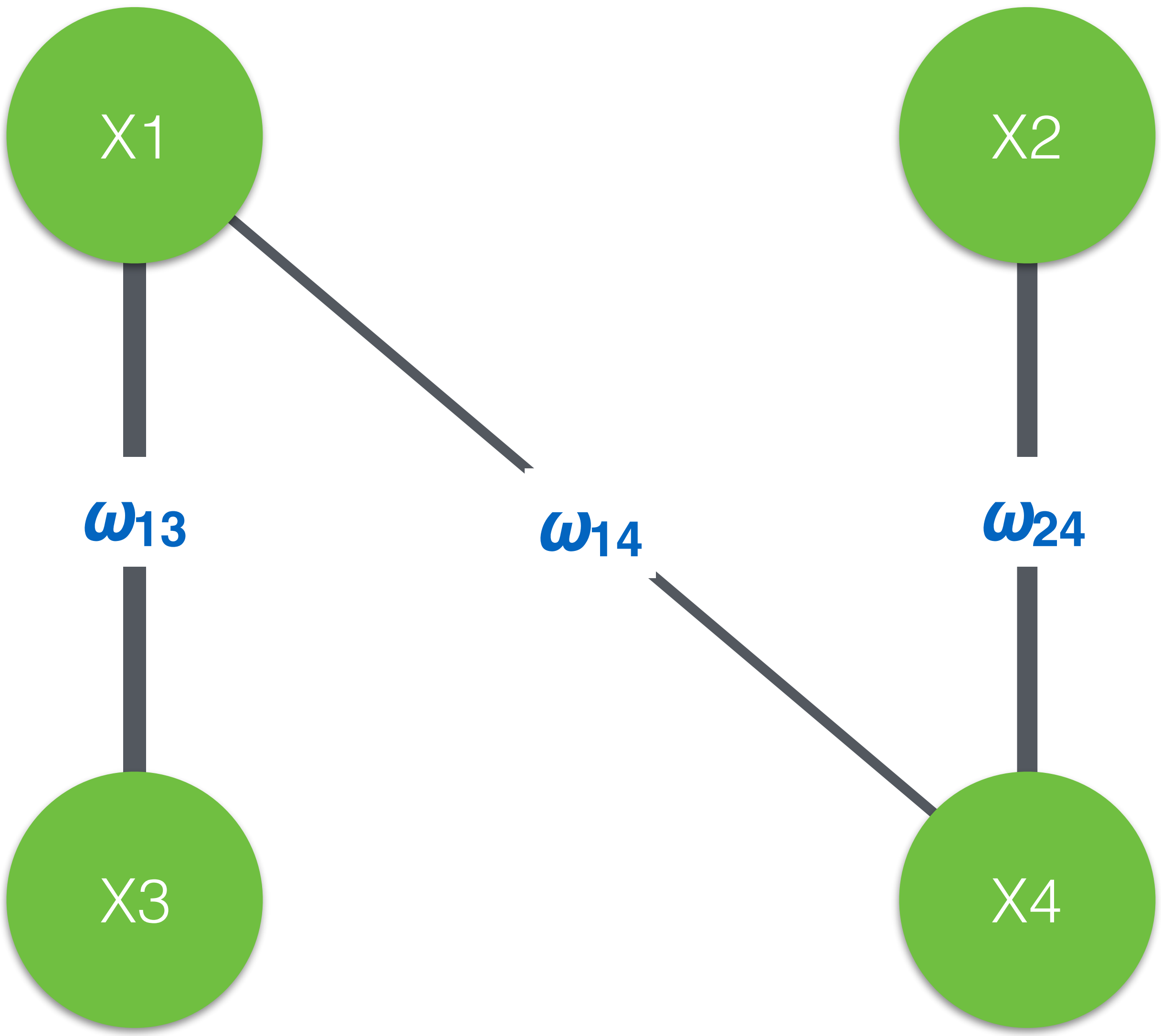


AND-rule:

if $\beta_{ij} \neq 0$ AND $\beta_{ji} \neq 0$

then $\omega_{ij} = (\beta_{ij} + \beta_{ji})/2$

else $\omega_{ij} = 0$



ω_{14} is the connection strength when controlled for all other variables

L_1-regularization

Performance

- High specificity
  - few false positives
- Moderate sensitivity
  - some false negatives
- Converges to the true network
  - with increasing sample size, more and more true edges are recovered
- For binary data: `IsingFit()` (in package IsingFit)
  - L_1 regularization on node wise logistic regressions
- For multivariate normal data: `EBICglasso()` (in package qgraph)
  - L_1 regularization on the inverse covariance matrix

L_1-regularization

These packages do everything (node-wise L1-regularized regression, choosing the tuning parameter, apply the AND-rule, etc) for you! You only have to enter your data in IsingFit() and EBICglasso() and you get the best fitting network! :-)

Performance

- High specificity
  - few false positives
- Moderate sensitivity
  - some false negatives
- Converges to the true network
  - with increasing sample size, more and more true edges are recovered
- For binary data: `IsingFit()` (in package IsingFit)
  - L_1 regularization on node wise logistic regressions
- For multivariate normal data: `EBICglasso()` (in package qgraph)
  - L_1 regularization on the inverse covariance matrix

Rule of thumb

How much observations do I need?

- Rule of thumb for binary data
  - The minimal number of observations: (p(p-1)/2 + p)*5
    - p: number of variables
    - p(p-1)/2 is the number of possible edges to be estimated
    - + p because the Ising model also estimates thresholds
- Rule of thumb for gaussian data
  - The minimal number of observations: (p(p-1)/2)*5

Bühlmann, P. & van de Geer, S. (2011). Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer.

Recap

- L_1 regularization is used to find the optimal balance between parsimony and goodness of fit and to circumvent multiple testing problems

- L_1 regularization sets some coefficients to exactly zero

- Connections are conditional dependencies (direct relationships after controlling for all others)

- Assuming that the data are realisations of a sparse network of pairwise interactions, these procedures converge to the true network

Literature

About performance of IsingFit (with elaborate Supplement about Ising model and regularization):

- Van Borkulo, C. D., Borsboom, D., Epskamp, S., Blanken, T. F., Boschloo, L., Schoevers, R. A., & Waldorp, L. J. (2014). A new method for constructing networks from binary data. *Scientific Reports*, 4(5918).

Tutorial about regularization:

- https://arxiv.org/abs/1607.01367

About performance EBICglasso:

- https://arxiv.org/pdf/1606.05771v1.pdf

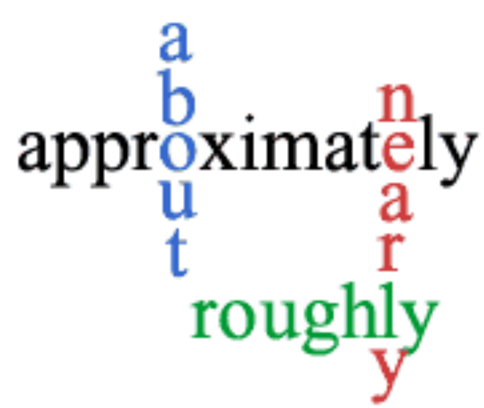
Practical

- Open Assignment_Day3_Part**2**.pdf
- Just follow the steps!
- Agree on time to start with third and final part of today…

Psych Systems

Psychological Networks
Amsterdam Winter School

Day 3, part 3
Recent advancements

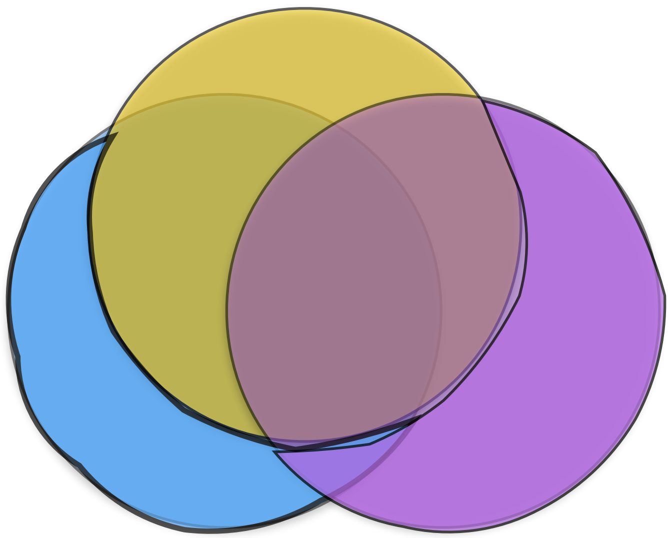


Claudia van Borkulo

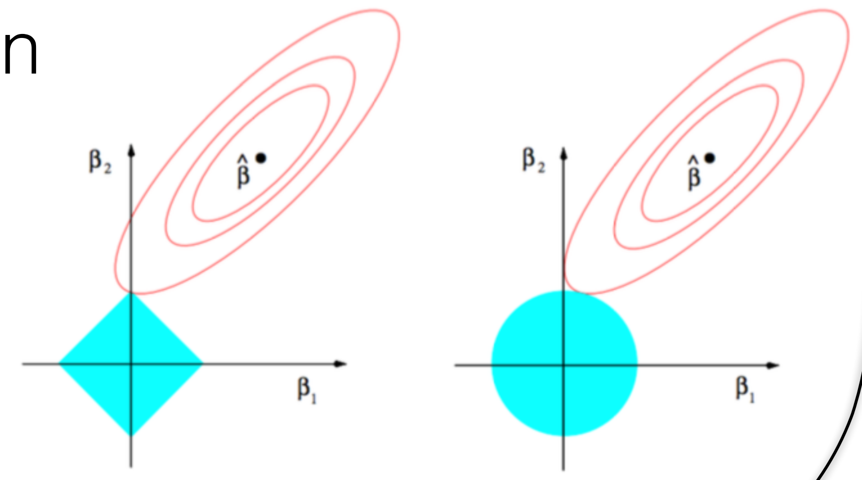
Februari 15th, 2017

Recap

- Conditional (in)dependence
  - a thorough understanding is important to work with graphical models (Ising model, Gaussian Graphical model)



- Estimation of graphical models
  - IsingFit, EBICglasso
  - regularization to find optimal balance between parsimony and goodness of fit



Outline

1. The basics: Conditional independence
2. Estimating graphical models with L1 regularization
3. **Recent advancements**
   - **Network stability**
   - **Network comparison**
   - **MGM**

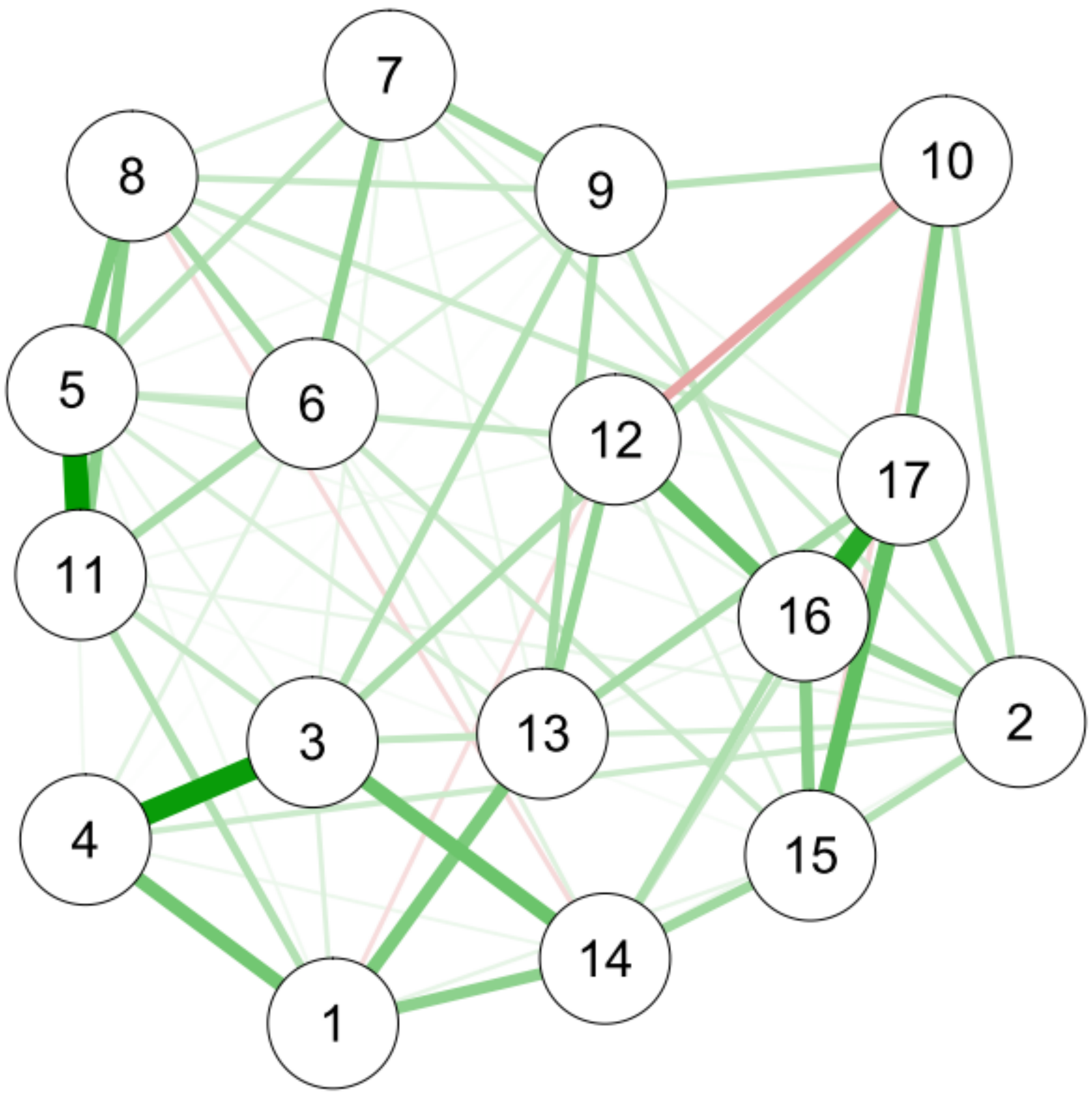
Network stability

Network stability

Suppose we want to write a paper on this data set

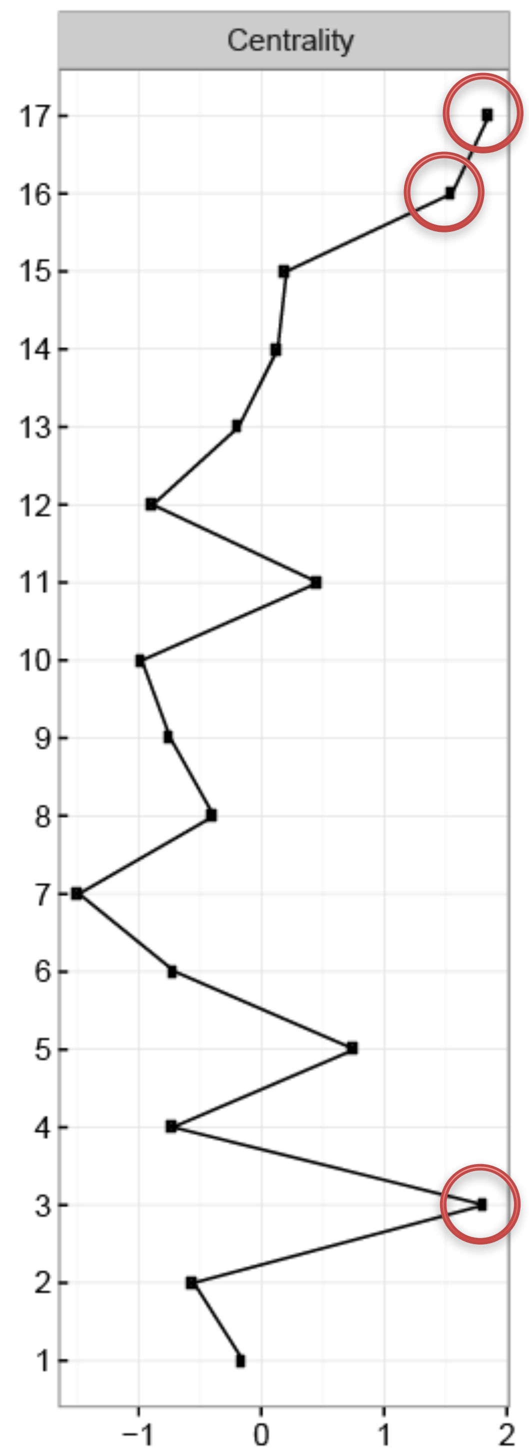
- 180 women with PTSD diagnosis, 17-item screener
- Data from DOI 10.1037/a0016227, freely available at https://datashare.nida.nih.gov/protocol/nida-ctn-0015

Dataset 1



1: Avoid reminders of the trauma
2: Bad dreams about the trauma
3: Being jumpy or easily startled
4: Being over alert
5: Distant or cut off from people
6: Feeling emotionally numb
7: Feeling irritable
8: Feeling plans wont come true
9: Having trouble concentrating
10: Having trouble sleeping
11: Less interest in activities
12: Not able to remember
13: Not thinking about trauma
14: Physical reactions
15: Reliving the trauma
16: Upset when reminded of trauma
17: Upsetting thoughts or images

Dataset 1

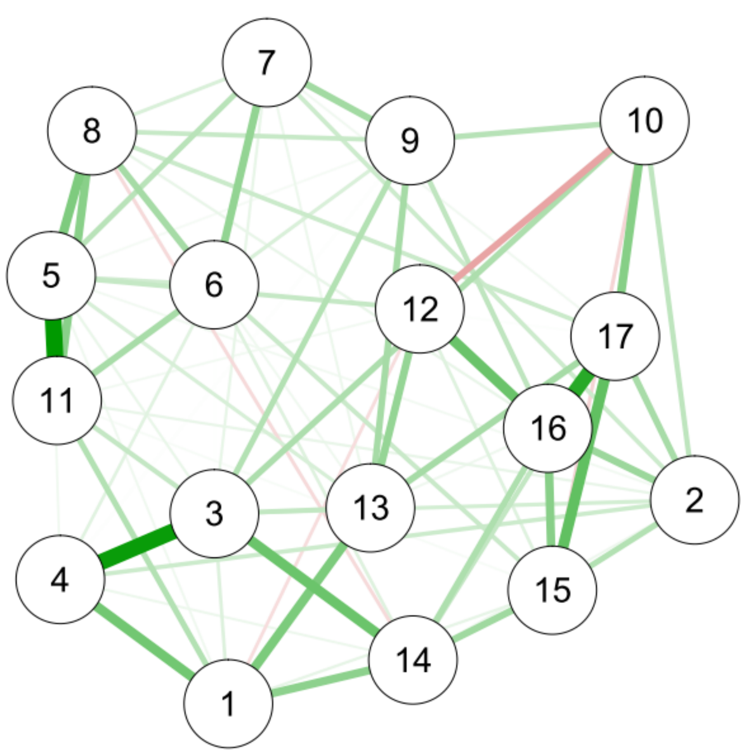


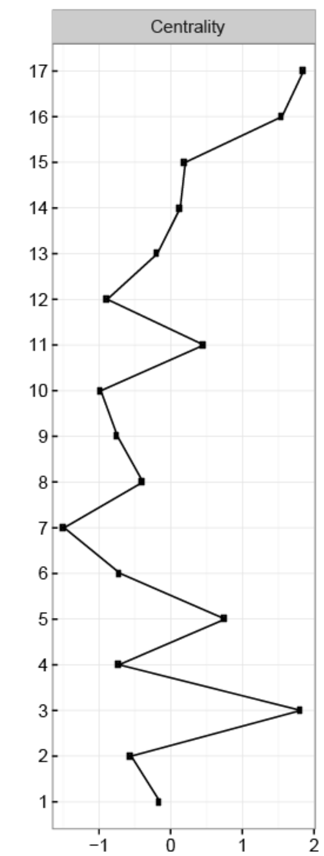
1: Avoid reminders of the trauma
2: Bad dreams about the trauma
3: Being jumpy or easily startled
4: Being over alert
5: Distant or cut off from people
6: Feeling emotionally numb
7: Feeling irritable
8: Feeling plans wont come true
9: Having trouble concentrating
10: Having trouble sleeping
11: Less interest in activities
12: Not able to remember
13: Not thinking about trauma
14: Physical reactions
15: Reliving the trauma
16: Upset when reminded of trauma
17: Upsetting thoughts or images

Dataset 1

Paper

- Strong positive connections between 3—4, 5—11, 16—17

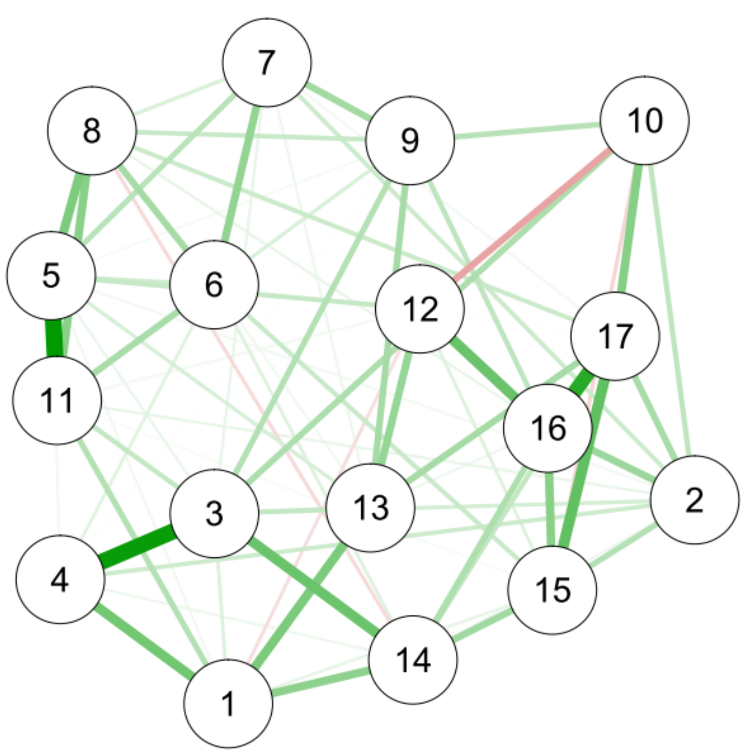


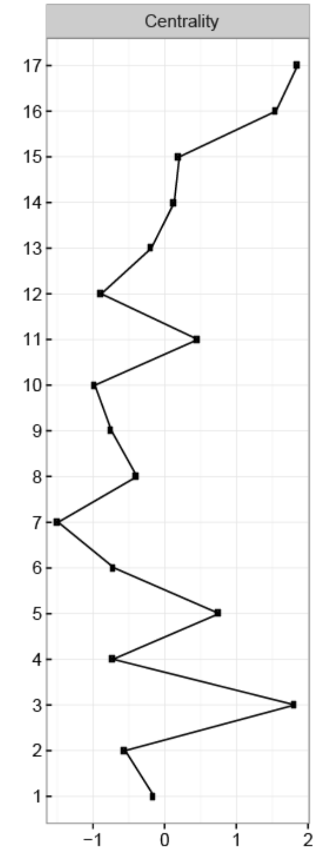


Dataset 1

Paper

- Strong positive connections between 3—4, 5—11, 16—17
- Strong negative edge between 10—12

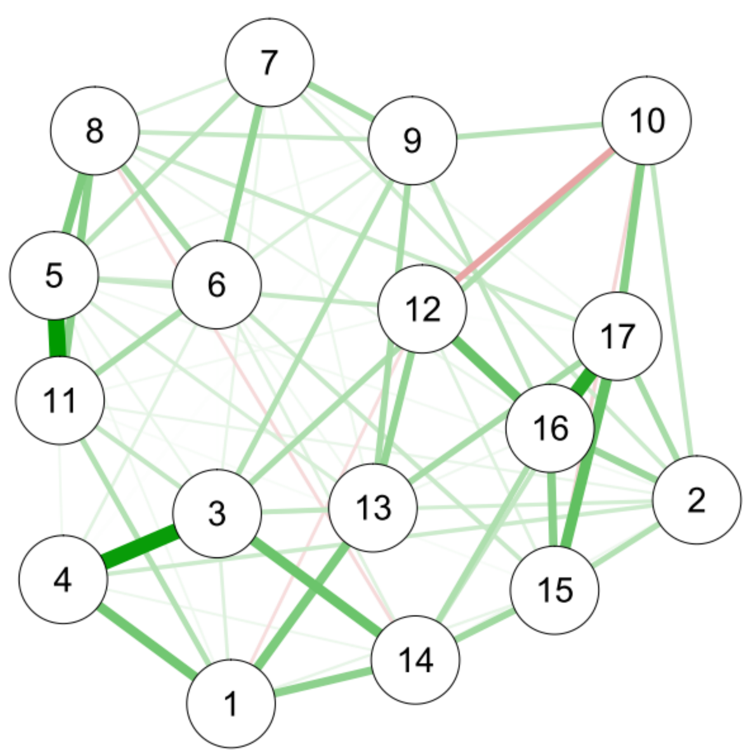


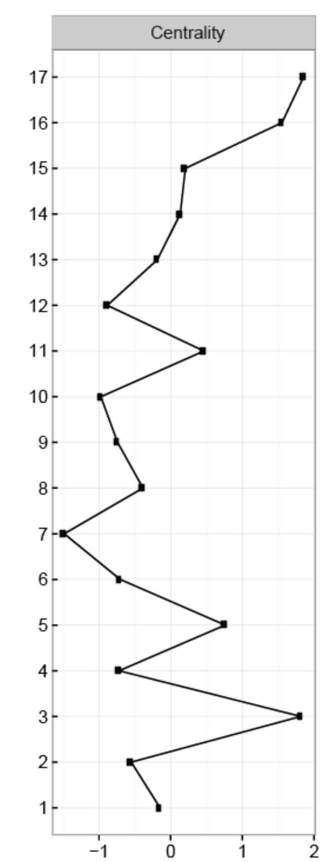


Dataset 1

Paper

- Strong positive connections between 3—4, 5—11, 16—17
- Strong negative edge between 10—12
- Most central nodes: 3, 16, 17 ⟶ consider as targets in intervention study





Dataset 1

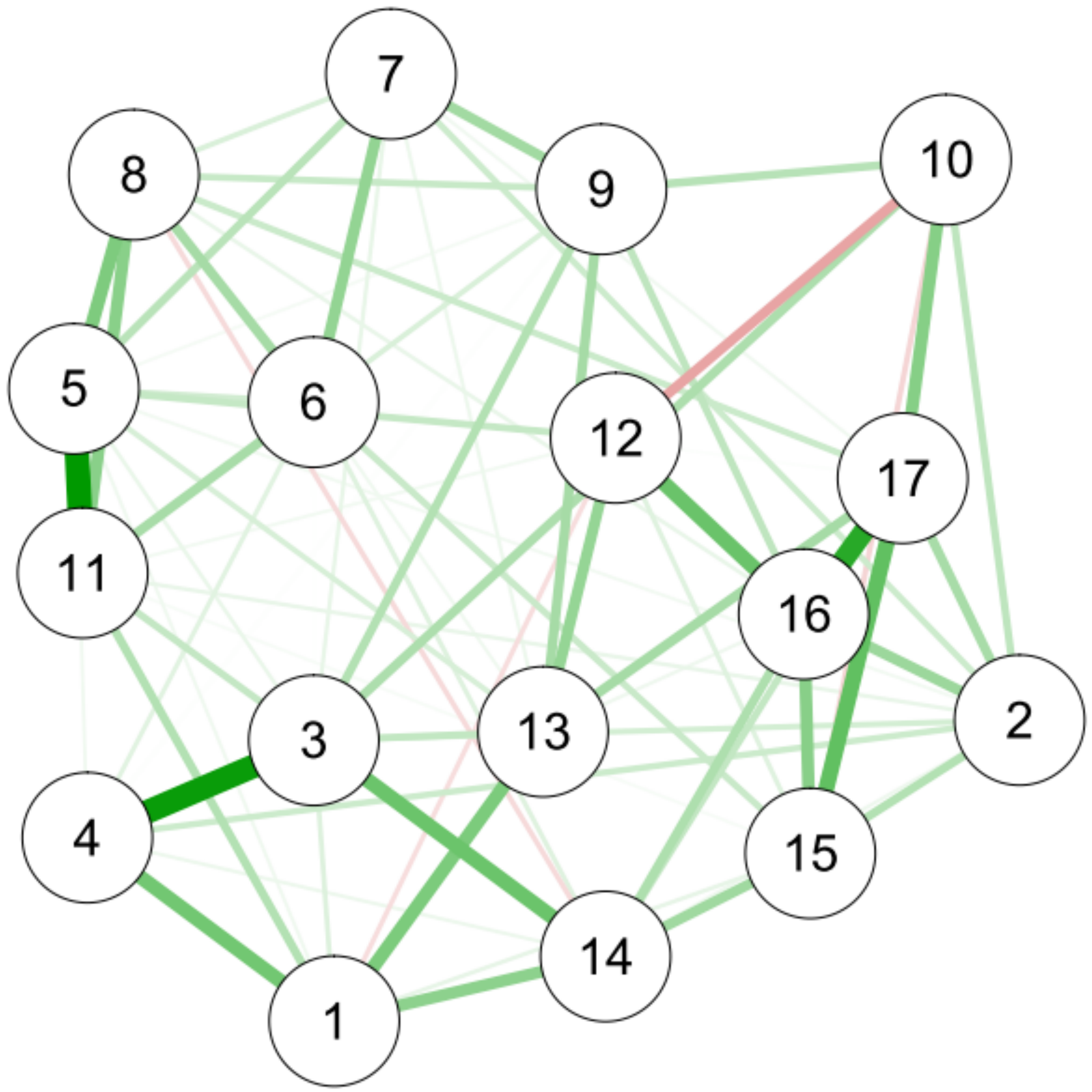
Paper published … partytime!

Dataset 1

Paper published … partytime!

Dataset 2

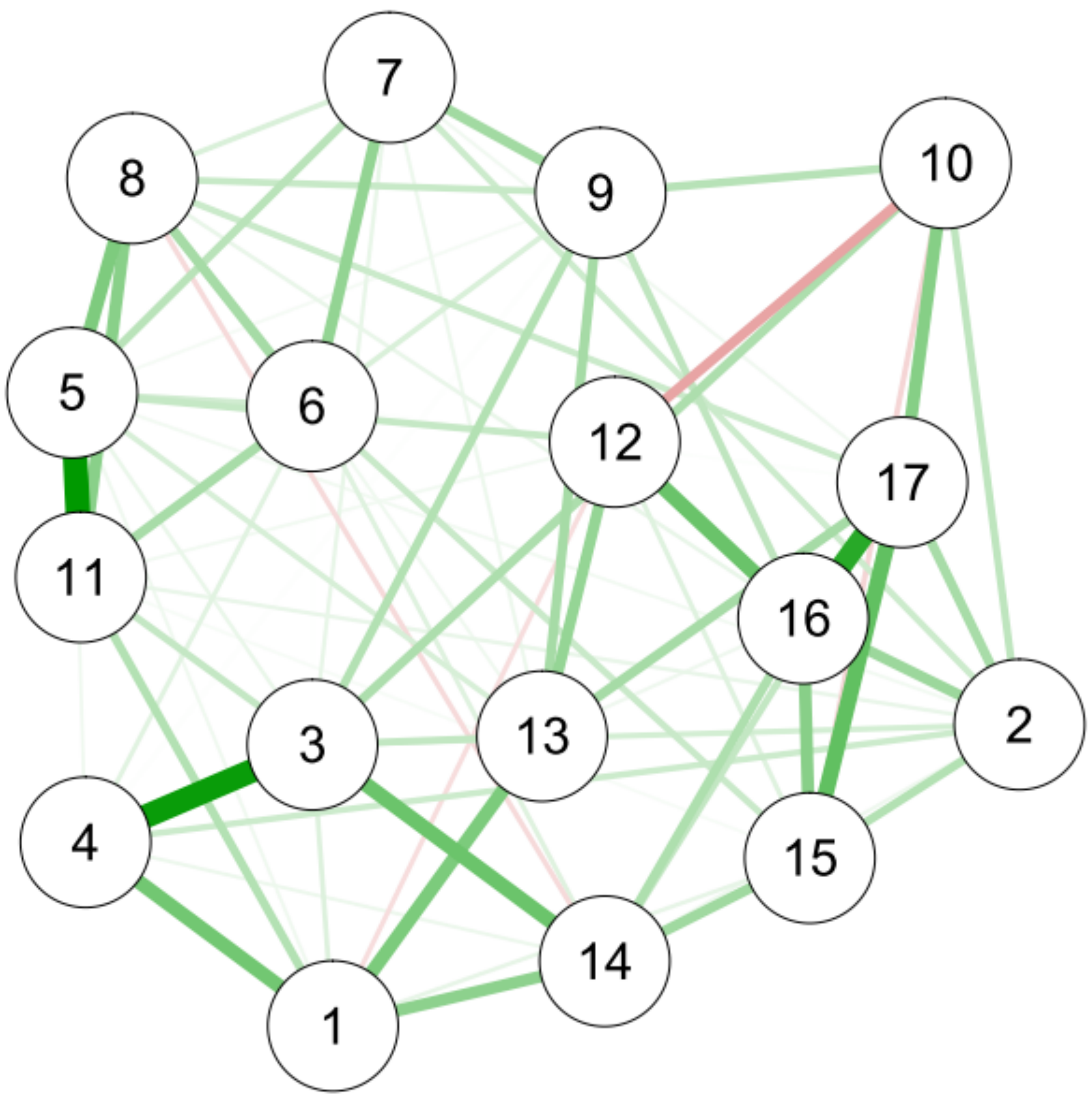
Now imagine we find another dataset, same sample size, female PTSD patients



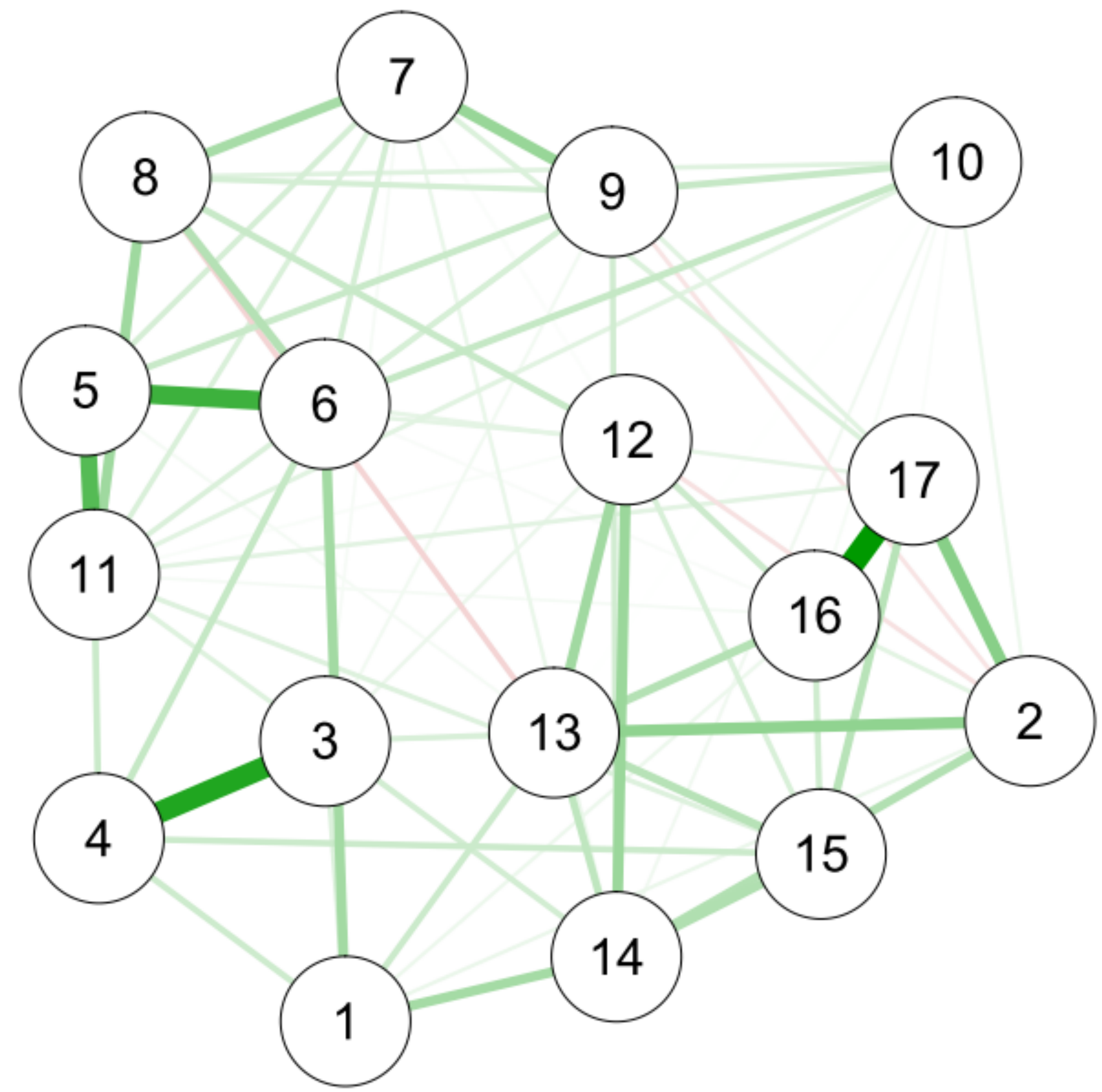
First dataset, n=180

Dataset 2

Now imagine we find another dataset, same sample size, female PTSD patients



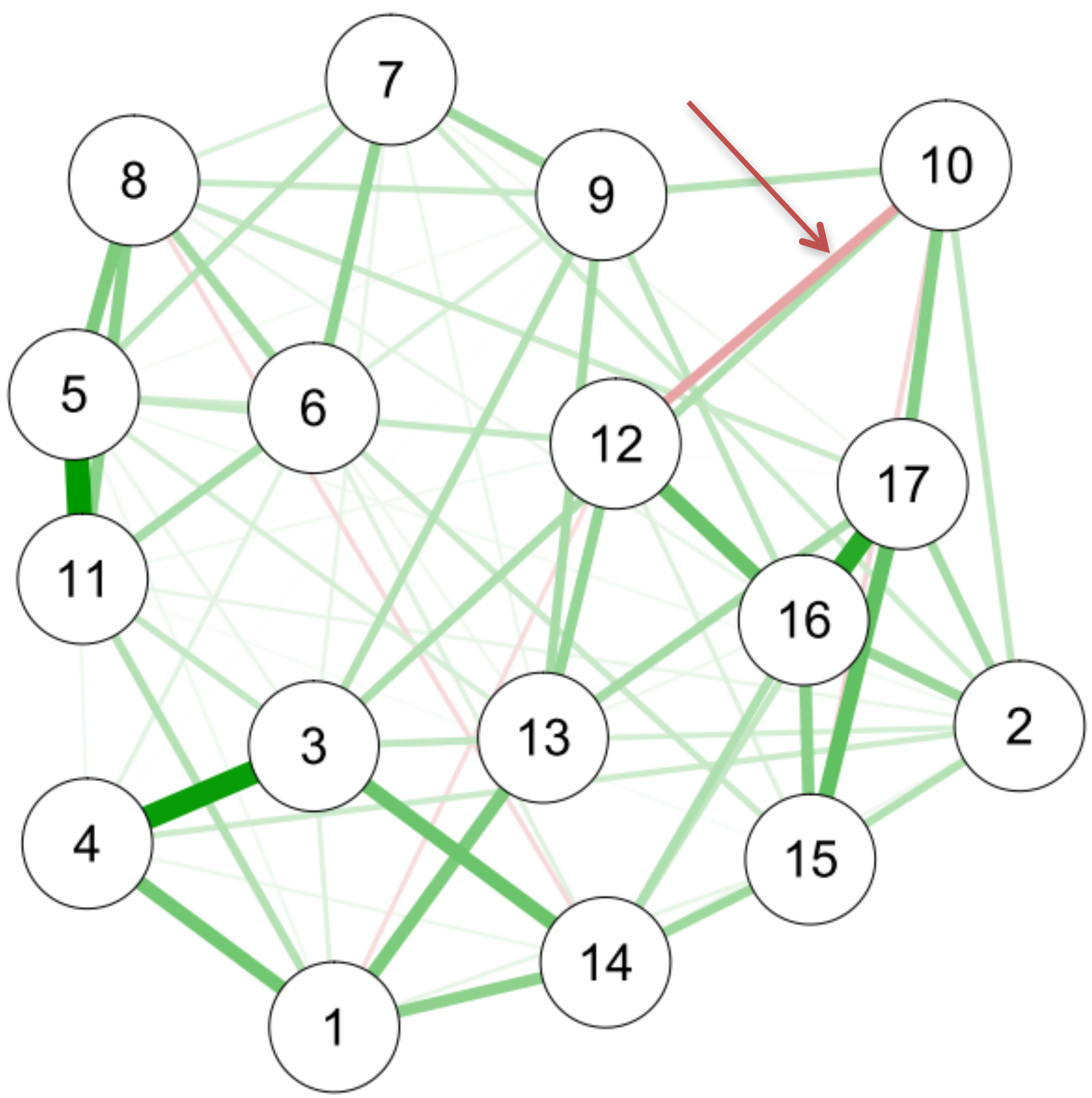
First dataset, n=180



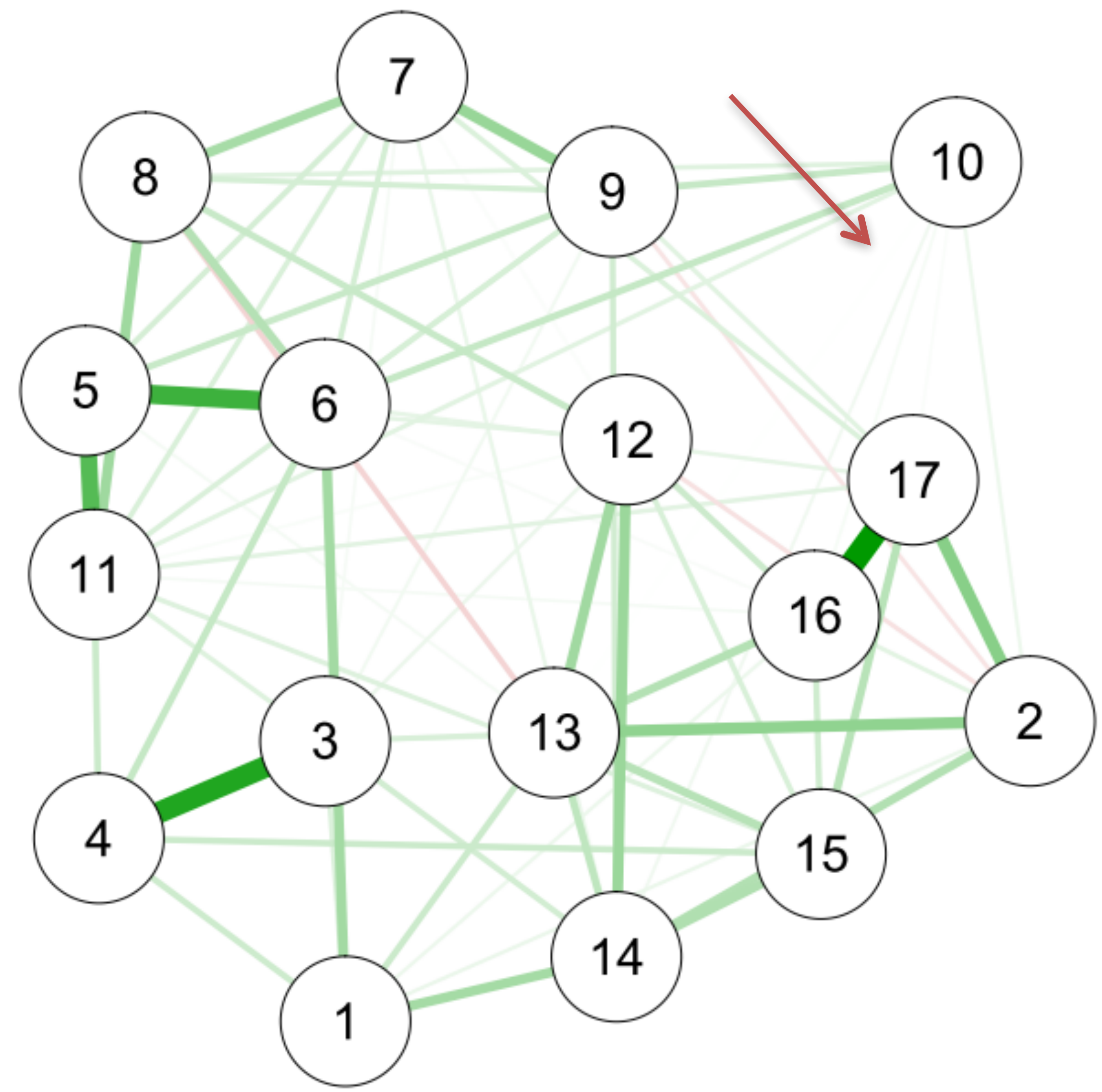
Second dataset, n=179

Dataset 2

Now imagine we find another dataset, same sample size, female PTSD patients



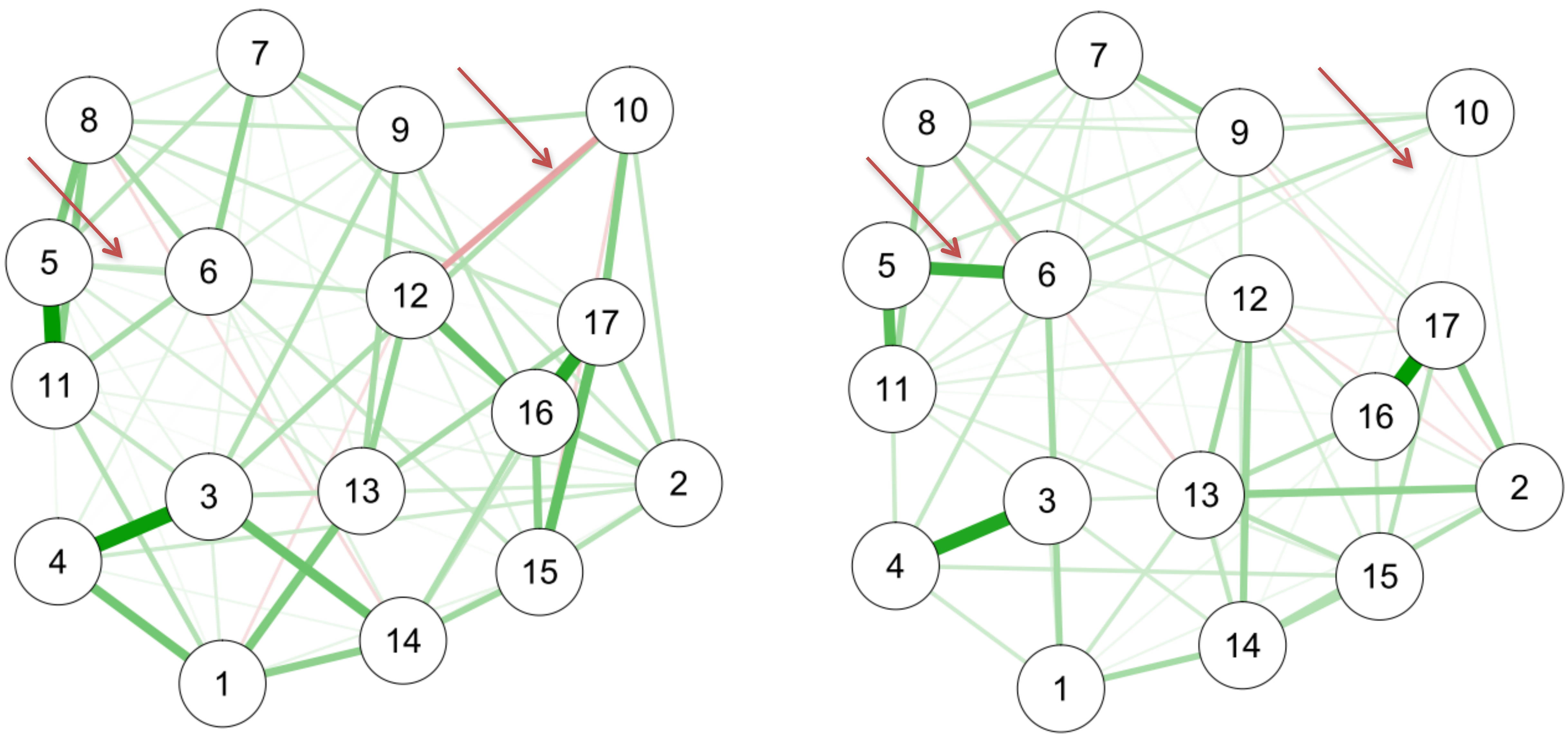
First dataset, n=180



Second dataset, n=179

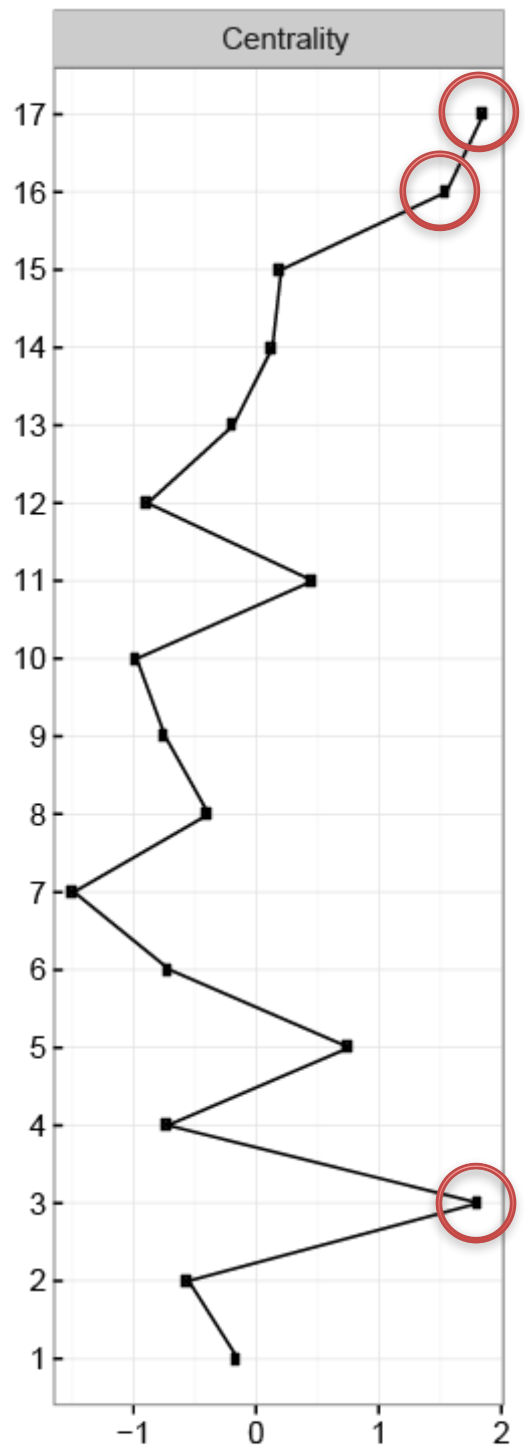
Dataset 2

Now imagine we find another dataset, same sample size, female PTSD patients

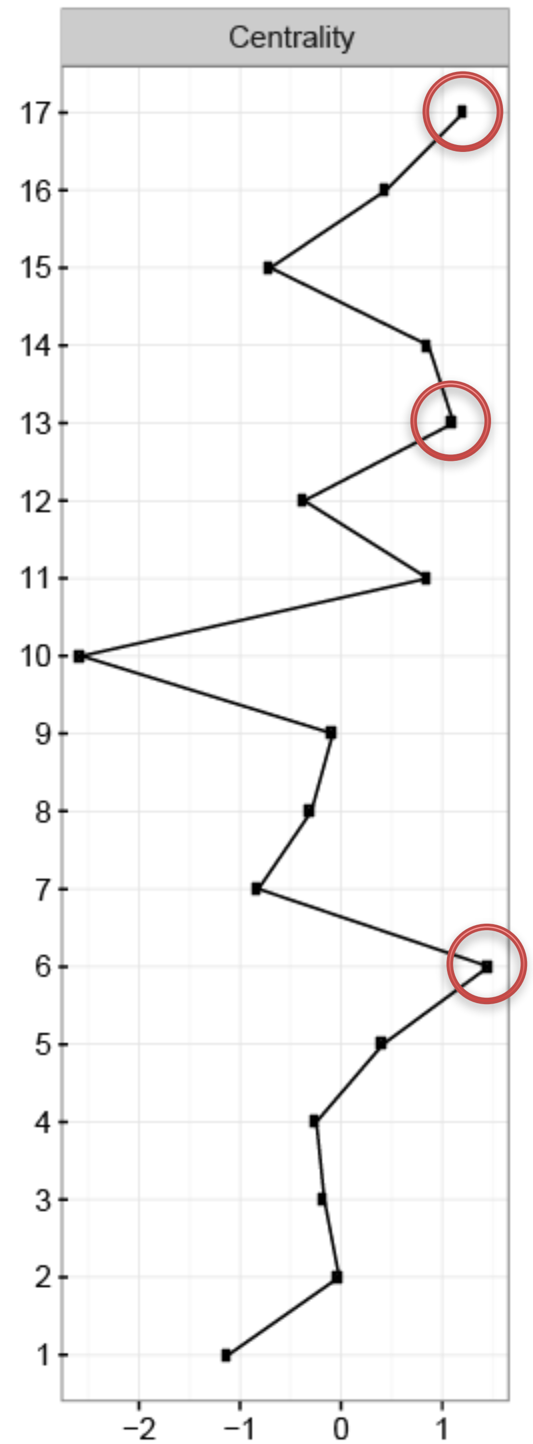


First dataset, n=180 Second dataset, n=179

Dataset 2



First dataset, n=180



Second dataset, n=179

Dataset 2

Dataset 2

Network stability

- To avoid a replicability crisis, we need to investigate and report how *stable* our parameter estimates are
- Especially relevant because our research may have clinical implications for patients
  - E.g.: what are the most central symptoms that ought to be treated?

Network stability

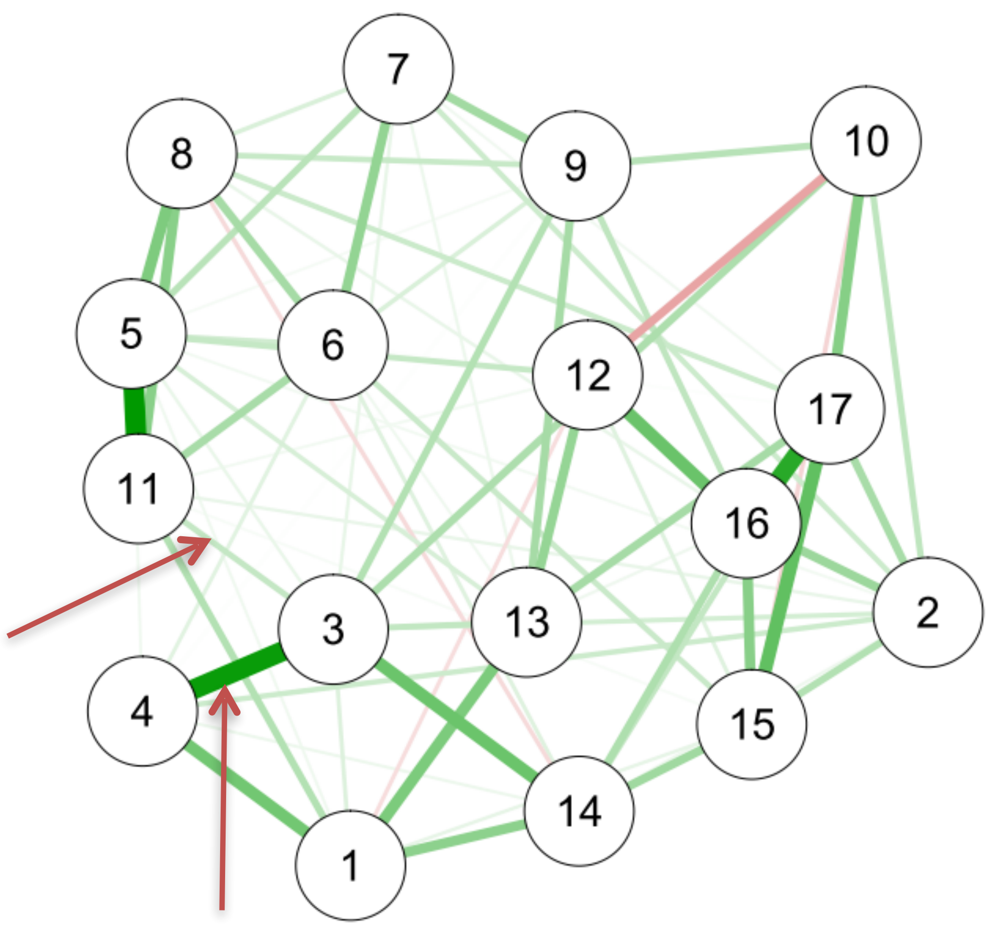
Two main questions:

- Stability of **edge weights**
- Stability of **centrality indices**

Network stability

Two main questions:

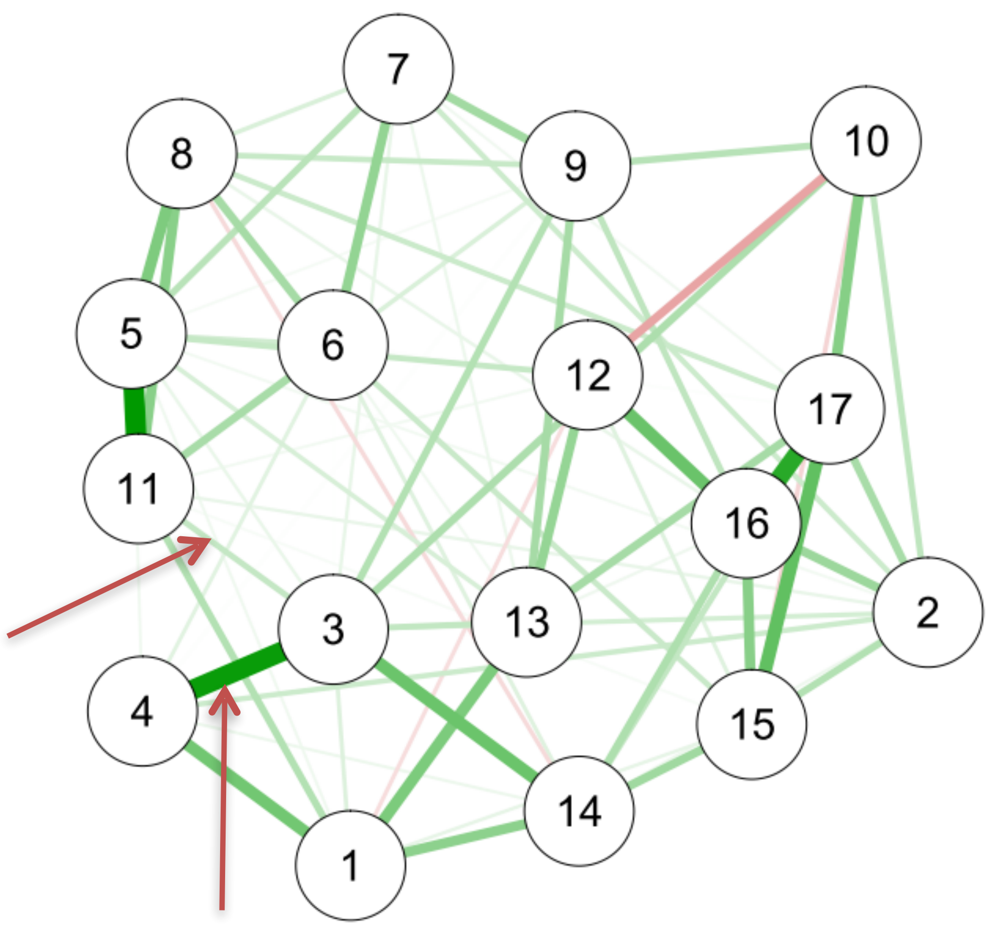
- Is edge 3—4 meaningfully larger than edge 3—11?

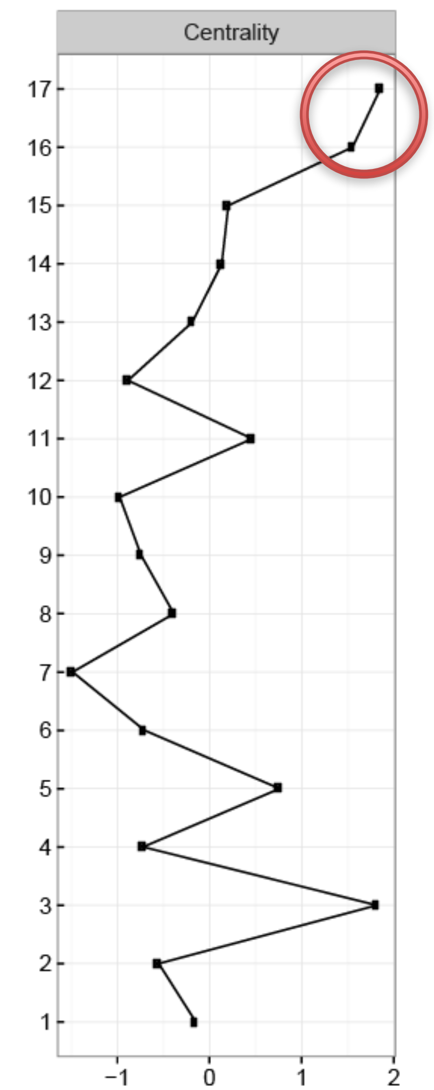


Network stability

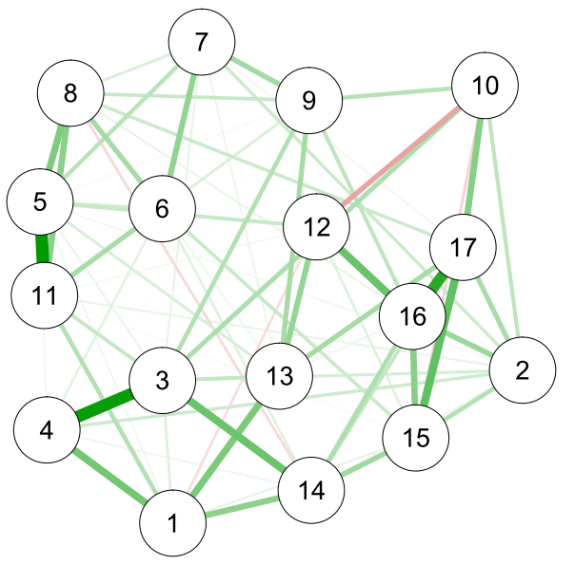
Two main questions:

- Is edge 3—4 meaningfully larger than edge 3—11?
- Is node 17 substantially more central than node 16?





EDGE WEIGHT STABILITY



Bootstrapping edge weights

Bootstrapping edge weights

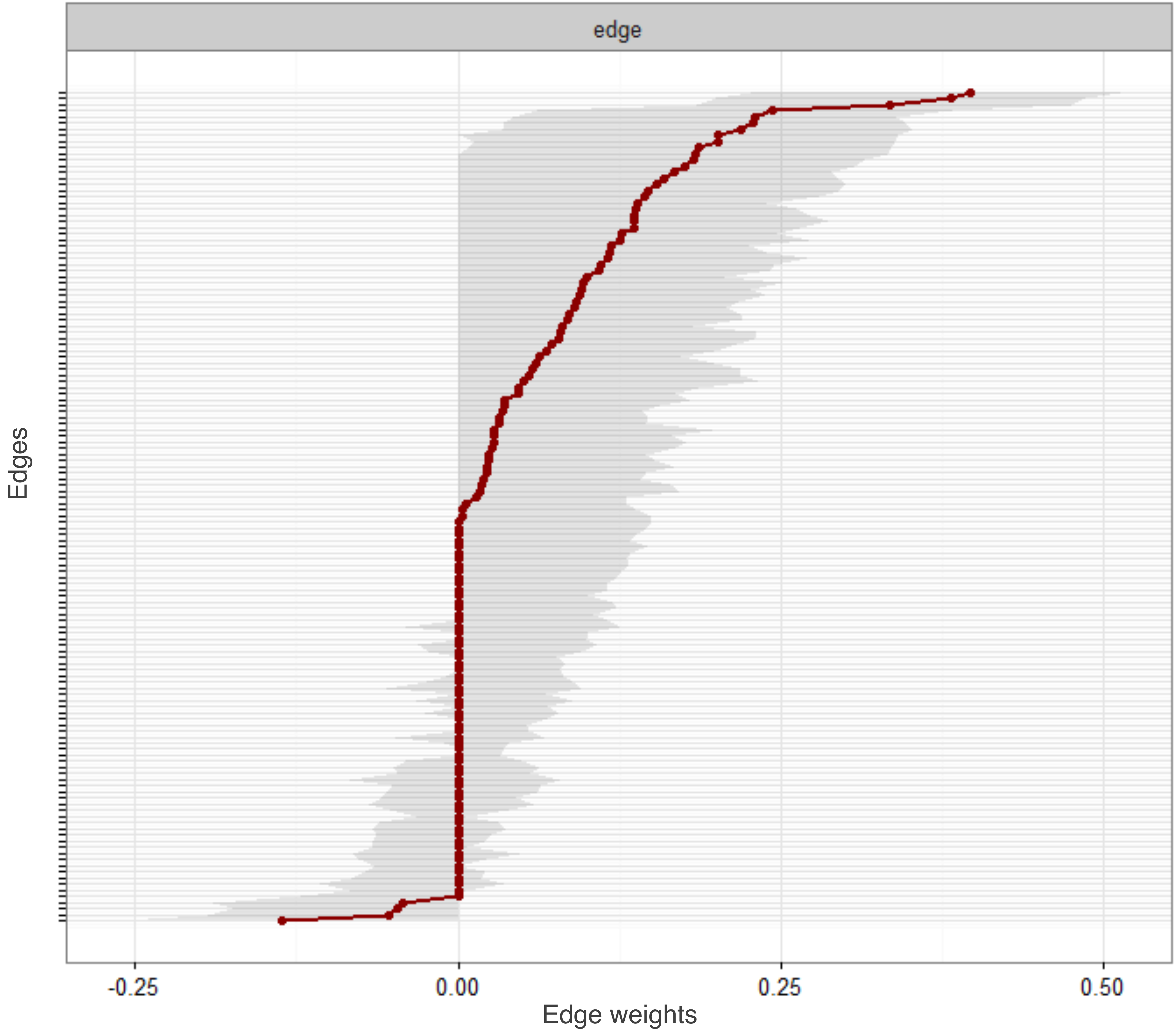
- Is edge 3—4 (0.42) stronger than edge 3—11 (0.14)?
- Obtain CI by *bootstrapping*

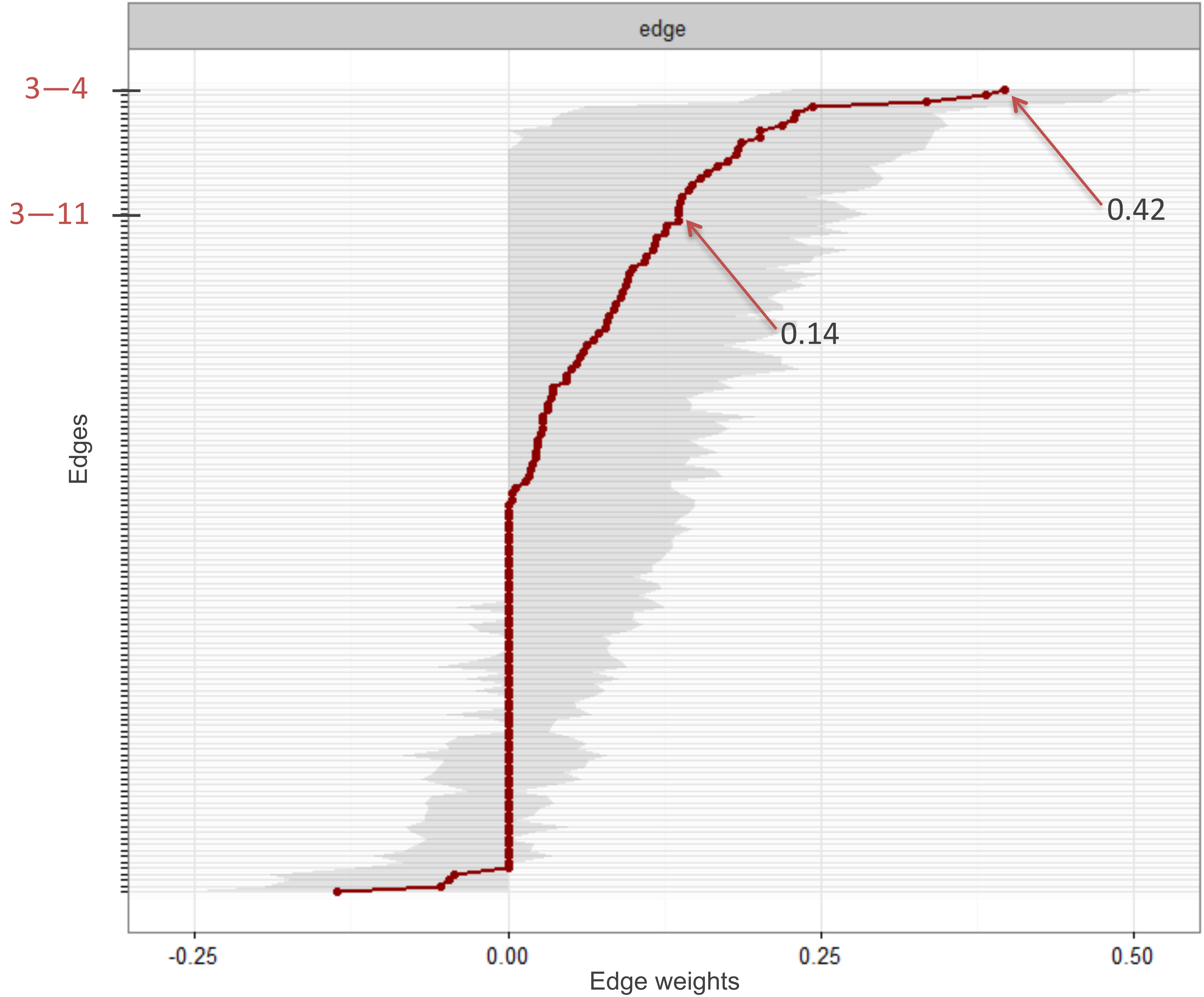
Bootstrapping edge weights

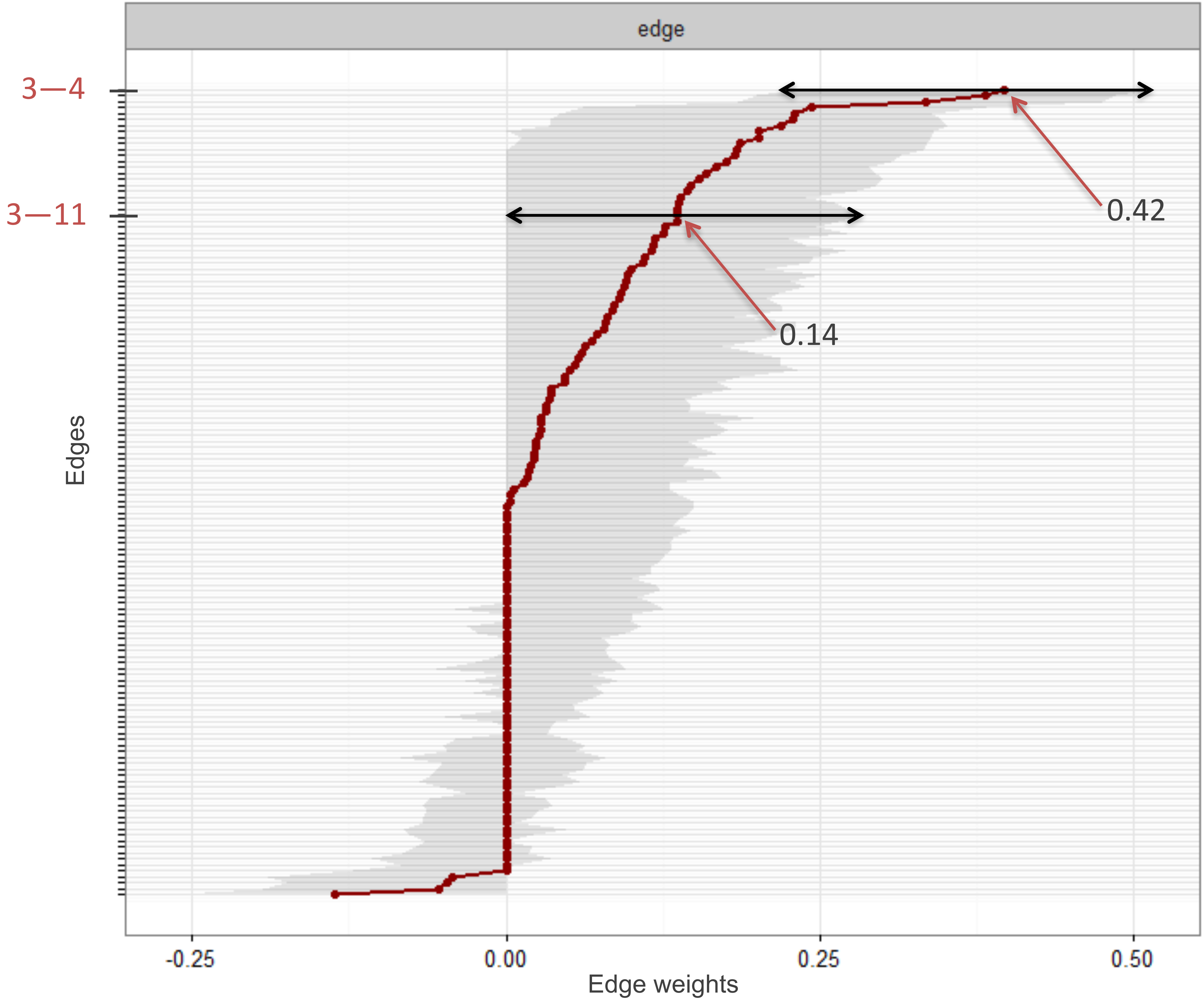
- Is edge 3—4 (0.42) stronger than edge 3—11 (0.14)?
- Obtain CI by *bootstrapping*

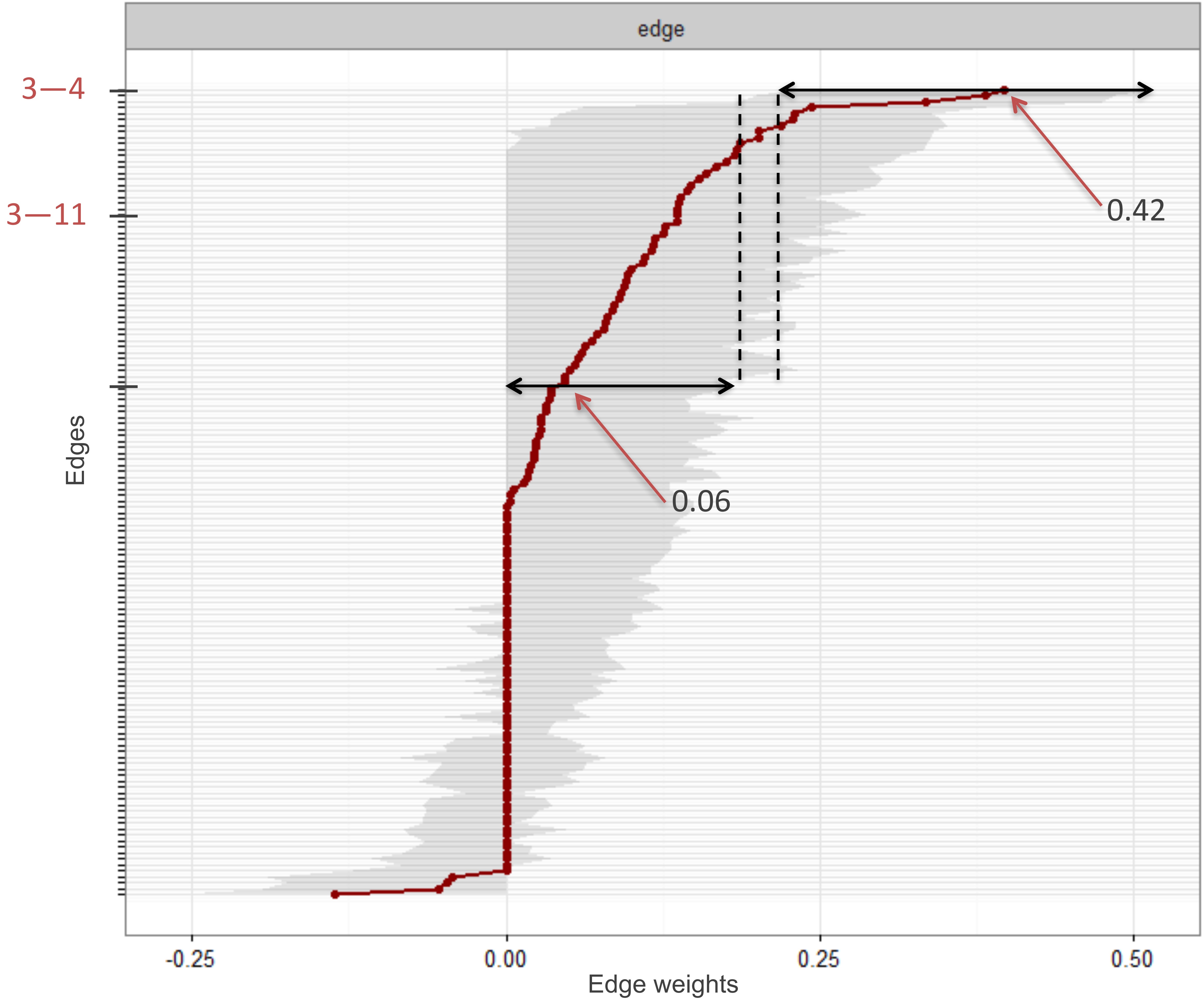
Boostrapping edge weights

- The edge weights of your sample are your *point estimates*
- Copy pasting your *sample* to create a ‘*population*’
- Take random samples (same size as original sample) from the ‘bootstrap’ population
- Estimate whatever you were estimating (edge weights)
- This gives you a distribution of estimated values and
- a confidence interval around your point estimate!


edge
Edges
-0.25
0.00
0.25
0.50
Edge weights


edge
3—4
3—11
Edges
0.42
0.14
-0.25
0.00
0.25
0.50
Edge weights

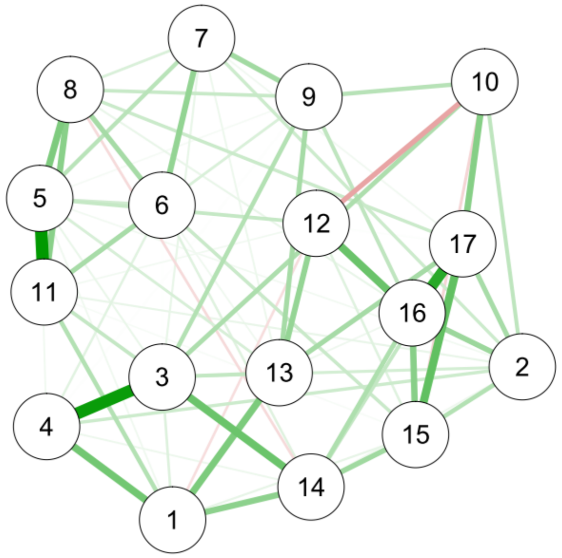

edge
3—4
3—11
0.42
0.14
Edges
-0.25
0.00
0.25
0.50
Edge weights


edge
3—4
3—11
0.42
0.06
Edges
-0.25
0.00
0.25
0.50
Edge weights

Most edges are not meaningfully different from each other because their CIs overlap.

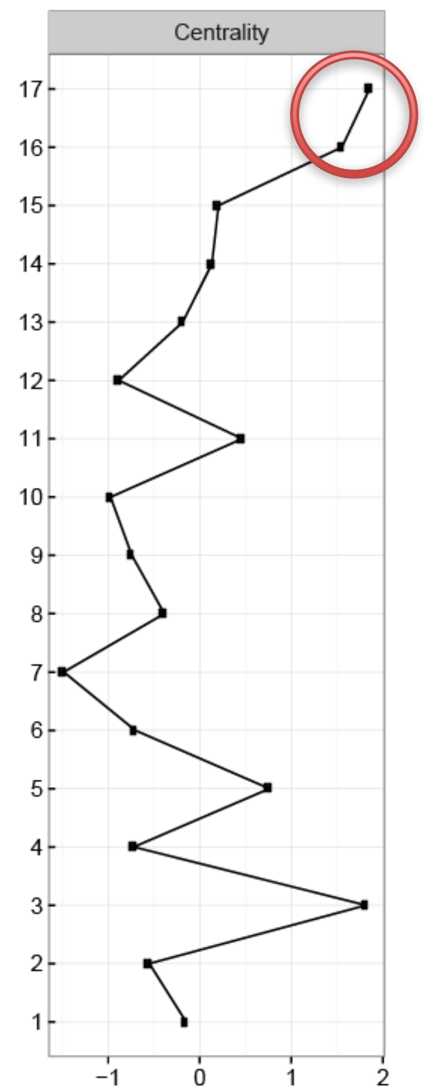
This is not really surprising: we are estimating 136 edge parameters with only 180 observations.

CENTRALITY STABILITY



Subset bootstrap

We now want to understand how stable the estimation of centrality indices is: e.g., is centralty of node 17 (1.16) substantially higher than the centrality of node 16 (0.99)



Subset bootstrap

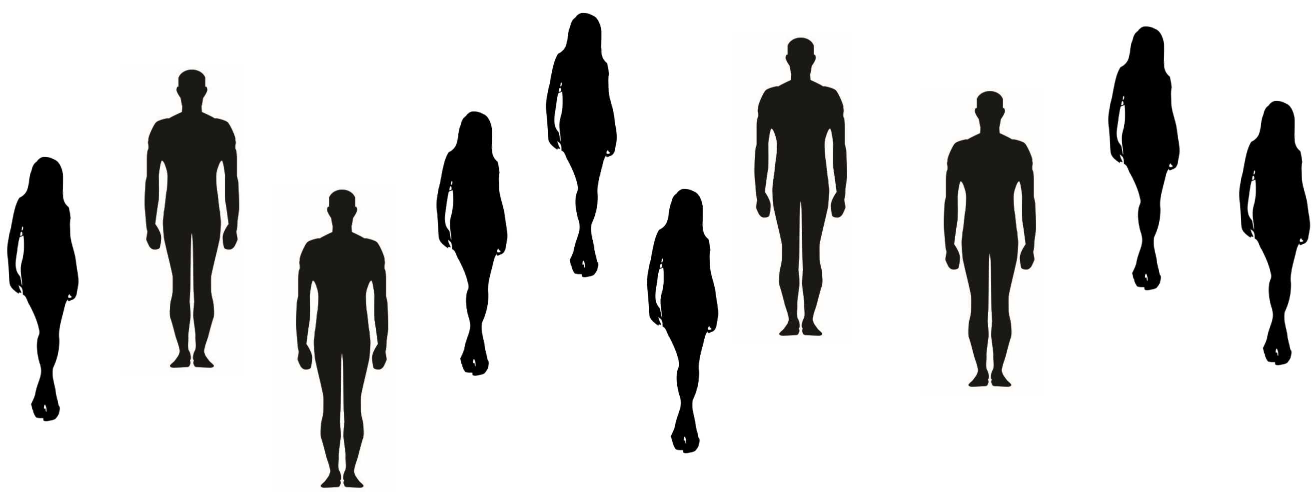
Subset bootstrap

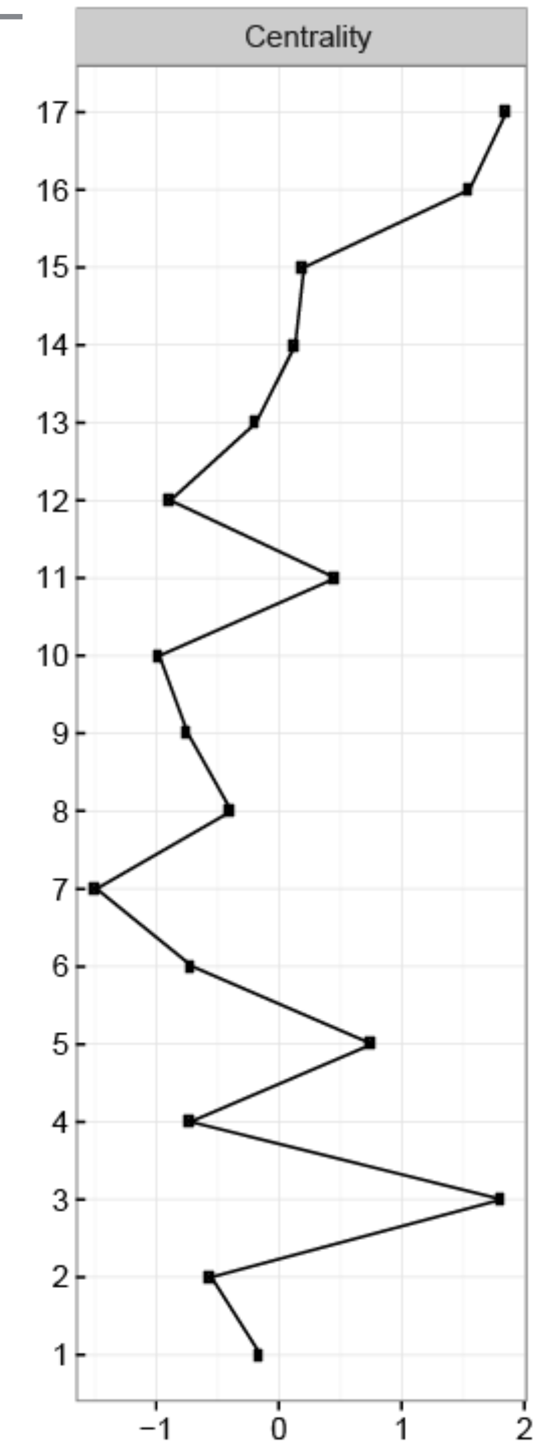
Subset bootstrap

Unfortunately, bootstrapping CIs around centrality estimates is not possible

Costenbader, E., & Valente, T. W. (2003)
DOI: 10.1016/S0378-8733(03)00012-1

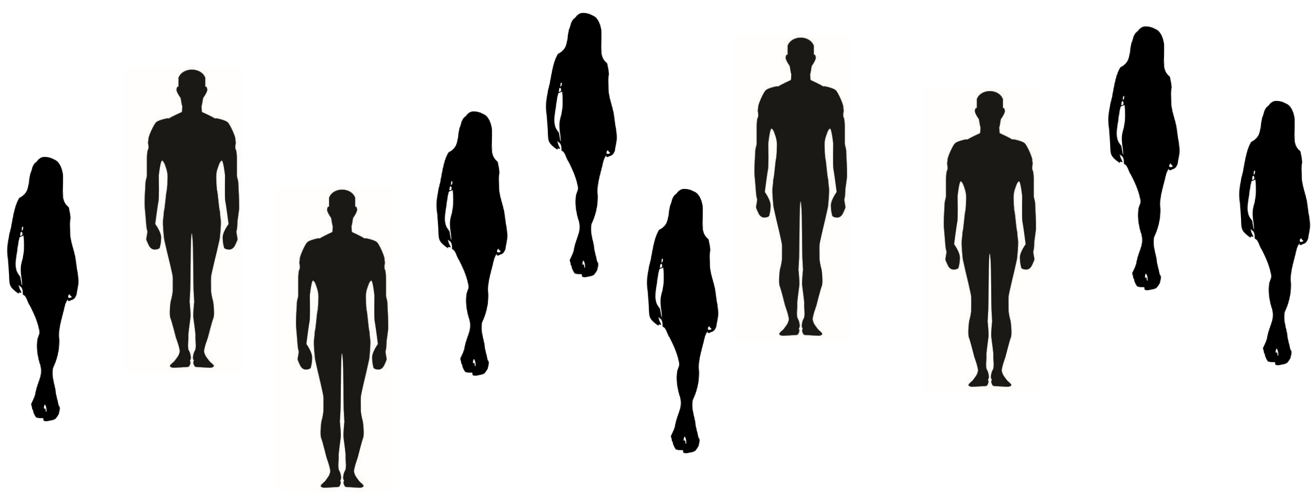
Subset bootstrap

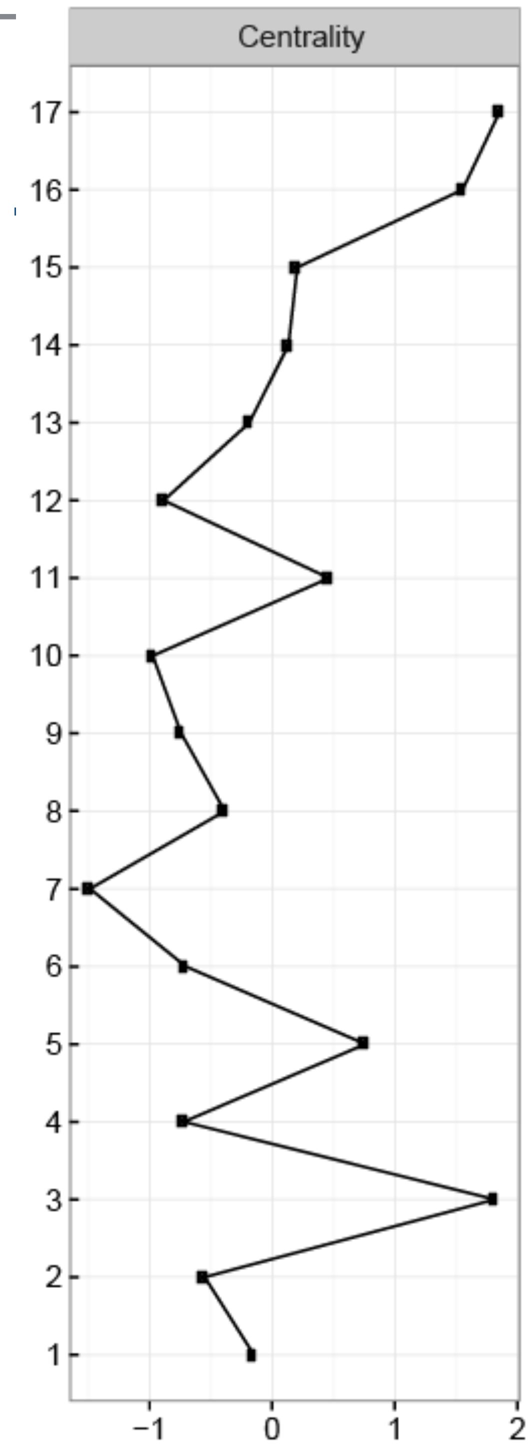




Subset bootstrap

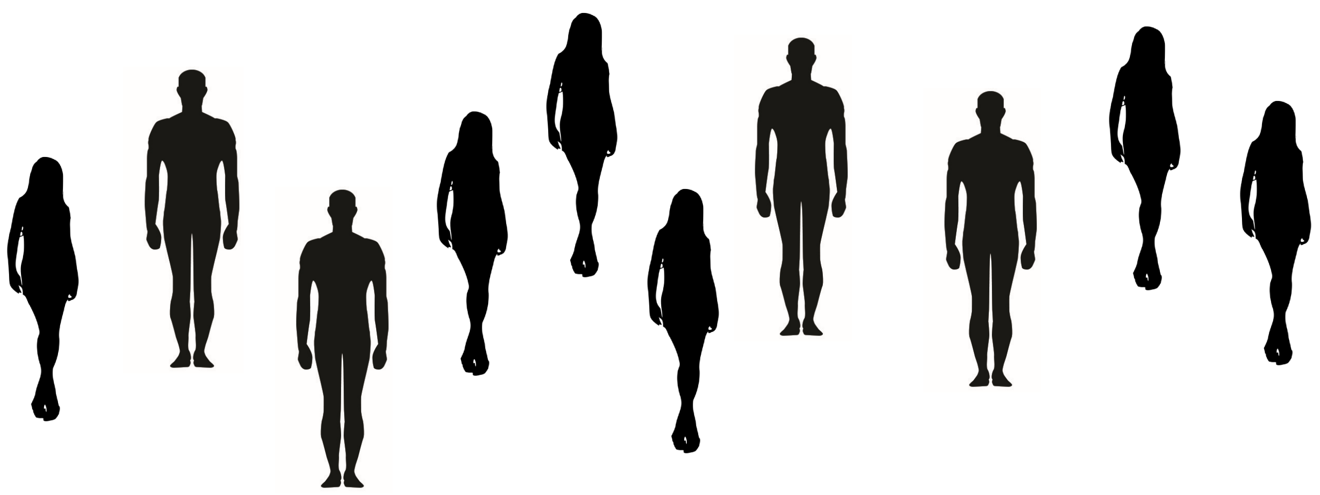
1. Obtain centrality for data (s17 > s3 > s16...





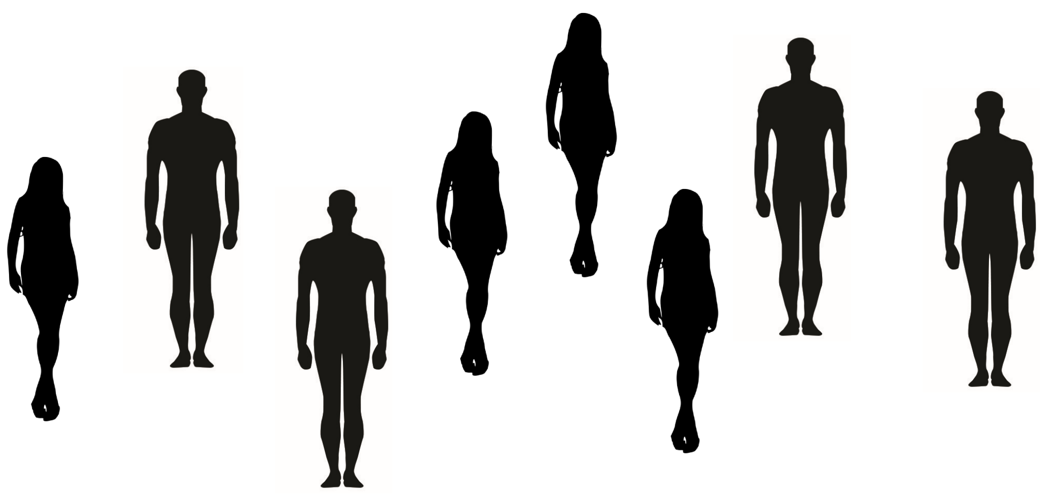
Subset bootstrap

1. Obtain centrality for data (s17 > s3 > s16...)



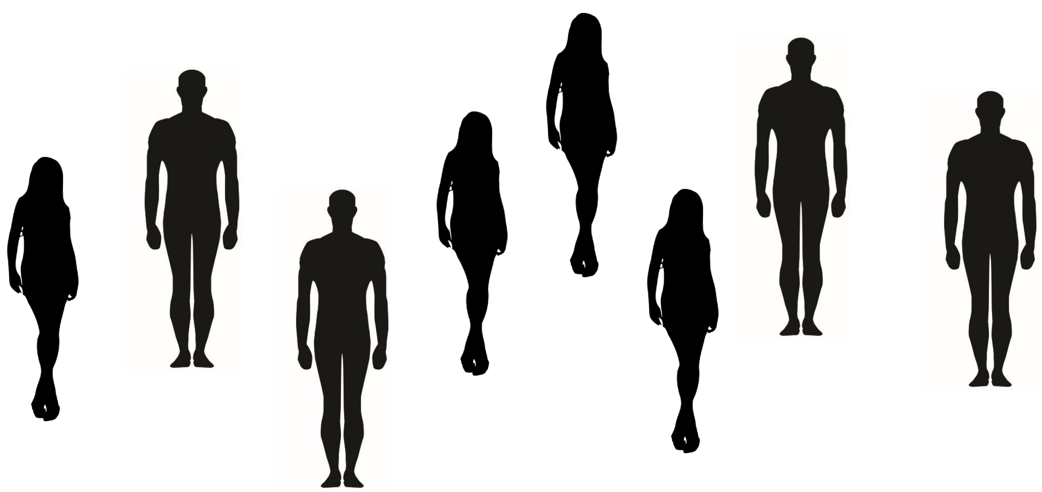
Subset bootstrap

1. Obtain centrality for data (s17 > s3 > s16...)
2. Subset data with 90% of the people



Subset bootstrap

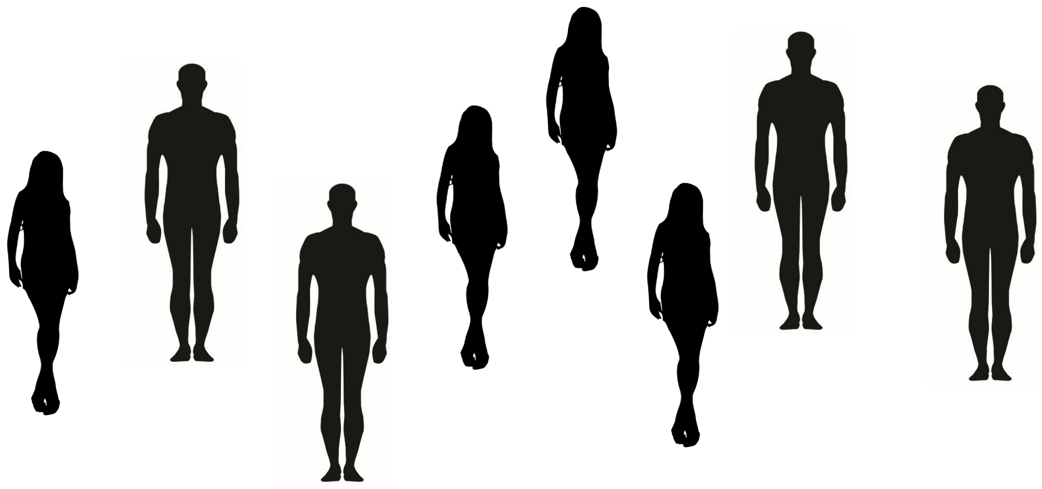
1. Obtain centrality for data (s17 > s3 > s16...)
2. Subset data with 90% of the people



3. Obtain centrality for 90% subset (s17 > s7 > s4...)

Subset bootstrap

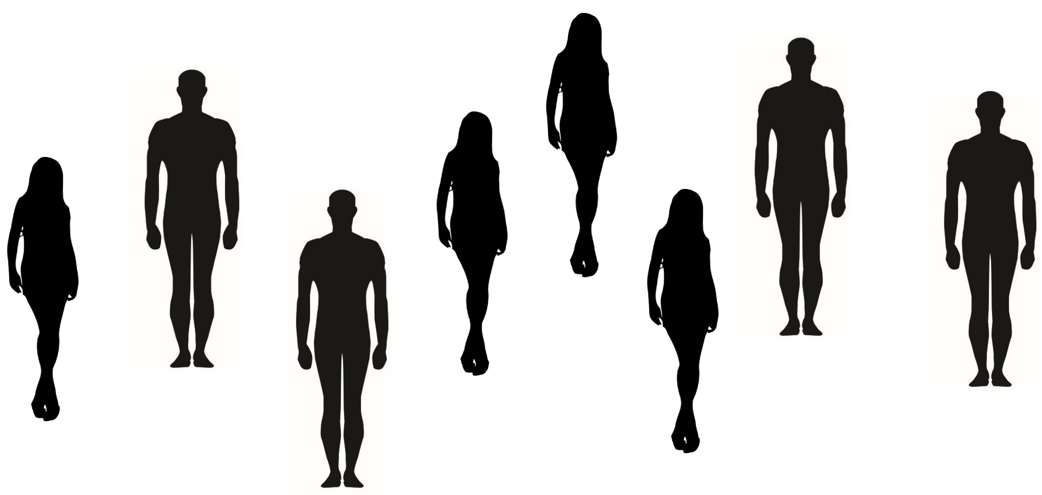
1. Obtain centrality for data (s17 > s3 > s16...)
2. Subset data with 90% of the people



3. Obtain centrality for 90% subset (s17 > s7 > s4...)
4. Subset data 80% of the people

Subset bootstrap

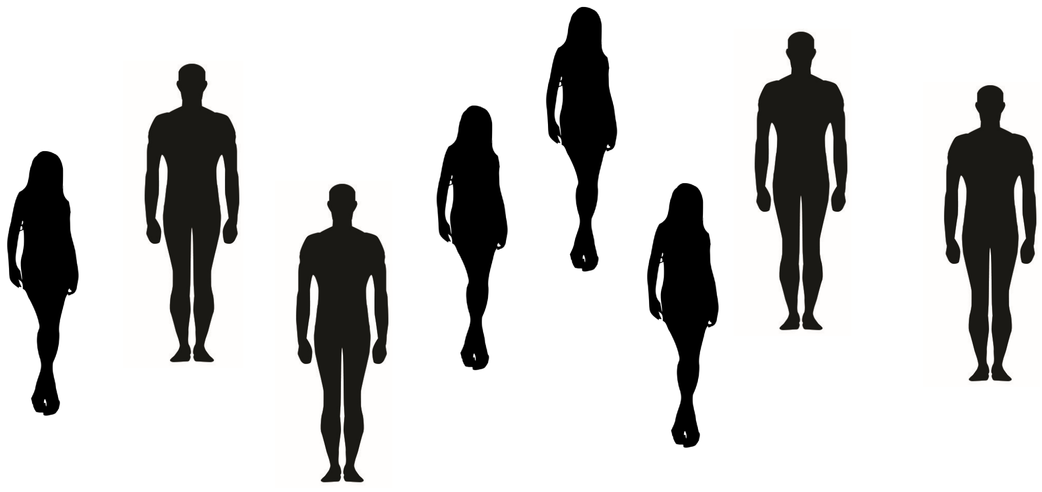
1. Obtain centrality for data (s17 > s3 > s16...)
2. Subset data with 90% of the people



3. Obtain centrality for 90% subset (s17 > s7 > s4...)
4. Subset data 80% of the people
5. Obtain centrality for 80% subset (s16 > s7 > s3...)

Subset bootstrap

1. Obtain centrality for data (s17 > s3 > s16...)
2. Subset data with 90% of the people



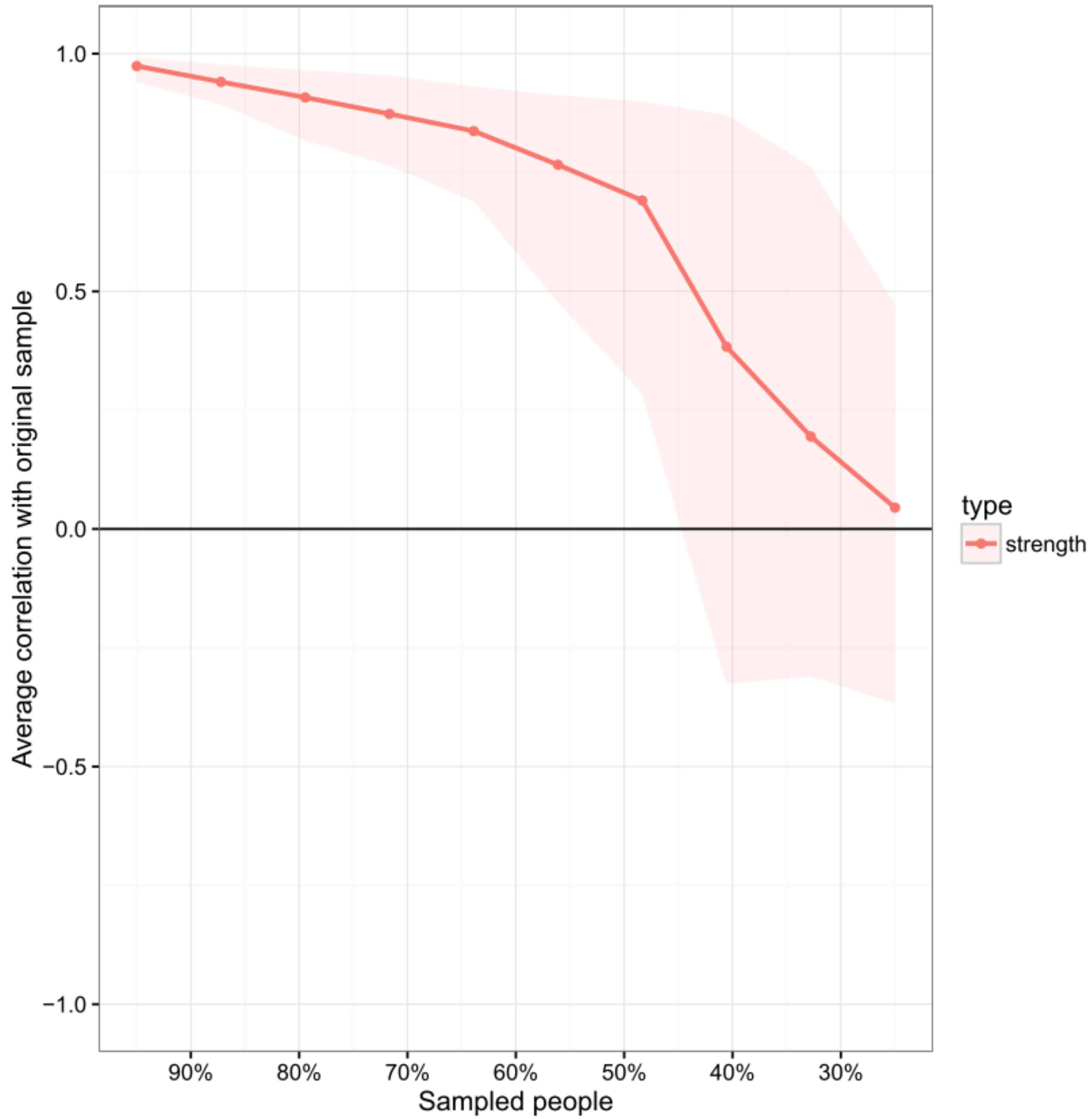
3. Obtain centrality for 90% subset (s17 > s7 > s4...)
4. Subset data 80% of the people
5. Obtain centrality for 80% subset (s16 > s7 > s3...)
6. ...

Subset bootstrap

So what we get is centrality for

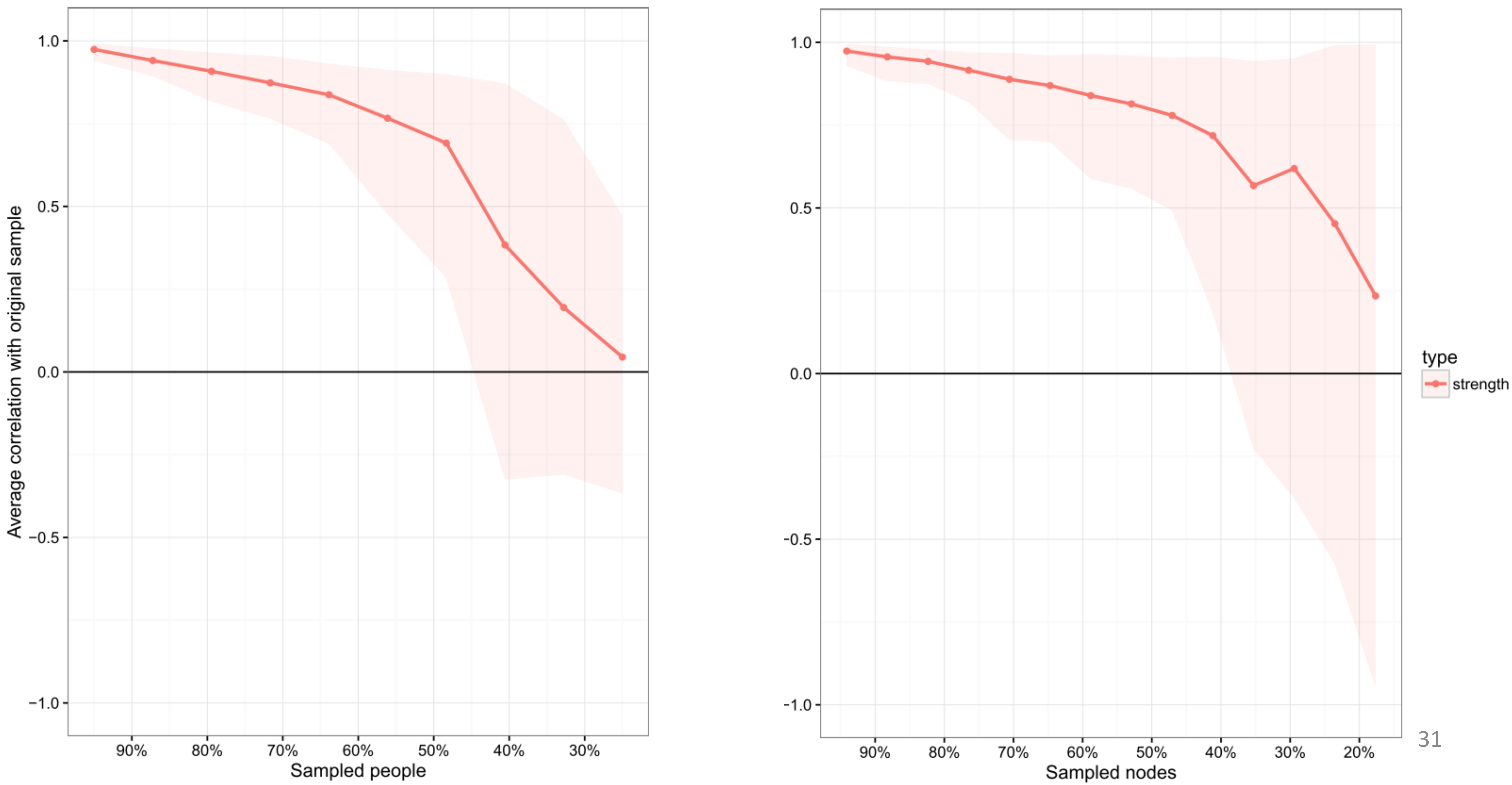
- Full data (s17 > s3 > s16...)
- N -10% (90%) data (s17 > s7 > s4...)
- N -20% (80%) data (s16 > s7 > s3...)
- N -30% (70%) data (s17 > s3 > s16...)
- N -40% (60%) data (s17 > s3 > s16...)
- N -50% (50%) data (s16 > s3 > s7...)
- N -60% (40%) data (s17 > s3 > s7...)
- N -70% (30%) data (s17 > s3 > s16...)
- N -80% (20%) data (s3 > s6 > s17...)
- N -90% (10%) data (s7 > s3 > s16...)

Subset bootstrap



Subset bootstrap

We can also subset nodes instead of people



Take home message

- For most statistical parameters or test statistics, it is very useful to understand how precisely they are estimated
  - Different ways to do that, one way is to bootstrap confidence intervals around the point estimates

Take home message

- For most statistical parameters or test statistics, it is very useful to understand how precisely they are estimated
  - Different ways to do that, one way is to bootstrap confidence intervals around the point estimates
- Investigating the stability of network parameters like edge weights will help us to understand how likely our networks generalize

Take home message

- For most statistical parameters or test statistics, it is very useful to understand how precisely they are estimated
  - Different ways to do that, one way is to bootstrap confidence intervals around the point estimates
- Investigating the stability of network parameters like edge weights will help us to understand how likely our networks generalize
- bootnet is a very first & preliminary step

Network stability

Thanks to Eiko Fried!

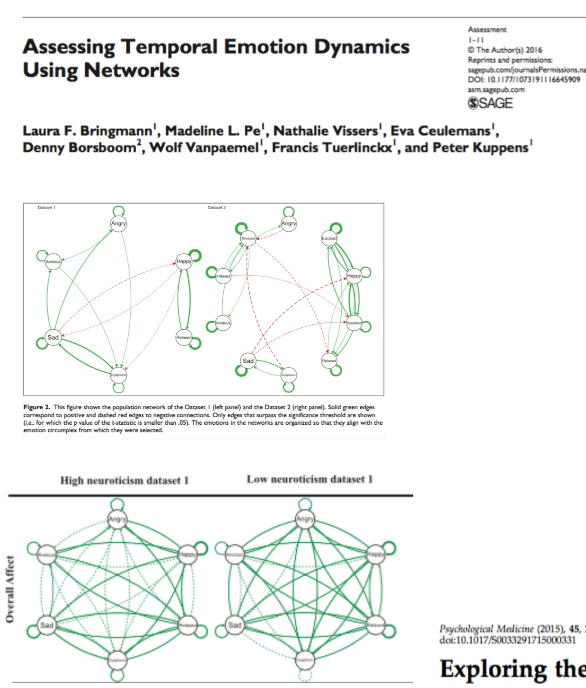
Epskamp, S., Borsboom, D., & Fried, E. I. (2016). Estimating Psychological Networks and their Stability: a Tutorial Paper. arXiv:1604.08462 [stat]. Retrieved from http://arxiv.org/abs/1604.08462

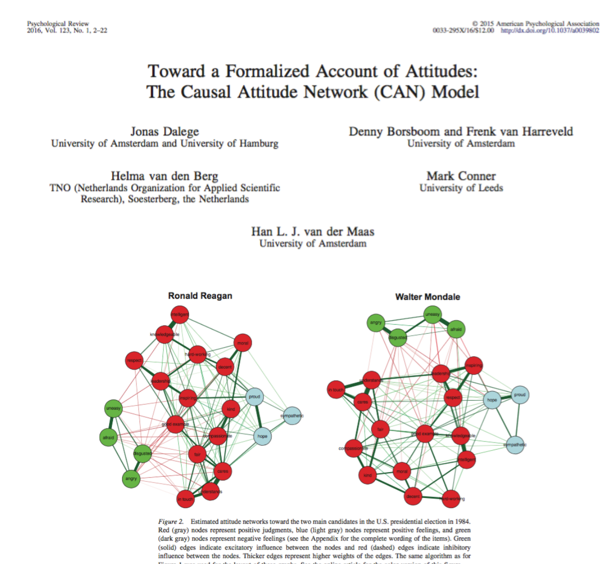
Network comparison

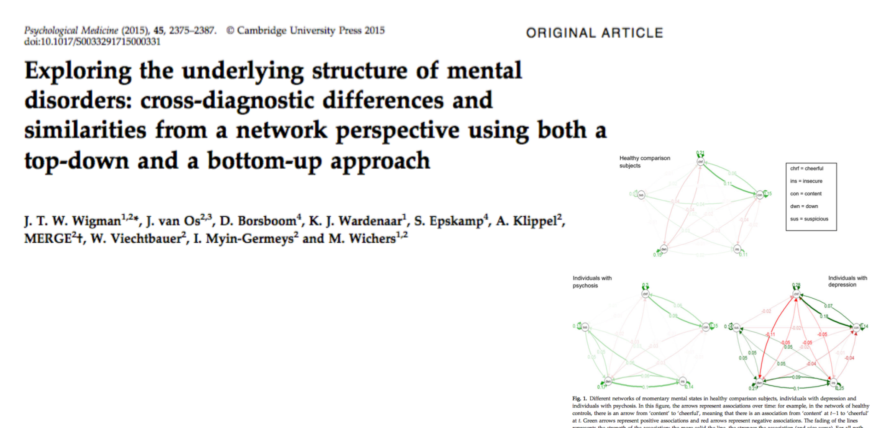
Can you spot the differences?

Network comparison

- Comparing network structures relied mainly on visual inspection
- There was no test to directly statistically assess the difference between two networks

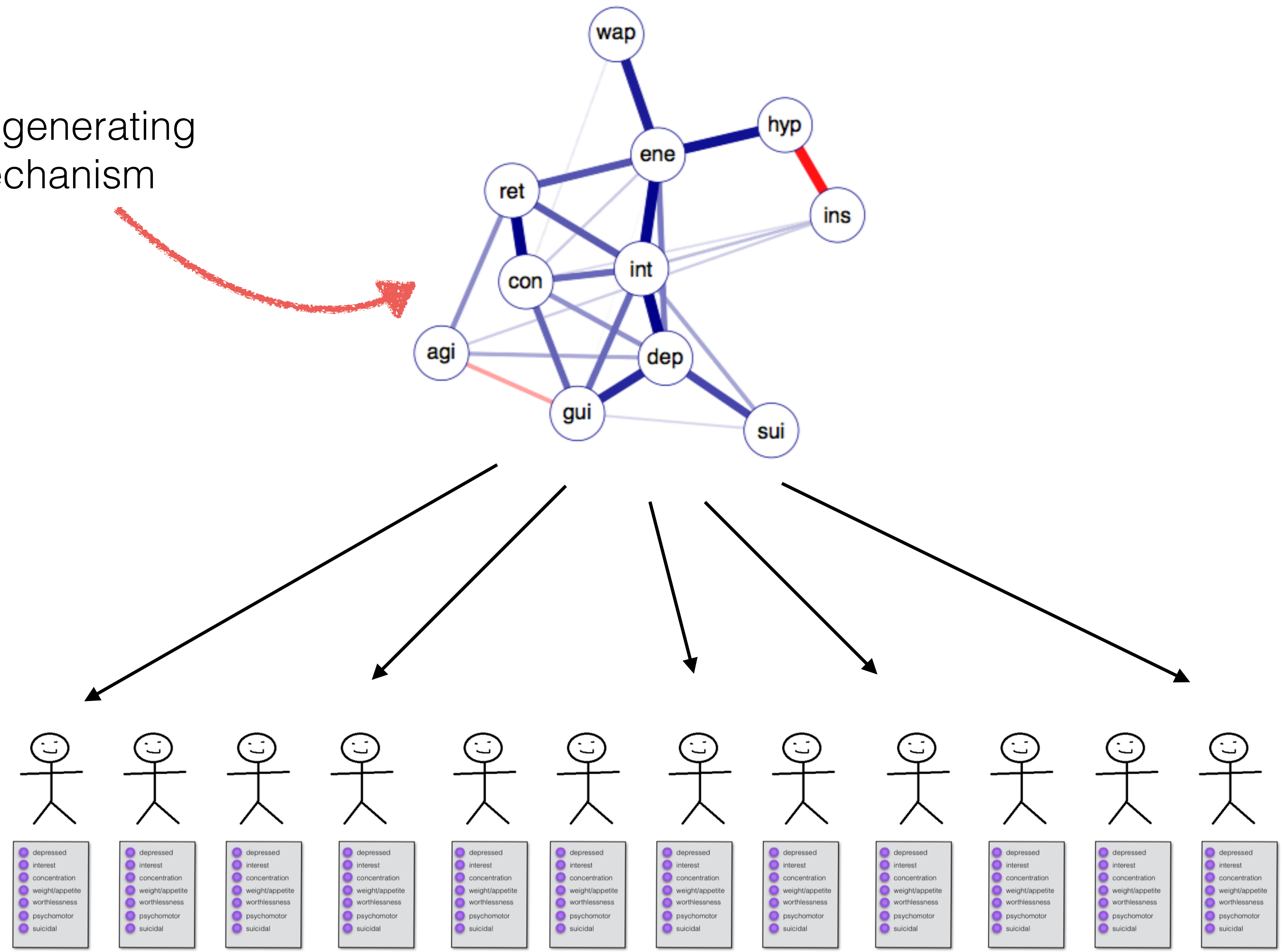






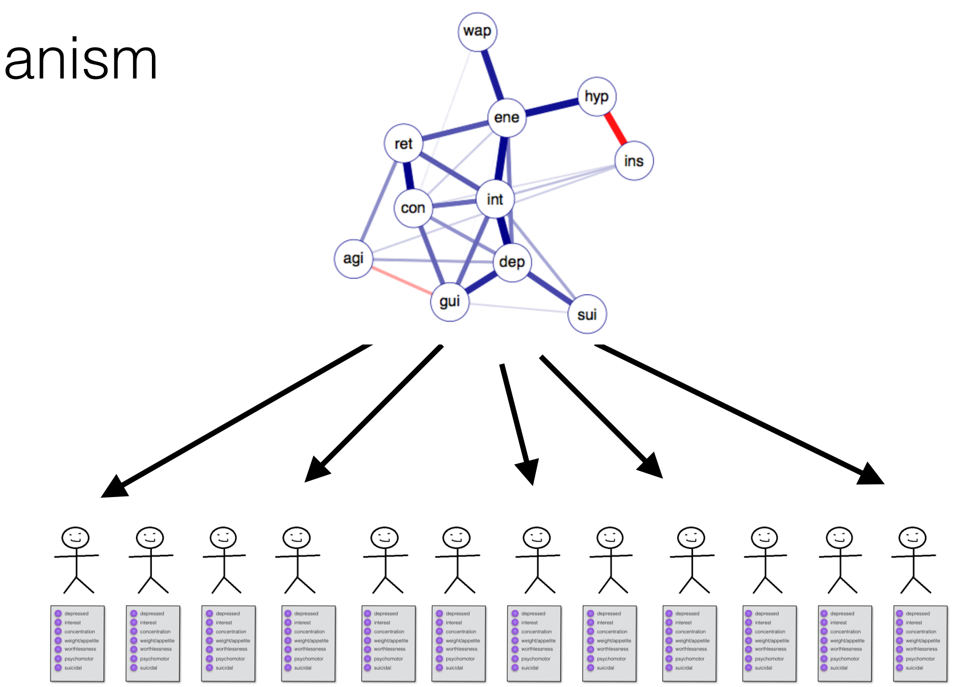
Comparing networks

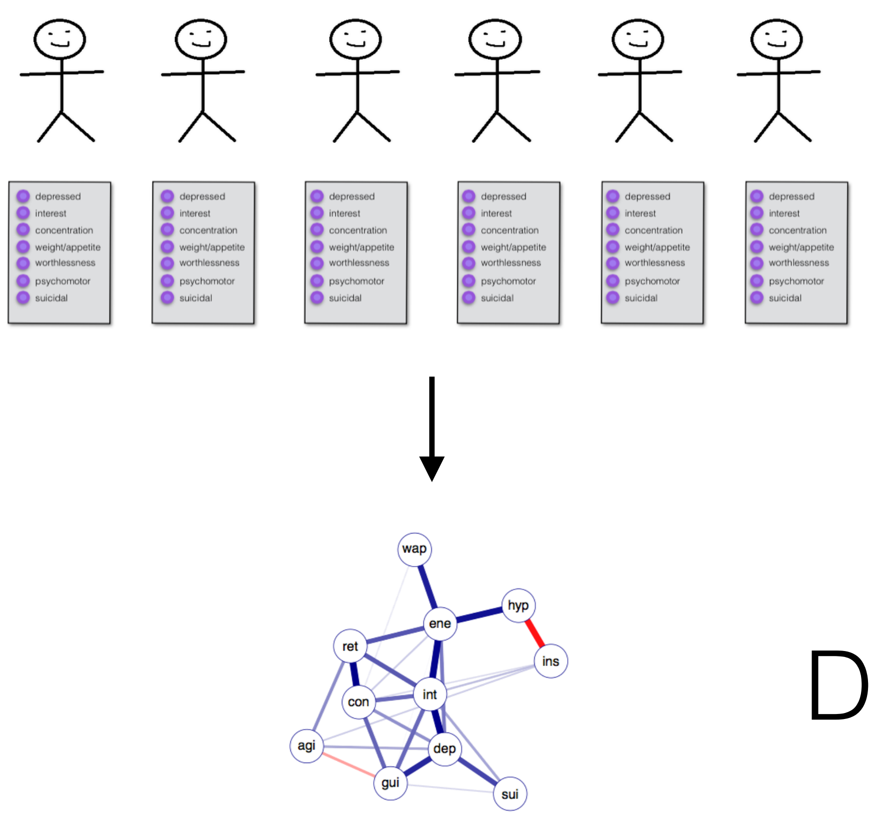
Data generating mechanism



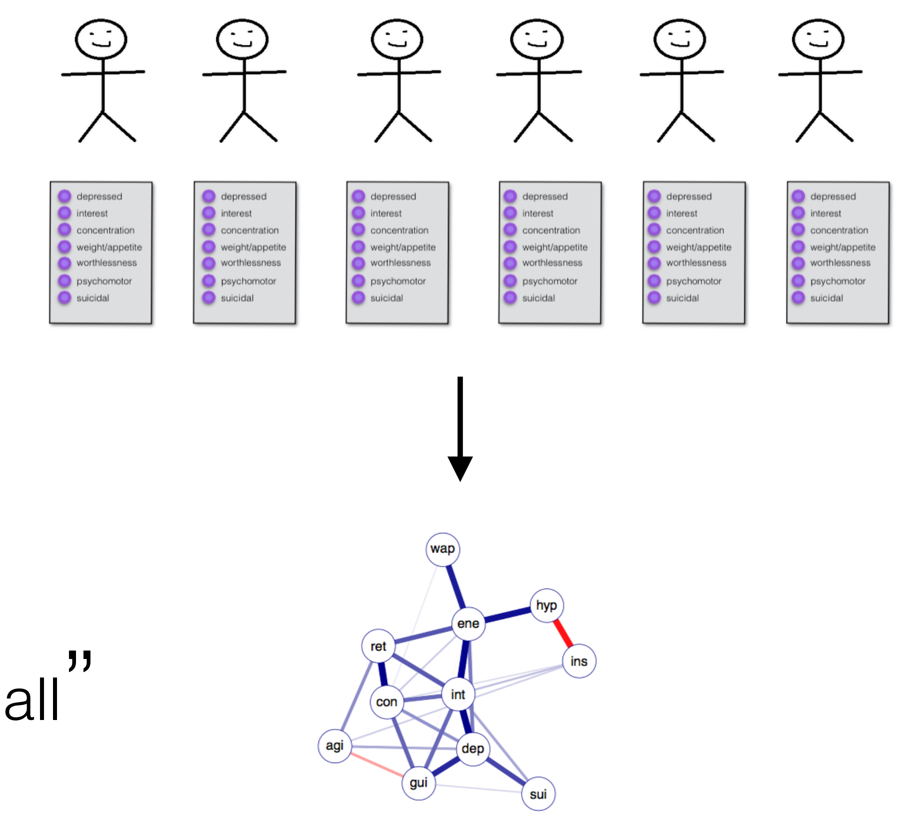
Comparing networks

One data generating mechanism

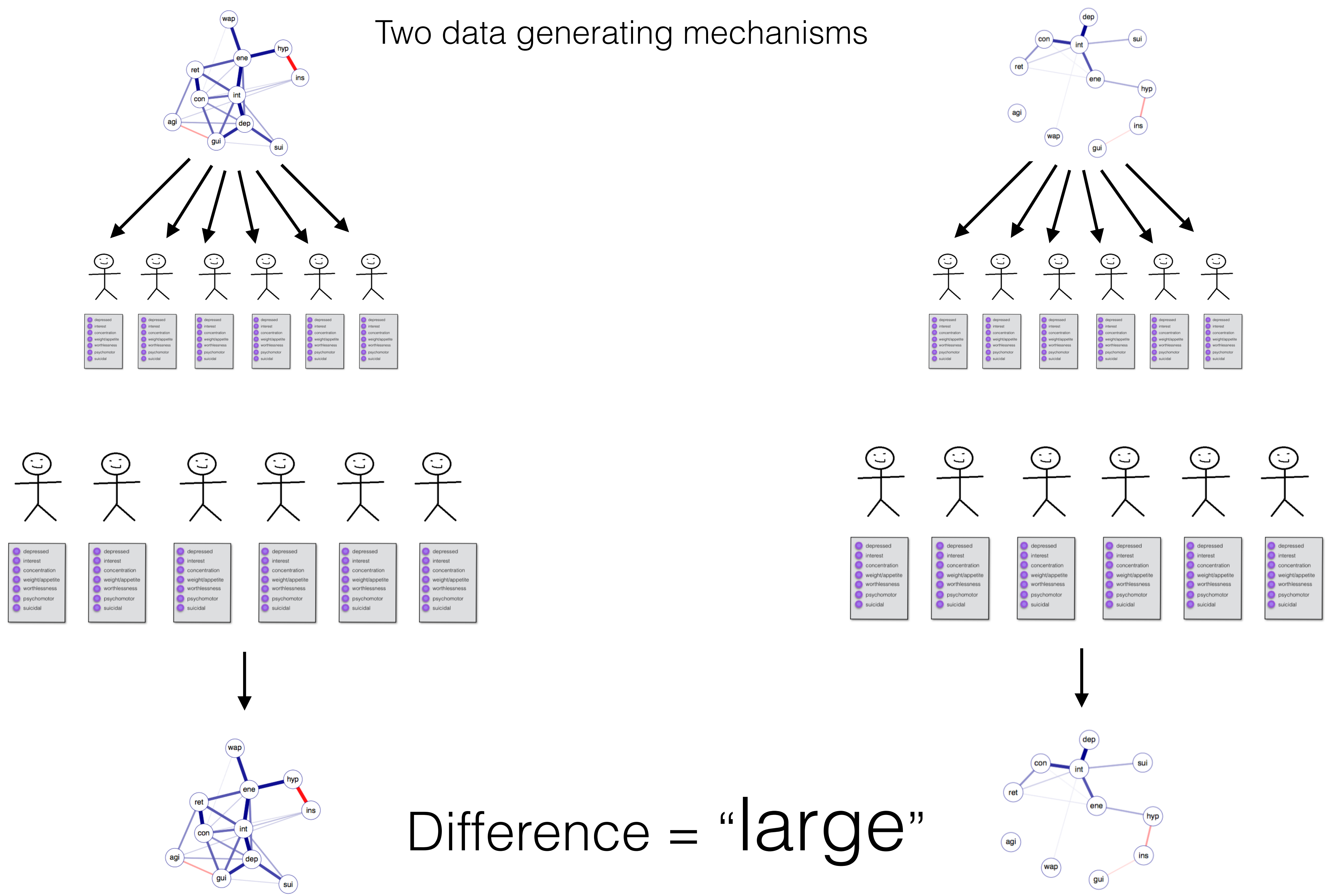




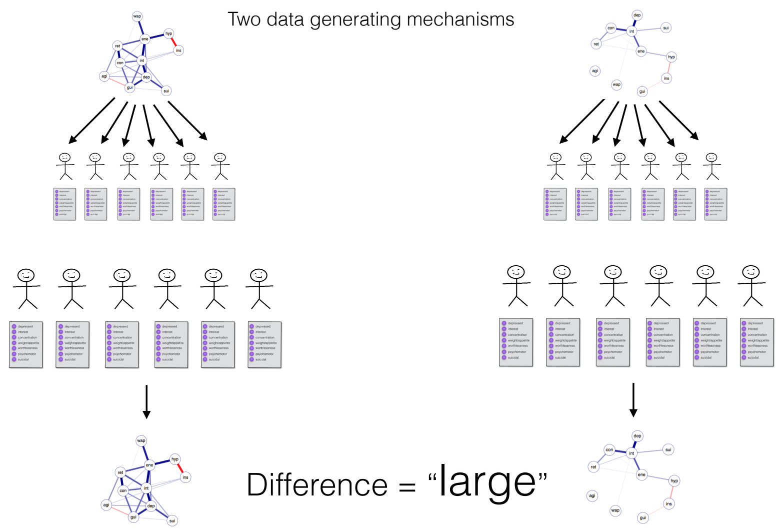
Difference = “small”



Comparing networks

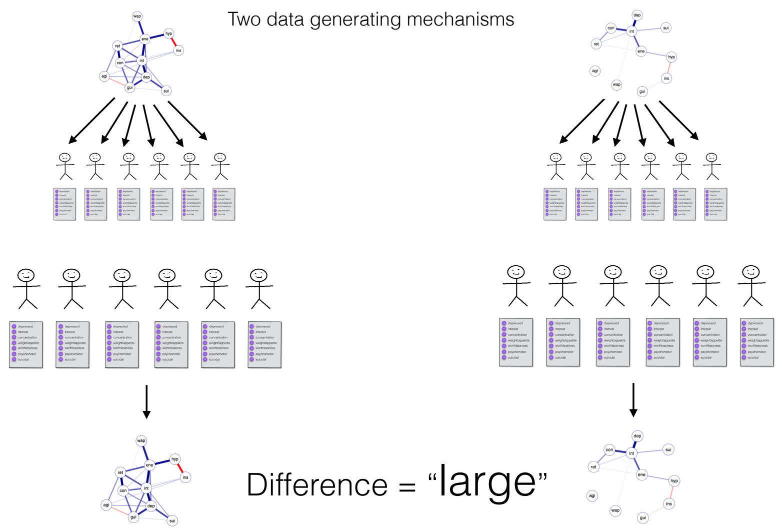


Comparing networks



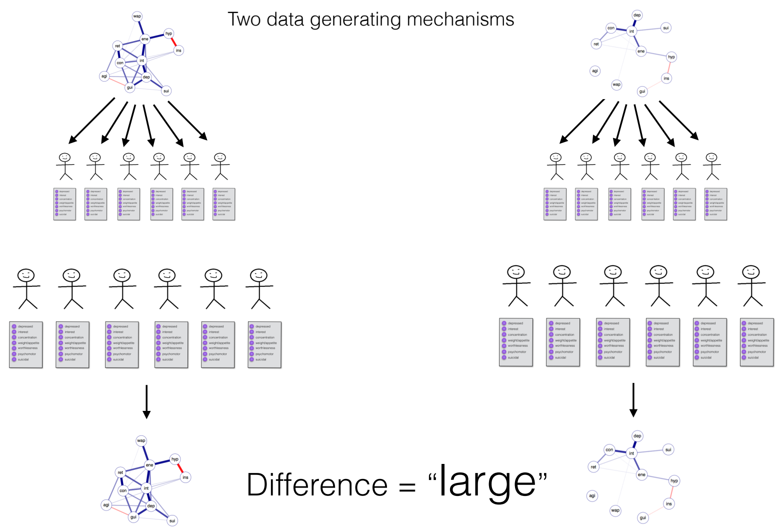
Comparing networks

- When is difference 'large'?



Comparing networks

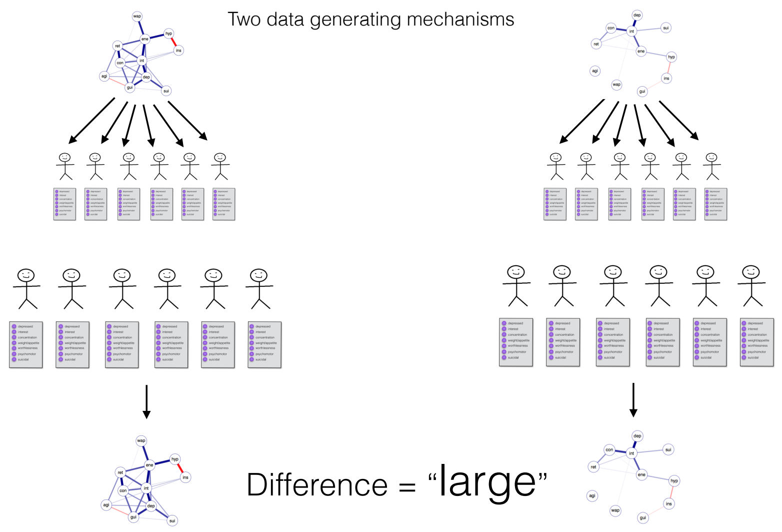
- When is difference ‘large’?
  - When it is larger than you would expect under the null hypothesis



Comparing networks

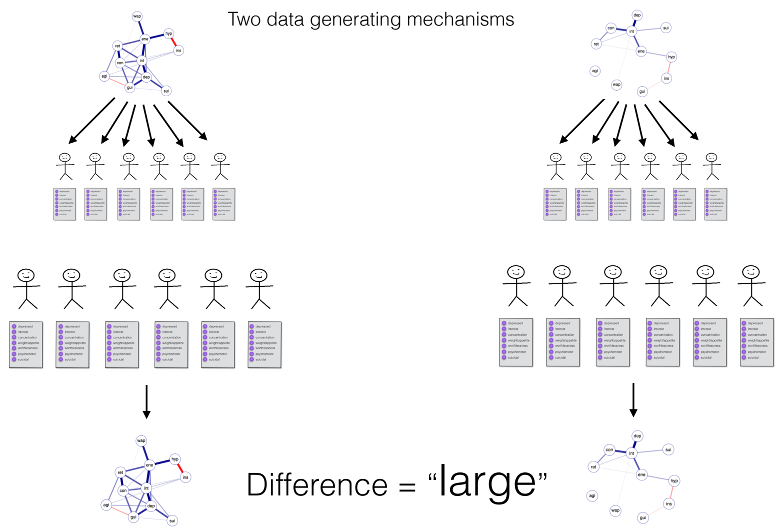
- When is difference ‘large’?
  - When it is larger than you would expect under the null hypothesis

- What is the null hypothesis?



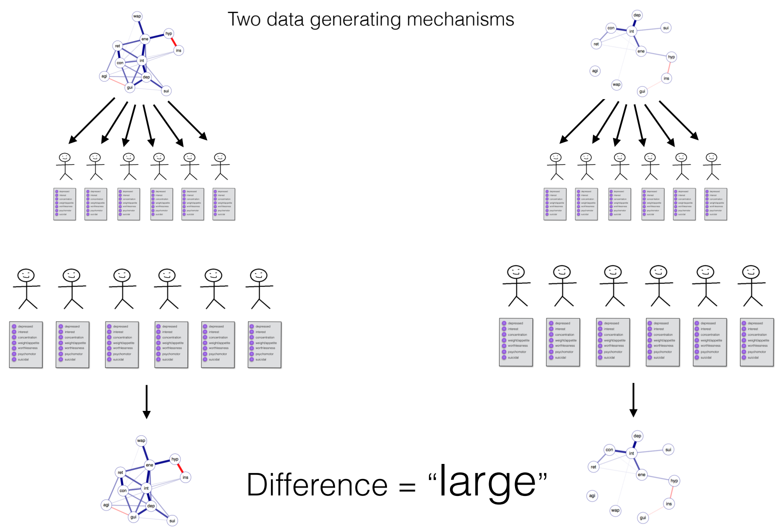
Comparing networks

- When is difference 'large'?
  - When it is larger than you would expect under the null hypothesis
- What is the null hypothesis?
  - All individuals come from the same population (with only one data generating mechanism)



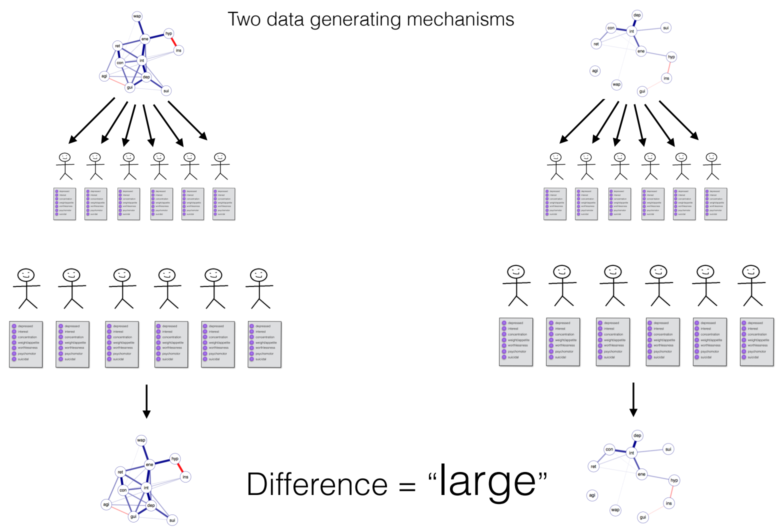
Comparing networks

- When is difference 'large'?
  - When it is larger than you would expect under the null hypothesis
- What is the null hypothesis?
  - All individuals come from the same population (with only one data generating mechanism)
- What do you expect under the null hypothesis?



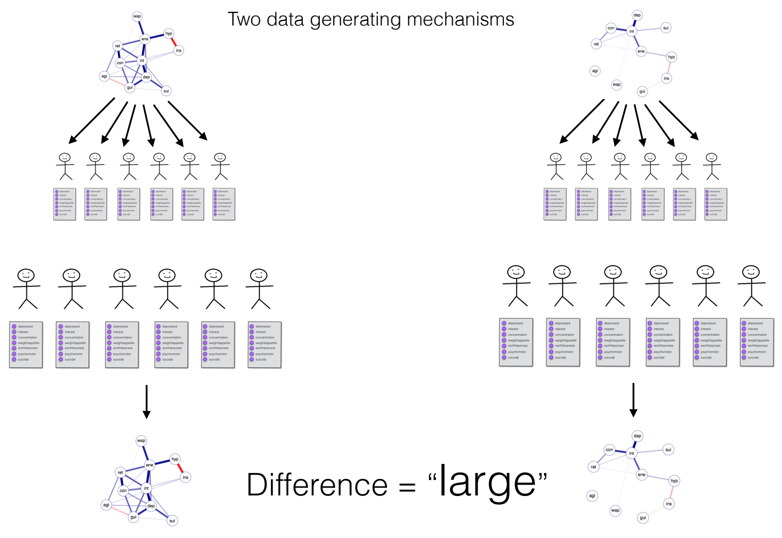
Comparing networks

- When is difference 'large'?
  - When it is larger than you would expect under the null hypothesis
- What is the null hypothesis?
  - All individuals come from the same population (with only one data generating mechanism)
- What do you expect under the null hypothesis?
  - It doesn't matter how individuals are arranged

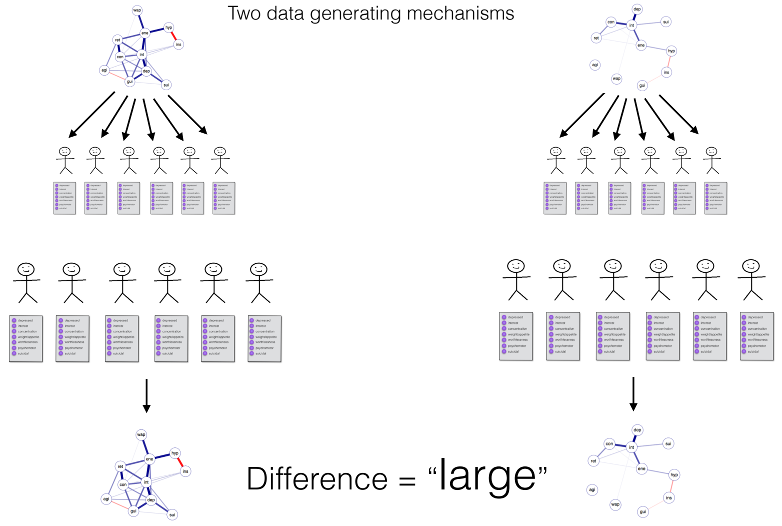


Comparing networks

- When is difference 'large'?
    - When it is larger than you would expect under the null hypothesis
- What is the null hypothesis?
    - All individuals come from the same population (with only one data generating mechanism)
- What do you expect under the null hypothesis?
    - It doesn't matter how individuals are arranged
- Let's see what happens if we repeatedly (randomly) rearrange individuals and calculate the 'difference'

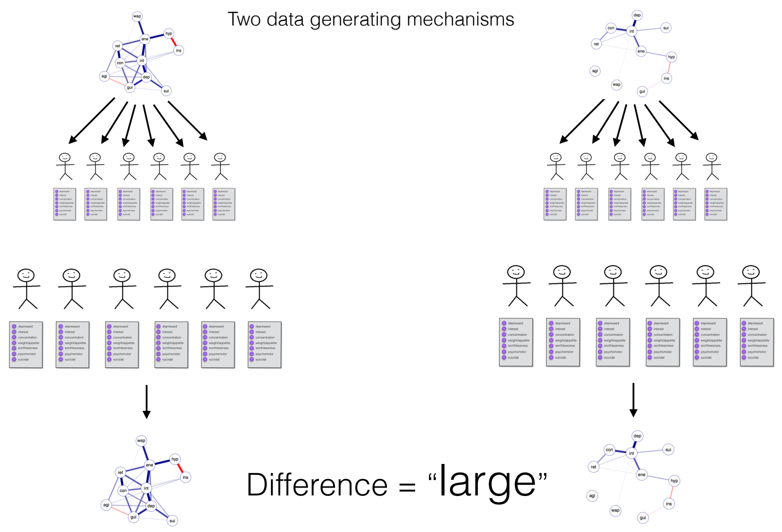


Comparing networks



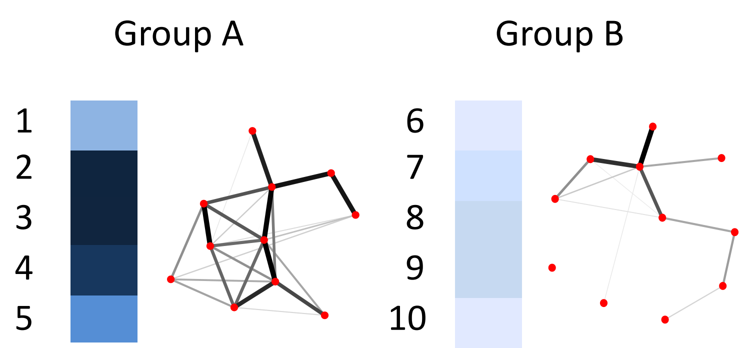
Comparing networks

- Life example?



Network Comparison Test

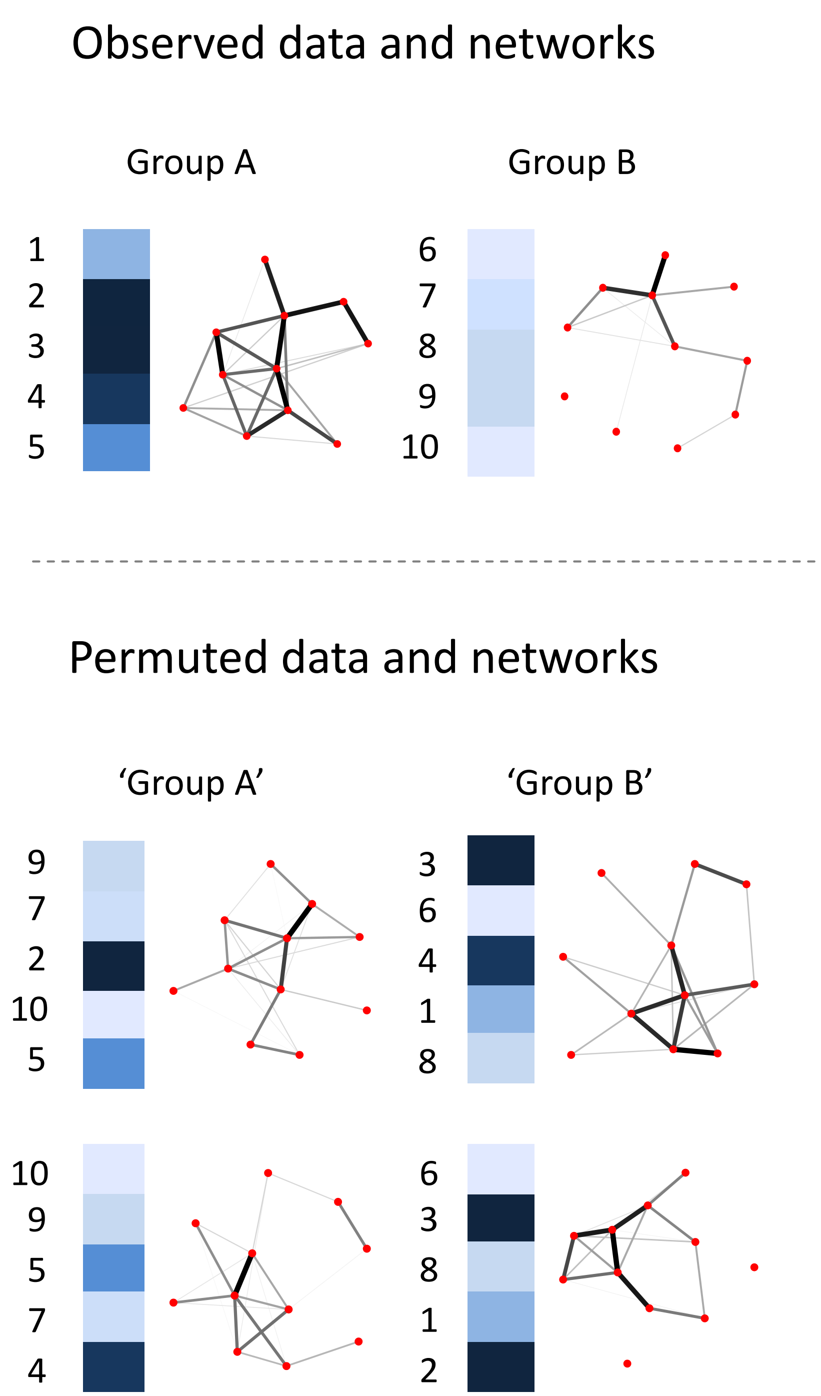
Observed data and networks



▲ observed difference

- NCT is implemented in **`R`**
- Currently suited for binary and continuous data
- Networks are estimated with **`IsingFit`** (binary data; Van Borkulo et al., 2014) or with **`EBICglasso`** (continous data; Epskamp et al., 2012)

Network Comparison Test

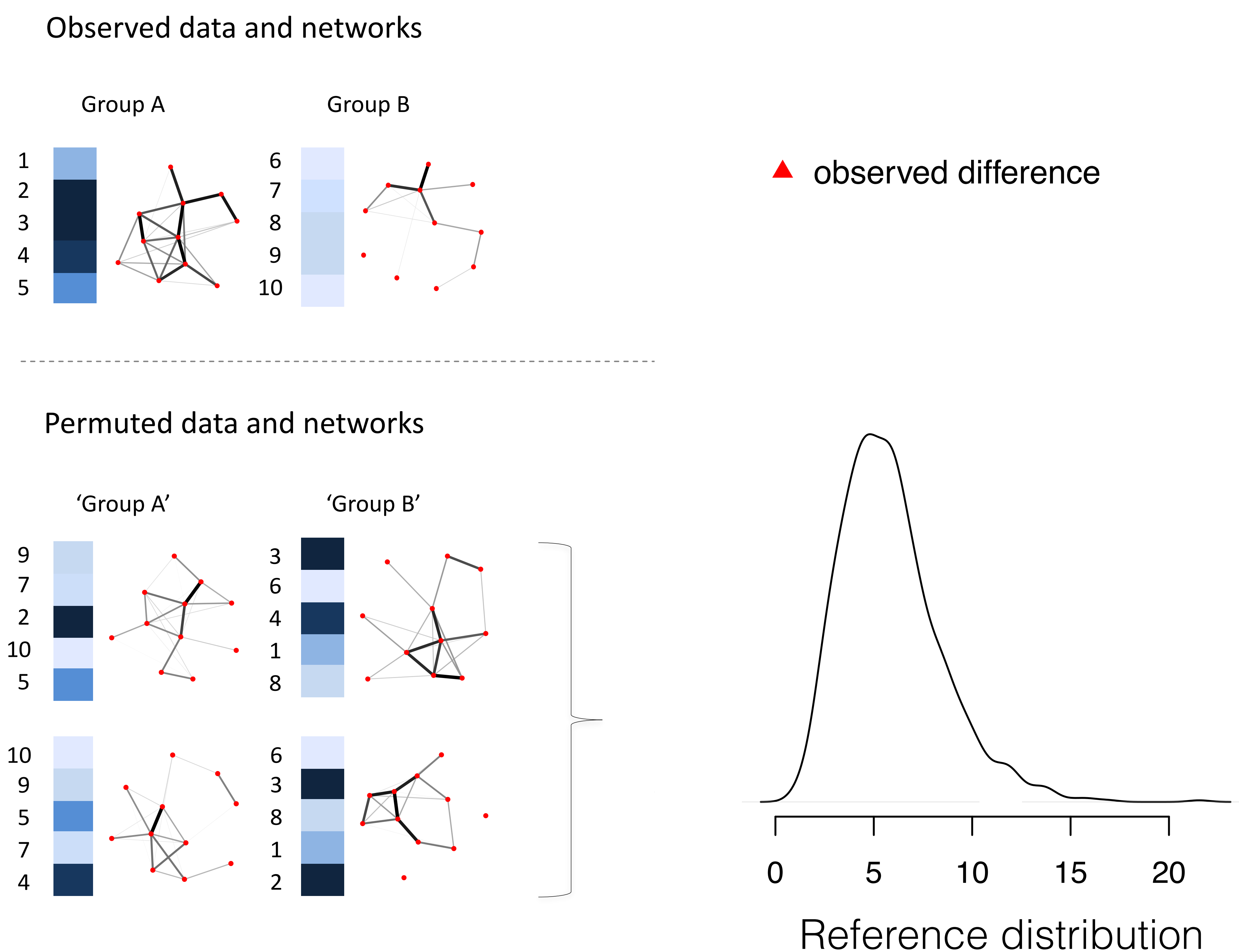


▲ observed difference

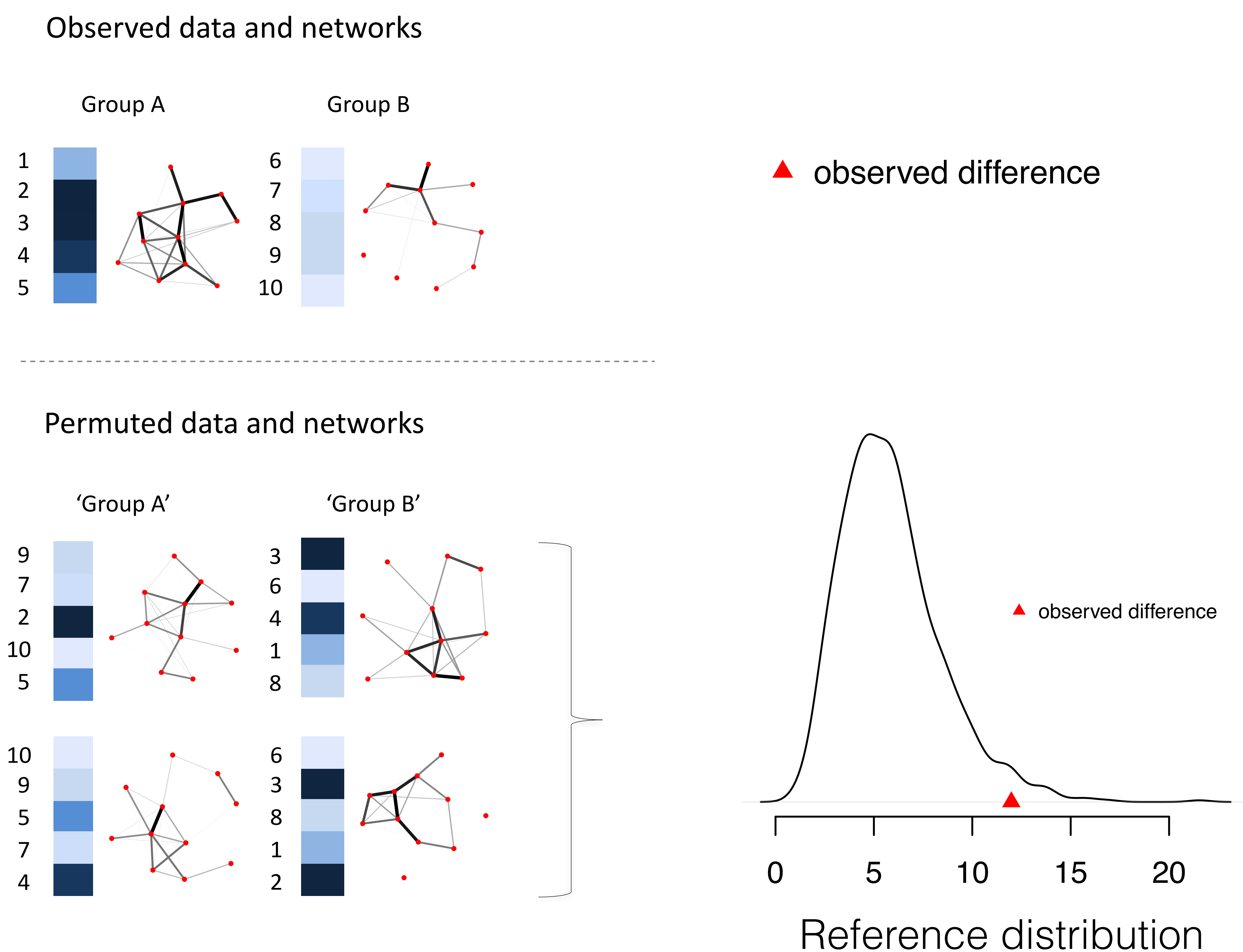
Null hypothesis

- It doesn't matter how you rearrange individuals
- Each individual stems from the same population

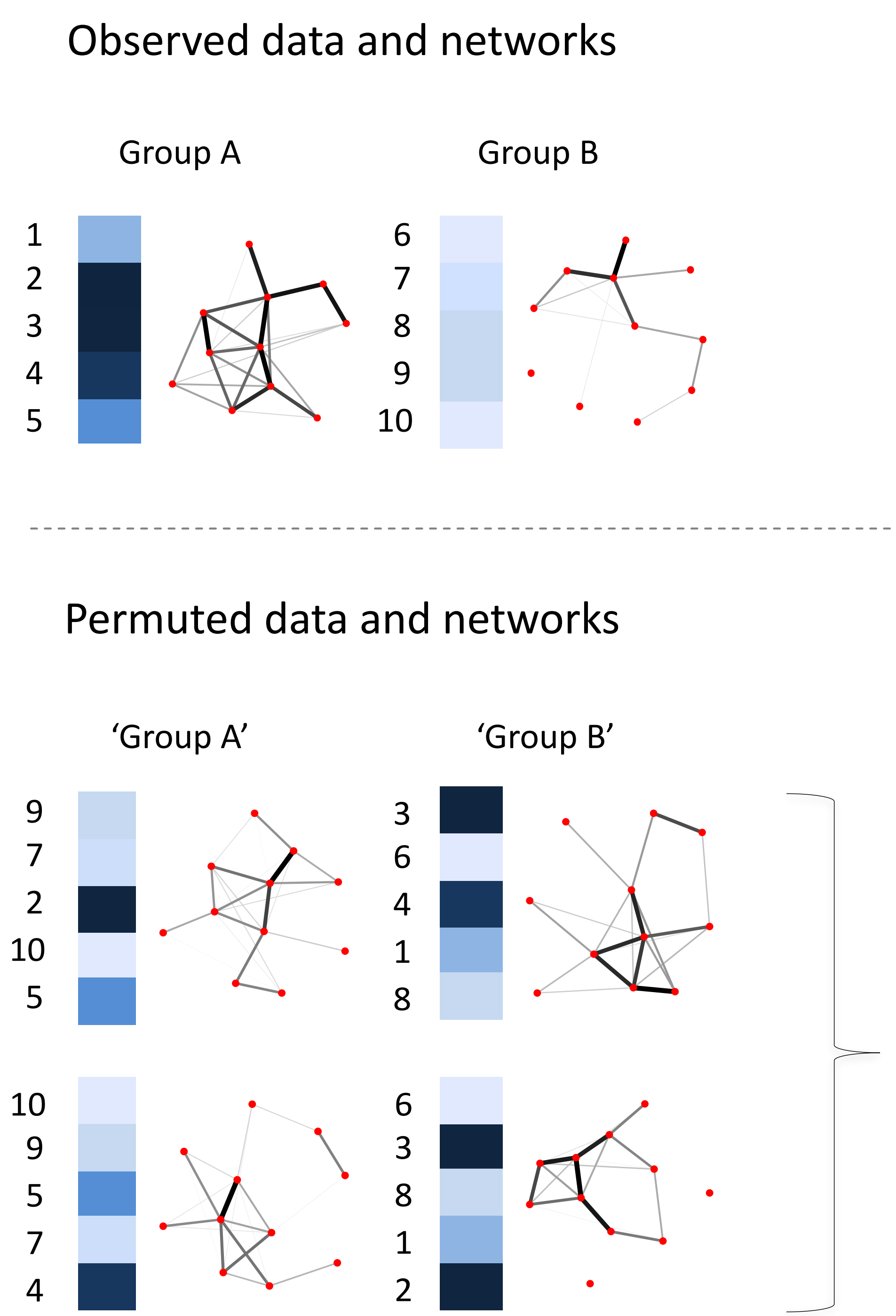
Network Comparison Test



Network Comparison Test

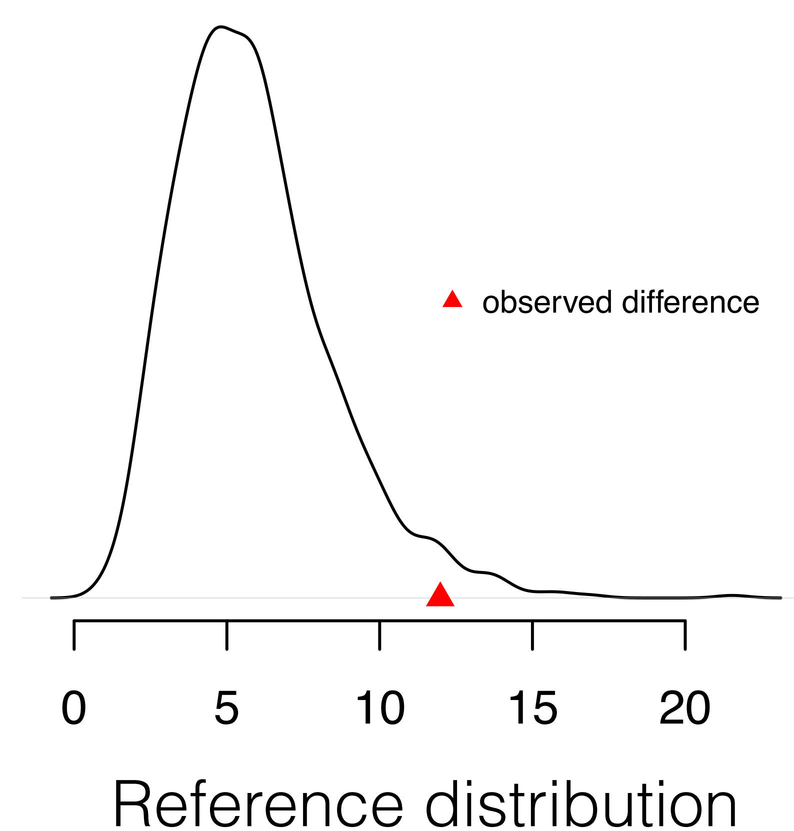


Network Comparison Test



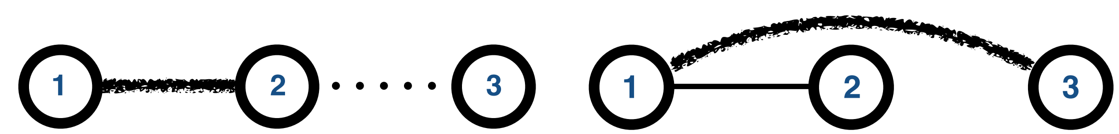
Tests on three invariance hypotheses:

1. Network structure
2. Global strength
3. Edge strength



Network Comparison Test

1. Network structure invariance hypothesis



- structure is completely identical across subpopulations
- distance measure (M) is based on the maximum or L_∞ norm
- similar to testing whether two distributions are similar (as in the Kolmogorov–Smirnov test)

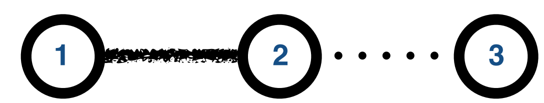
-	0.1	0.5
-	-	0.1
-	-	-

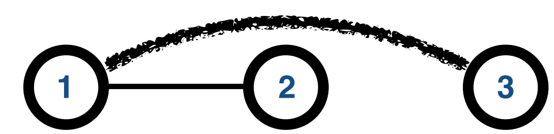
$$D_{ij} = |A_{1ij} - A_{2ij}|$$

$$M(G_1,G_2) = max\ (D_{ij})$$

Network Comparison Test

2. Global strength invariance hypothesis





- overall level of connectivity is identical across subpopulations
- distance measure (S) is based on difference in global strength

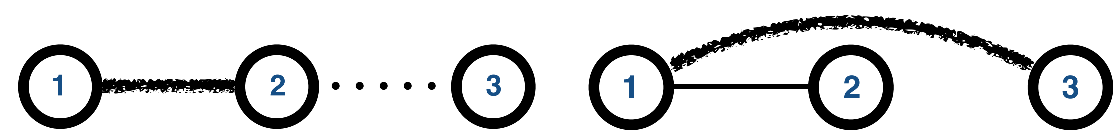
-	0.3	0
-	-	0.1
-	-	-

-	0.2	0.5
-	-	0
-	-	-

$$S(G_1, G_2) = |(\sum|A_{1ij}| - \sum|A_{2ij}|)|$$

Network Comparison Test

3. Edge strength invariance hypothesis



- a specific edge is identical across subpopulations
- distance measure (E) is based on difference in connection strength

-	0.3	0
-	-	0.1
-	-	-

-	0.2	0.5
-	-	0
-	-	-

$$D_{ij} = |A_{1ij} - A_{2ij}|$$

$$E(\beta_{ij}^{G1}, \beta_{ij}^{G2}) = |\beta_{ij}^{G1} - \beta_{ij}^{G2}|$$

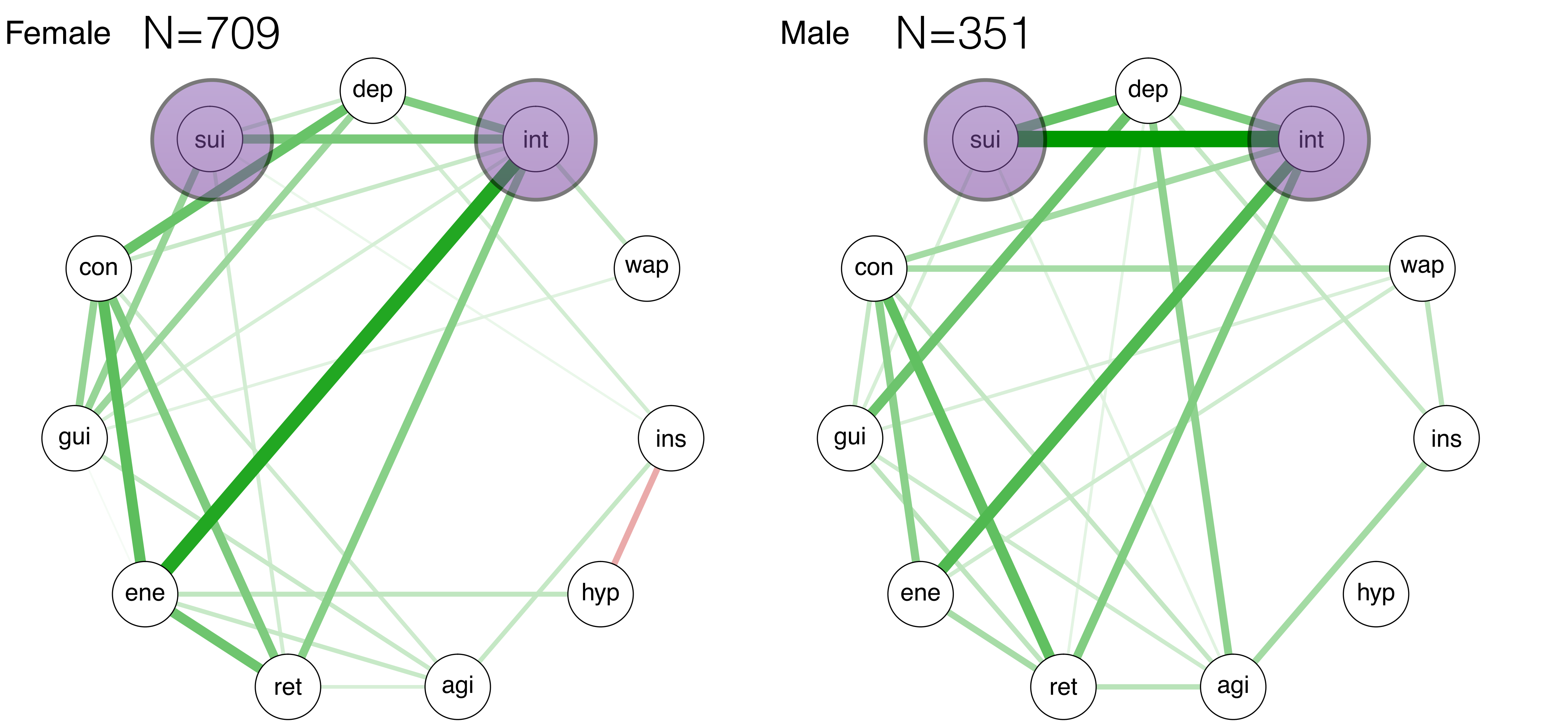
β_{ij}: a particular edge

Network Comparison Test

1. Network structure invariance hypothesis
    - structure is completely identical across subpopulations
    - distance measure (M) is based on the maximum or L_∞ norm
2. Global strength invariance hypothesis
    - overall level of connectivity is identical across subpopulations
    - distance measure (S) is based on difference in global strength
3. Edge strength invariance hypothesis
    - a specific edge is identical across subpopulations
    - distance measure (E) is based on difference in connection strength

Real data

NCT(datamen,datawomen, binary.data=TRUE, it=1000, test.edges=TRUE, edges=list(c(2,11)))



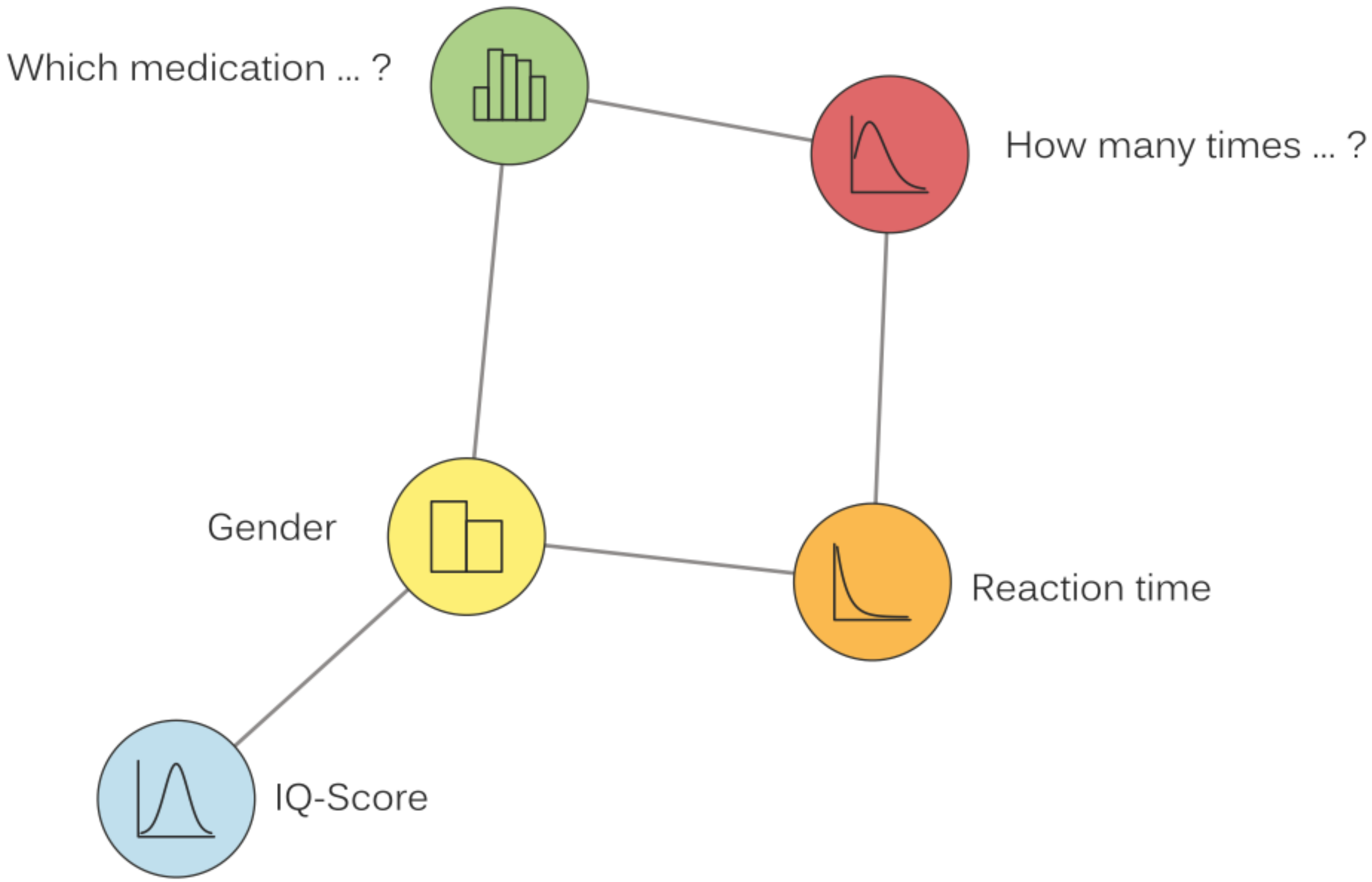
Netherlands Study of Depression and Anxiety (NESDA; Penninx et al., 2008)

Real data

Summary

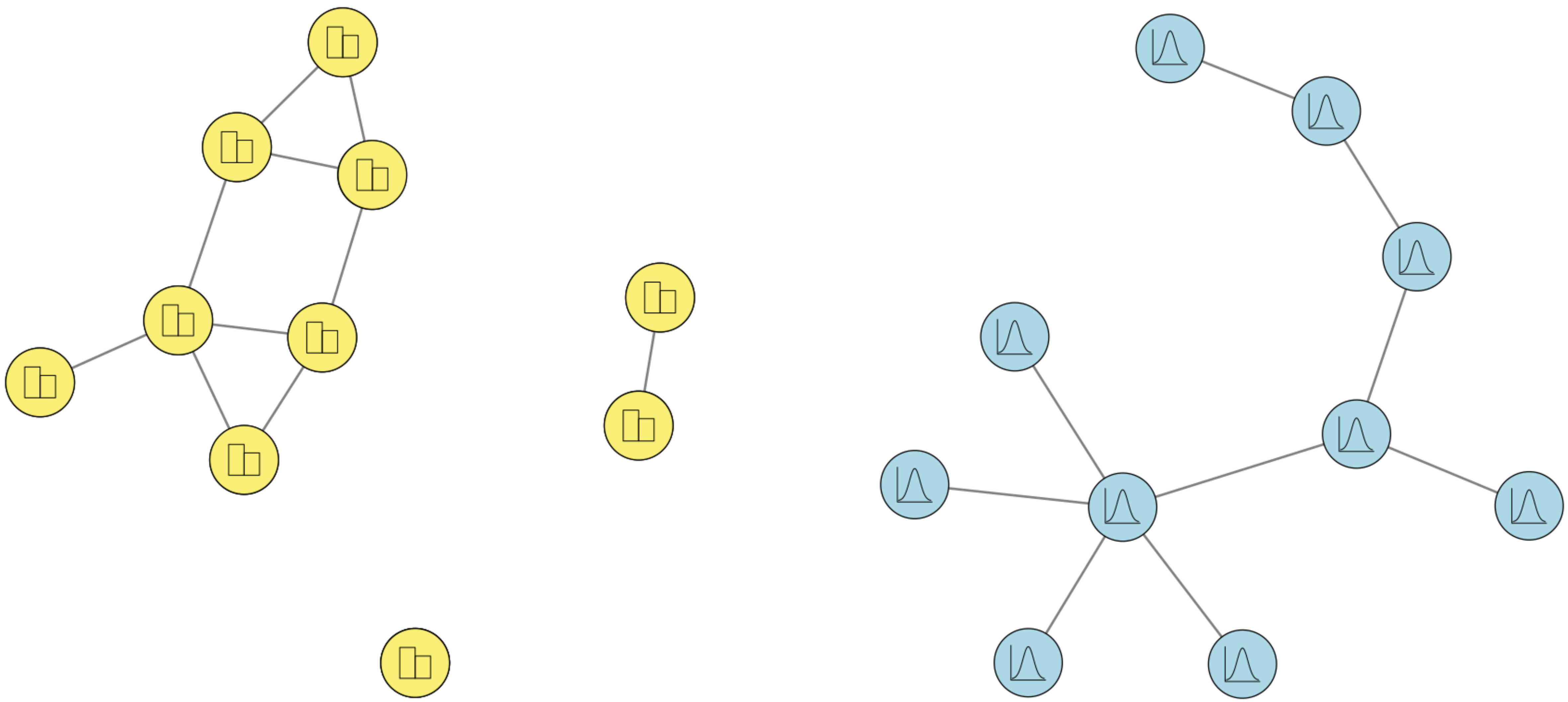
- Network structure: no difference ($p=.251$)
- Global strength: no difference ($p=.909$)
- Edge strength: sui-int significantly stronger in network of males ($p=.027$)

Mixed graphical models



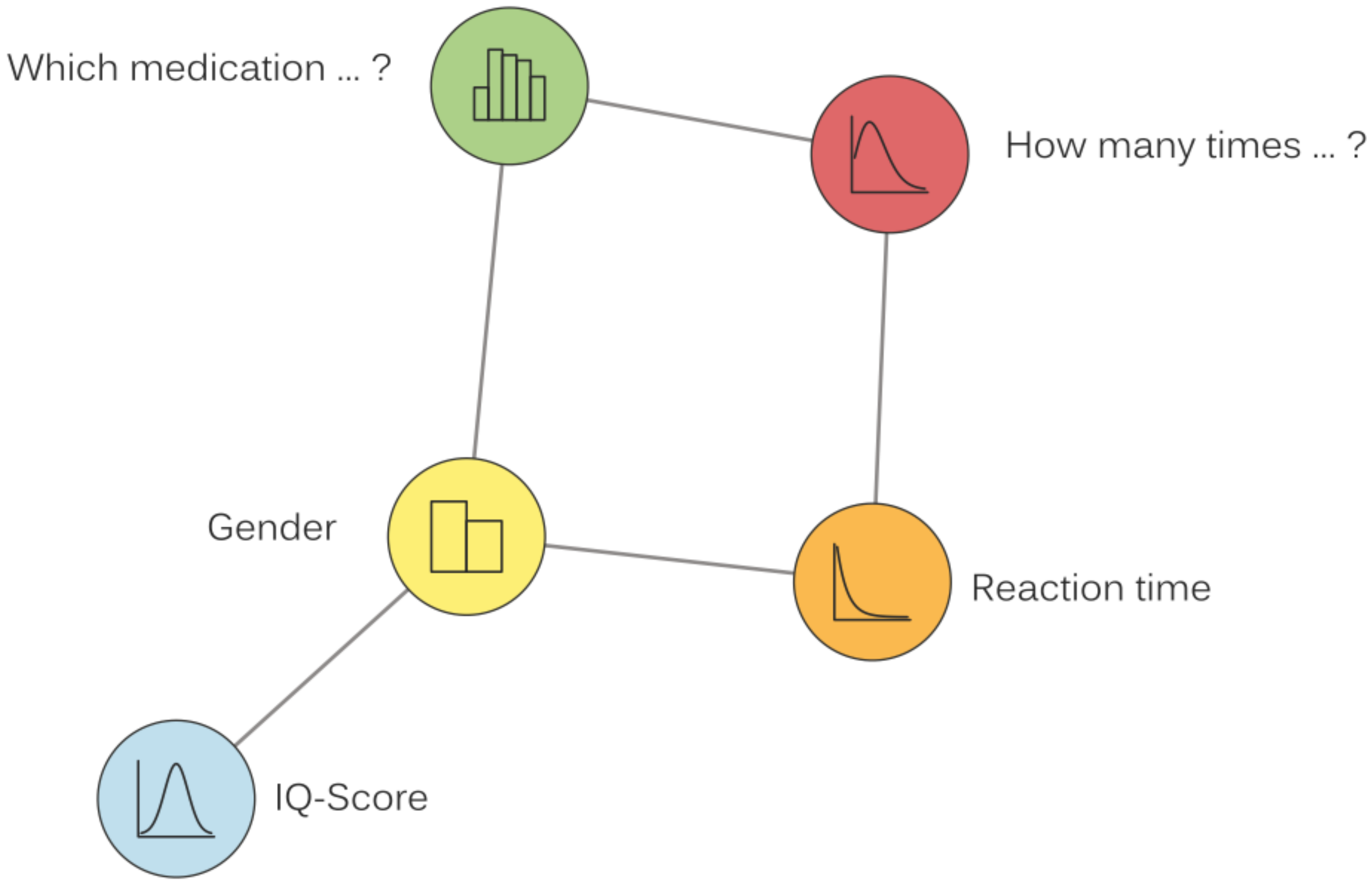
Mixed graphical models

We currently fit networks with either binary or gaussian data



Mixed graphical models

But psychological data are often mixed



MGM

- Novel *R* package **mgm** (mixed graphical models) allows us to fix mixed data
  - By Jonas Haslbeck, UvA, http://jmbh.github.io/papers_software
- Further reading:
  - Post 1: http://jmbh.github.io/Estimation-of-mixed-graphical-models/
  - Post 2: http://jmbh.github.io/Interactions-between-categorical-Variables-in-mixed-graphical-models/
  - Papers on Jonas' homepage

MGM

Call packages and get example

```
library("mgm")  # Estimate mixed graphical models
library("httr") # For GET() function to download data

url='http://jmbh.github.io/figs/efpsa_workshop/
    autism_datalist.RDS' GET(url, write_disk
    "autism_datalist.RDS", overwrite=TRUE))

Autism_data <- readRDS('autism_datalist.RDS')
```

MGM

```
Autism_data$colnames # variable names
Autism_data$type     # define variable types
Autism_data$lev      # define variables level
```

```
> Autism_data$colnames       # variable names
 [1] "Gender"                              "IQ"
 [3] "Age diagnosis"                       "Openness about Diagnosis"
 [5] "Success selfrating"                  "Well being"
 [7] "Integration in Society"              "No of family members with autism"
 [9] "No of Comorbidities"                 "No of Physical Problems"
[11] "No of Treatments"                    "No of Medications"
[13] "No of Care Units"                    "Type of Housing"
[15] "No of unfinished Educations"         "Type of work"
[17] "Workinghours"                        "No of Interests"
[19] "No of Social Contacts"               "Good Characteristics due to Autism"
[21] "No of Transition Problems"           "Satisfaction: Treatment"
[23] "Satisfaction: Medication"            "Satisfaction: Care"
[25] "Satisfaction: Education"             "Satisfaction: Work"
[27] "Satisfaction: Social Contacts"       "Age"
```

```
> Autism_data$type           # define variable types
 [1] "c" "g" "g" "c" "g" "c" "c" "p" "p" "p" "p" "p" "p" "c" "p" "c" "g" "p" "p" "p" "p" "g" "g" "g"
[25] "g" "g" "c" "g"
```

```
> Autism_data$lev            # define variables level
 [1] 2 1 1 2 1 5 3 1 1 1 1 1 1 2 1 4 1 1 1 1 1 1 1 1 1 1 3 1
```

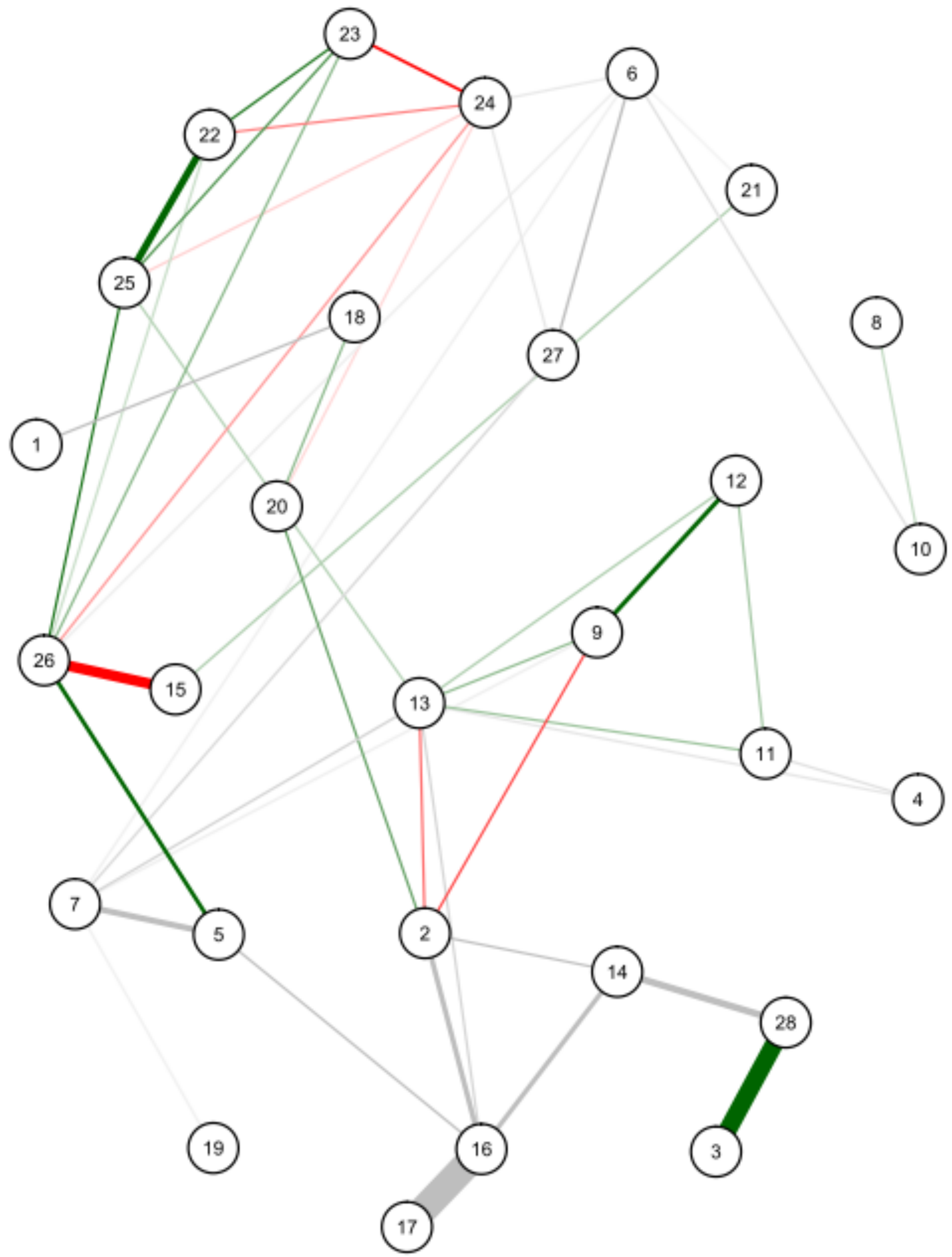
MGM

Estimate & visualize network

```
fitMGM <- mgmfit(Autism_data$data, Autism_data$type,
Autism_data$lev, d=2)

qgraph(fitMGM$wadj, nodeNames=Autism_data$colnames,
layout='spring', edge.color=fitMGM$edgecolor, legend.cex=.
3, vsize=3, legend.cex=1)
```

MGM



1: Gender
2: IQ
3: Age diagnosis
4: Openness about Diagnosis
5: Success selfrating
6: Well being
7: Integration in Society
8: No of family members with autism
9: No of Comorbidities
10: No of Physical Problems
11: No of Treatments
12: No of Medications
13: No of Care Units
14: Type of Housing
15: No of unfinished Educations
16: Type of work
17: Workinghours
18: No of Interests
19: No of Social Contacts
20: Good Characteristics due to Autism
21: No of Transition Problems
22: Satisfaction: Treatment
23: Satisfaction: Medication
24: Satisfaction: Care
25: Satisfaction: Education
26: Satisfaction: Work
27: Satisfaction: Social Contacts
28: Age

Practical

- Open Assignment_Day3_Part**3**.pdf
- Just follow the steps!
- If you go through the exercises quickly, you can try stability analysis or network comparison on your own data (if you data is cross-sectional)

Recap of Day 3

Today you have learned

- about conditional (in)dependence

  when a node is connected to another node it means that they are still associated after controlling for all other variables

- about advanced network estimation

  regularization to find optimal balance between parsimony and goodness of fit

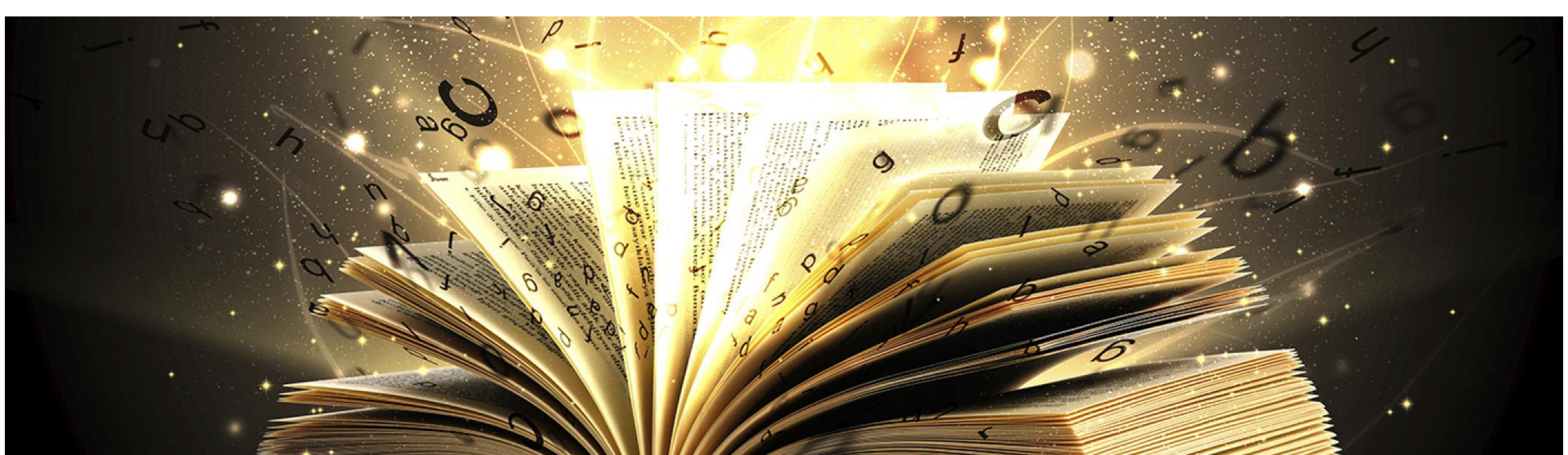
- about network stability, network comparison, mix graphical models

Literature

About network stability:
Epskamp, S., Borsboom, D., & Fried, E. I. (2016). Estimating Psychological Networks and their Stability: a Tutorial Paper. arXiv:1604.08462 [stat]. Retrieved from http://arxiv.org/abs/1604.08462

A paper that applied NCT:
van Borkulo, C., Boschloo, L., Borsboom, D., Penninx, B. W. J. H., Waldorp, L. J., & Schoevers, R. A. (2015). Association of Symptom Network Structure With the Course of Depression. *JAMA Psychiatry, 72*(12), 1219. http://doi.org/10.1001/jamapsychiatry.2015.2079

Submitted paper on performance of NCT:
van Borkulo, C. D., Waldorp, L. J., Boschloo, L., Kossakowski, J., Tio, P., Schoevers, R. A., & Borsboom, D. *Distinguishing between networks on three aspects: A permutation test*.



Contact information

cvborkulo@gmail.com

https://cvborkulo.com

https://www.researchgate.net/profile/Claudia_Van_Borkulo

https://nl.linkedin.com/in/claudia-van-borkulo-4b227a70

Let’s call it a day!