BERD Academy

A Connected World

Data Analysis for Real World Network Data

ERGM

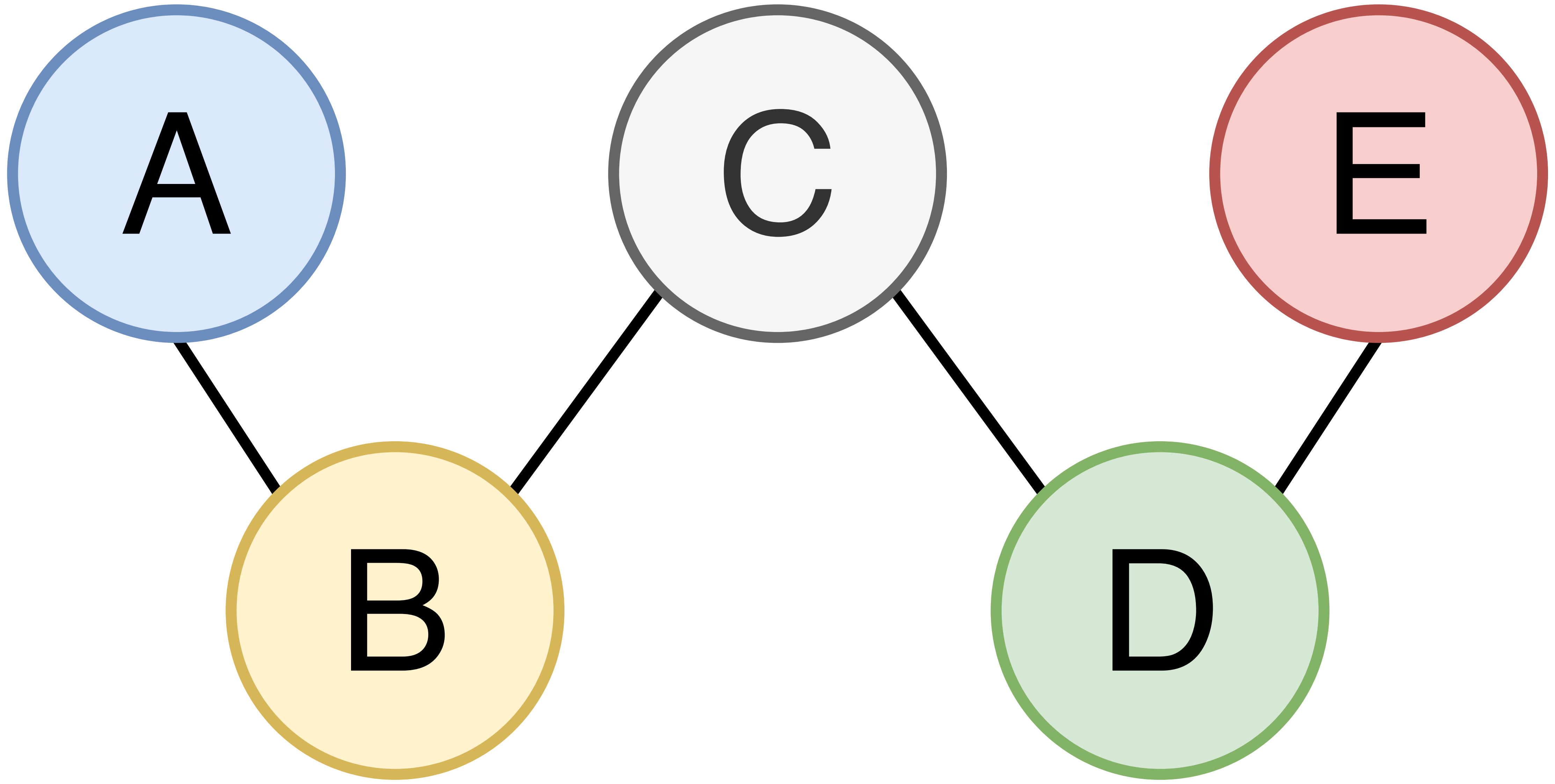
08.12.2022

Motivation

Modeling Networks

Why do we want to study networks?

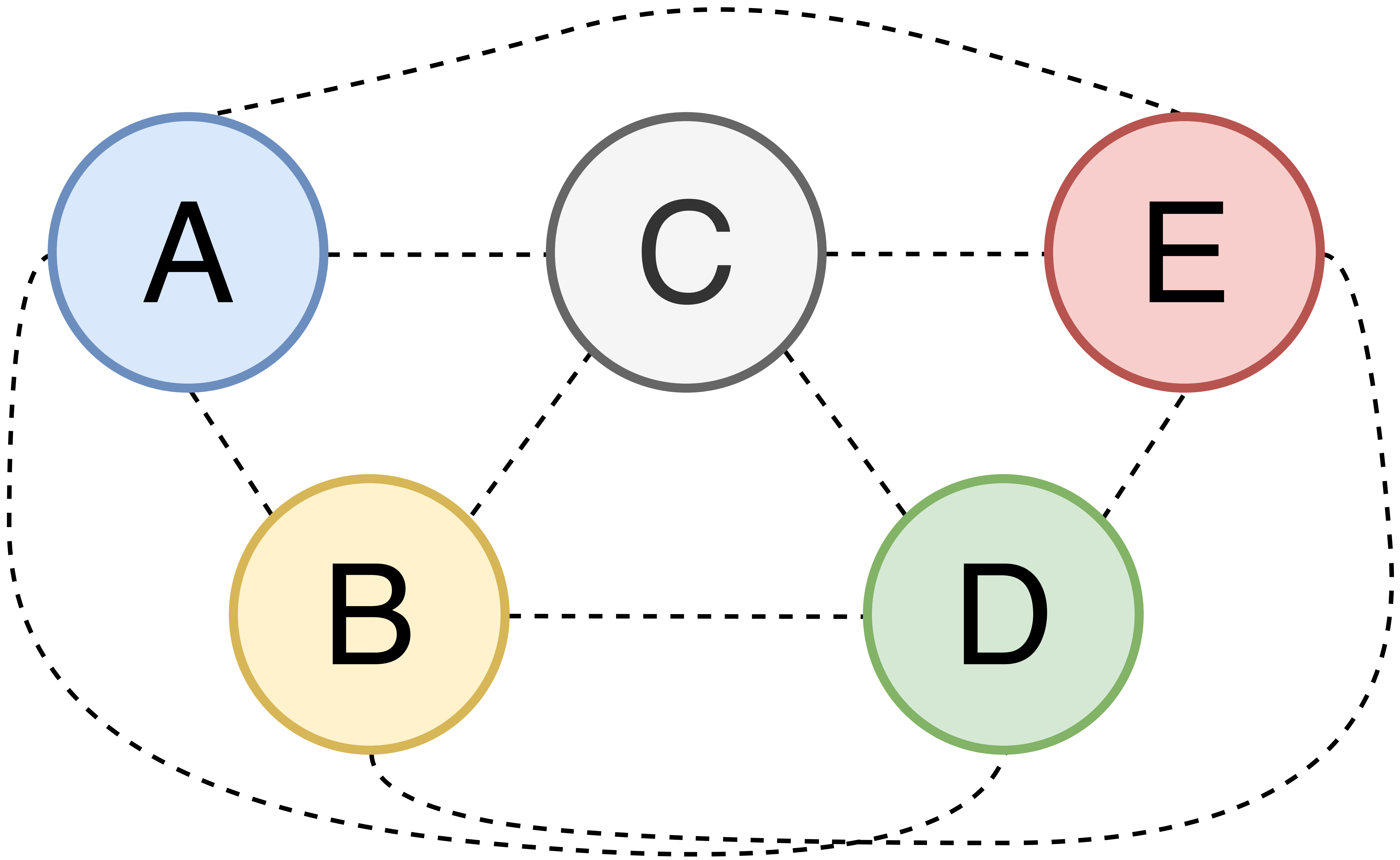
- What graphs would have been observable?



Modeling Networks

Why do we want to study networks?

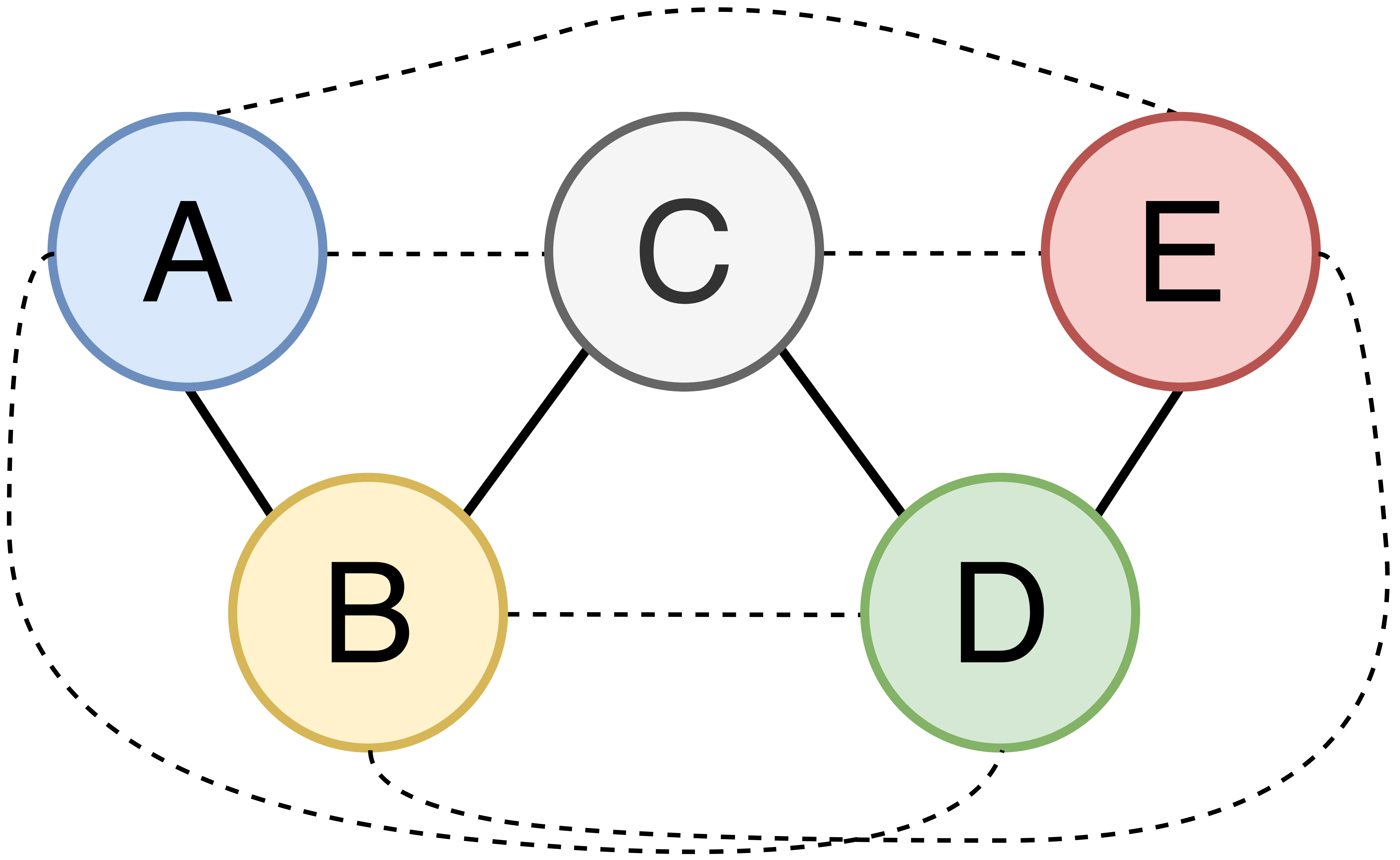
- What graphs would have been observable?



Modeling Networks

Why do we want to study networks?

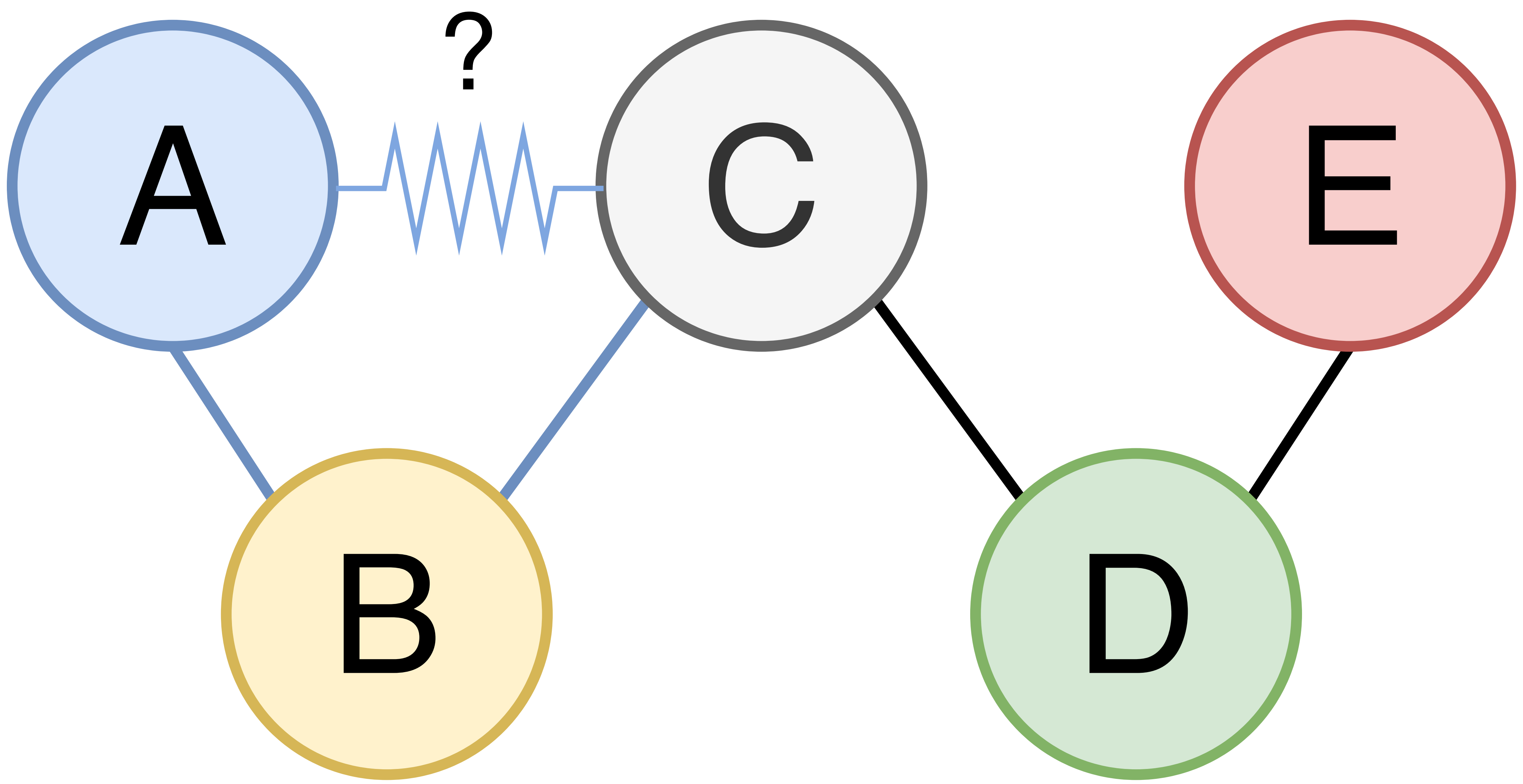
- What graphs would have been observable? ⇒ Why this one?



Modeling Networks

Why do we want to study networks?

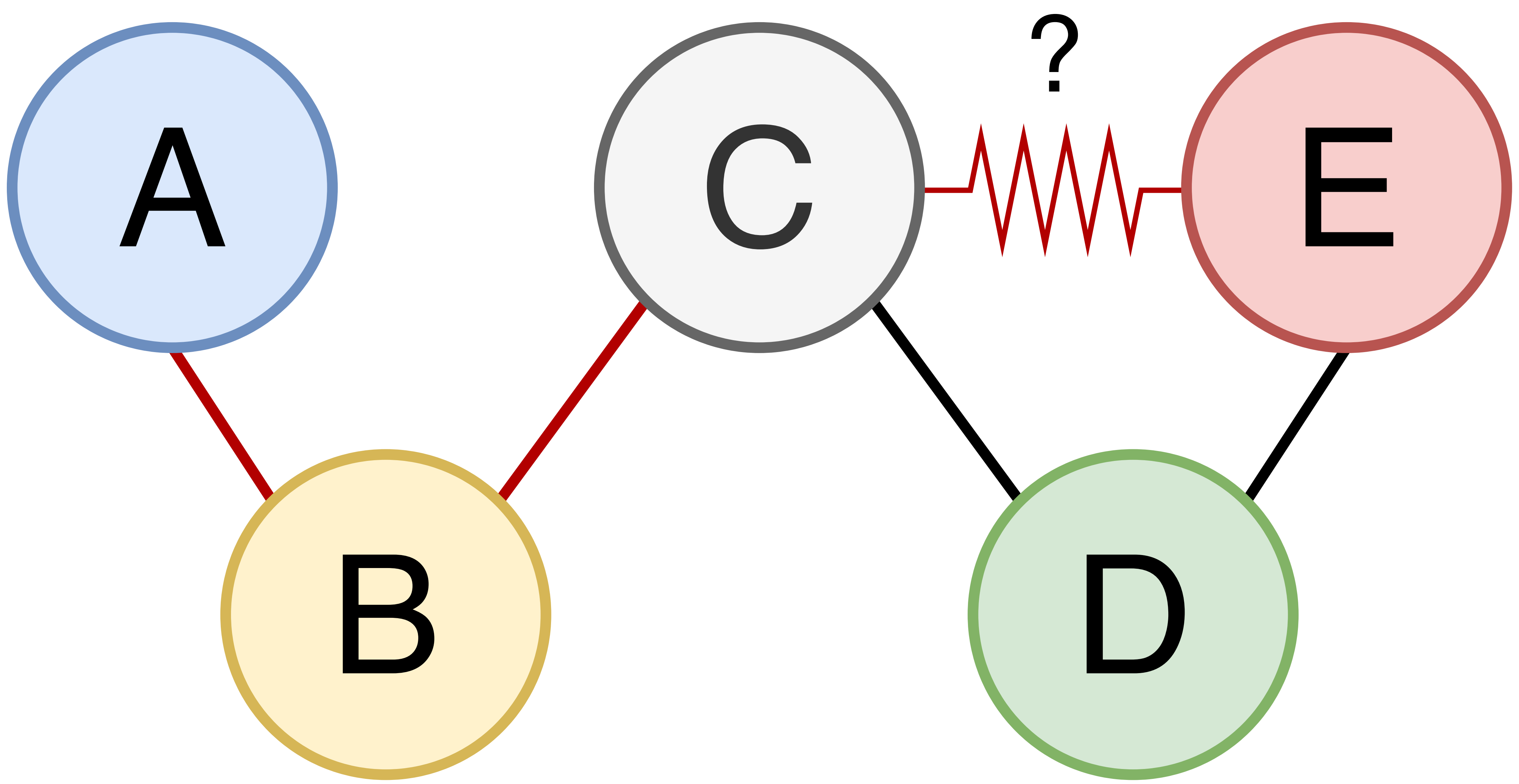
- Are there structural mechanisms at play? (e.g., transitivity)
- The existence of one edge might change the probability of another edge



Modeling Networks

Why do we want to study networks?

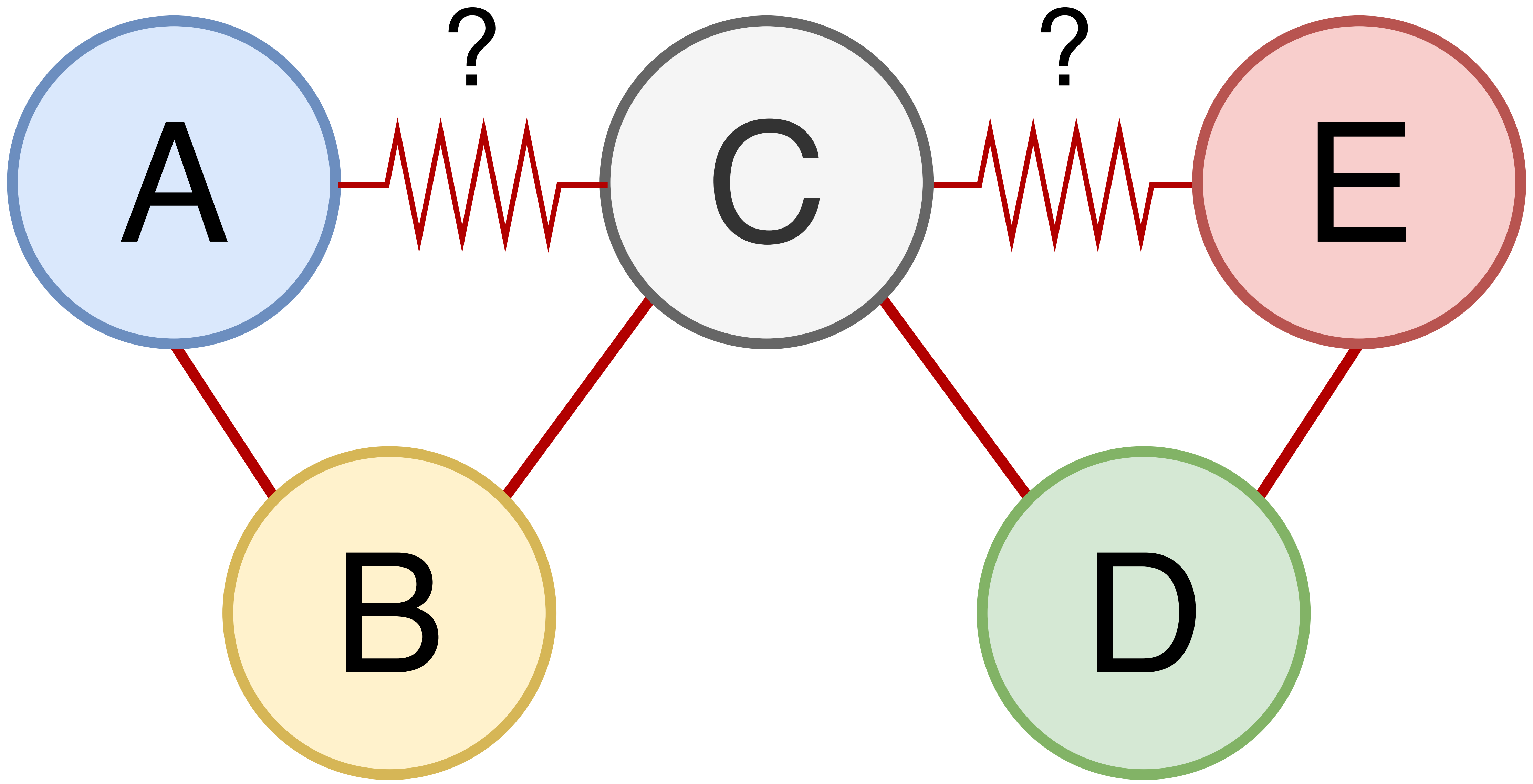
- Are there structural mechanisms at play? (e.g., transitivity)
- The existence of one edge might change the probability of another edge



Modeling Networks

Why do we want to study networks?

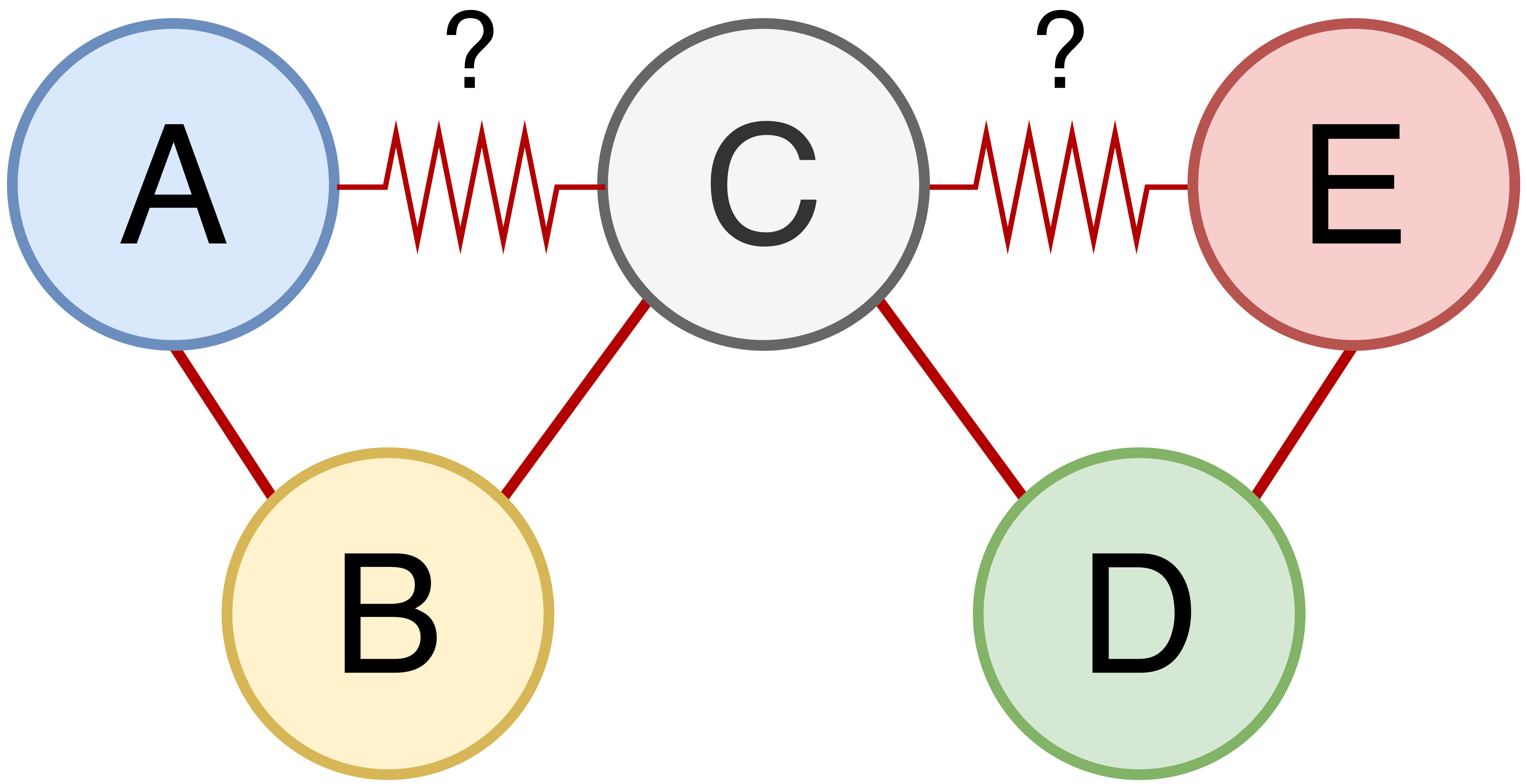
- Do all actors behave the same way?



Modeling Networks

Why do we want to study networks?

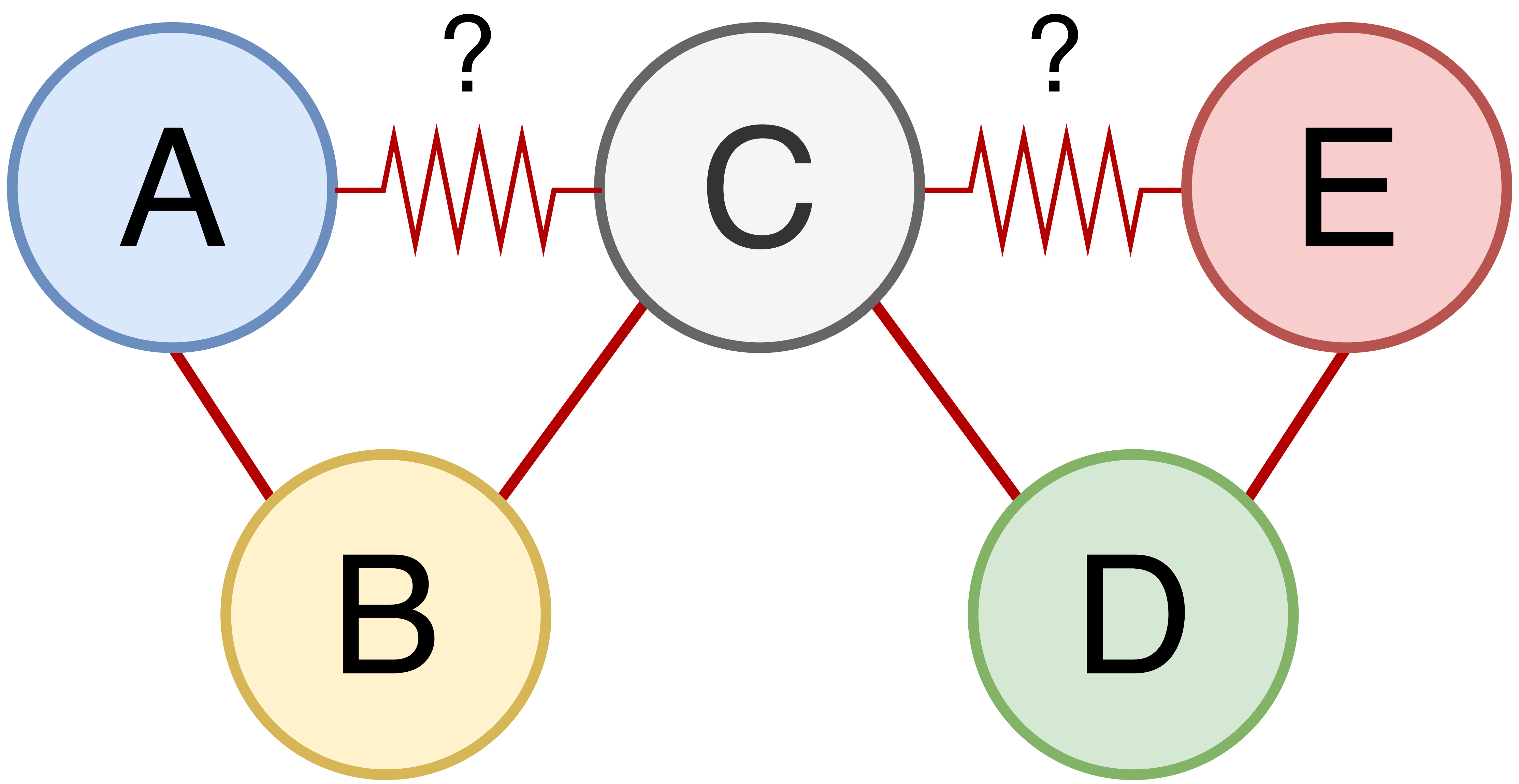
- Do all actors behave the same way?
- **Time of independent observations is over ⇒ non-iid setting**



Modeling Networks

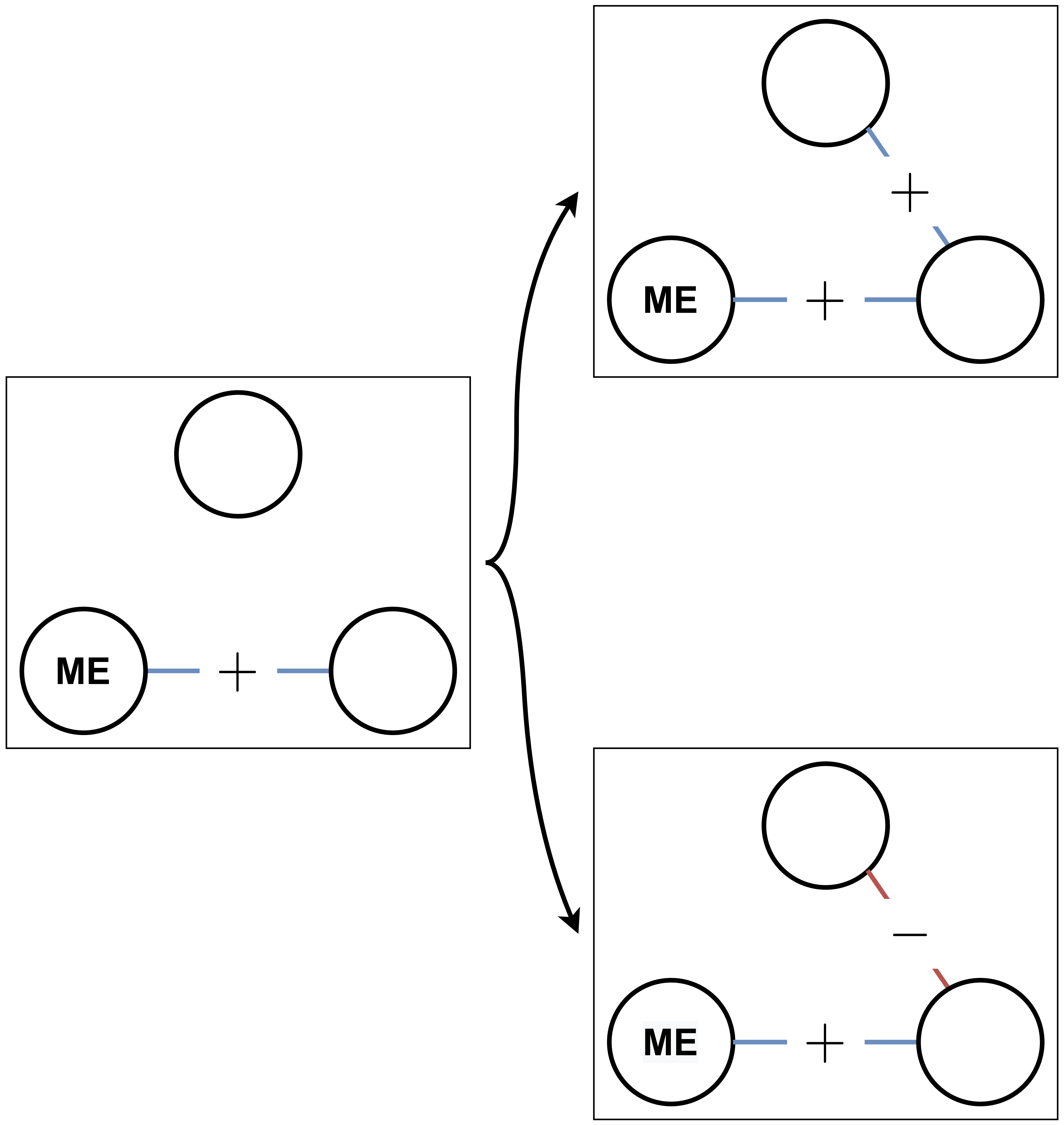
Why do we want to study networks?

- Time of independent observations is over ⇒ simultaneous dependence
- When is this the case? ⇒ When studying network theories



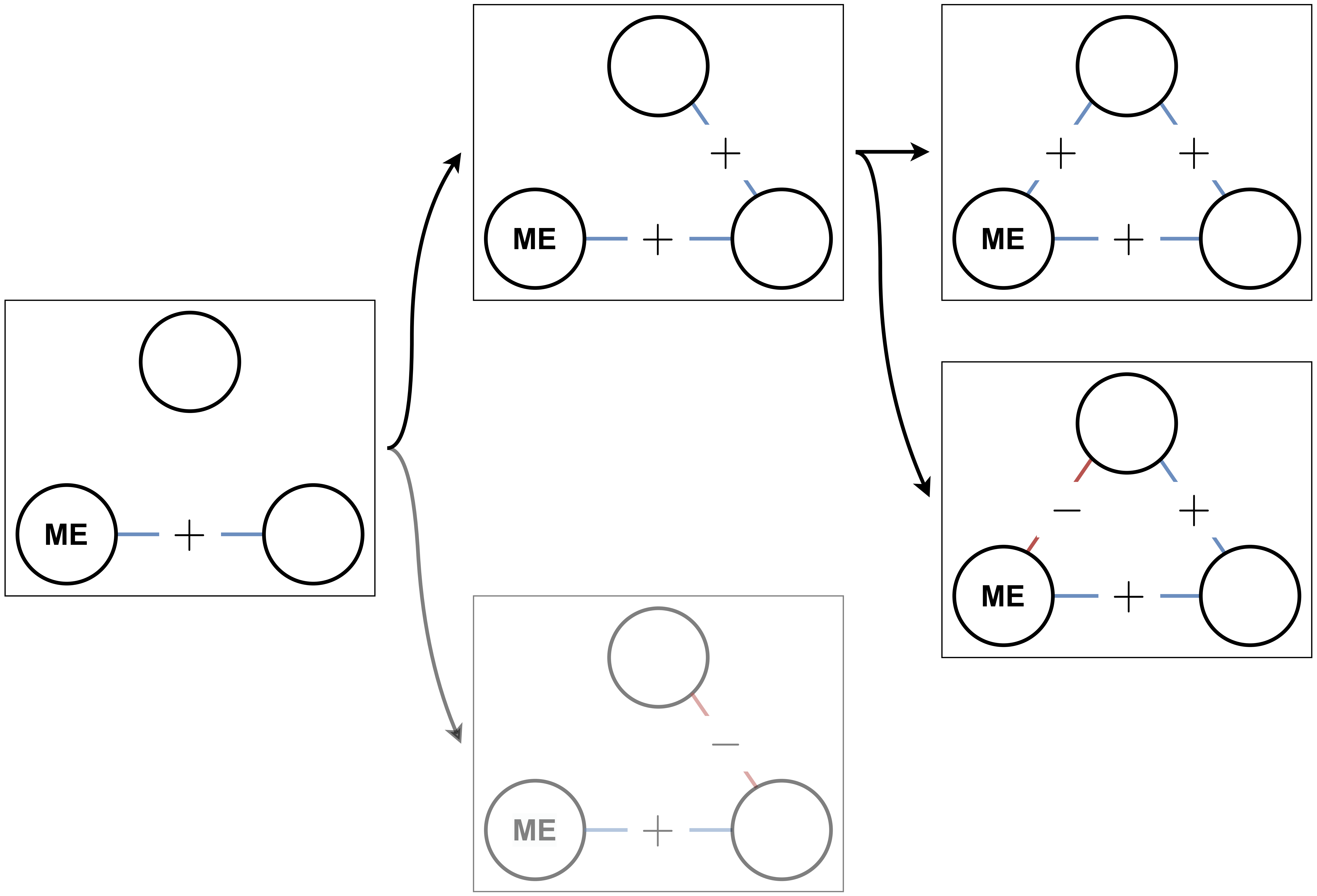
Network Theories

What’s structural balance theory?



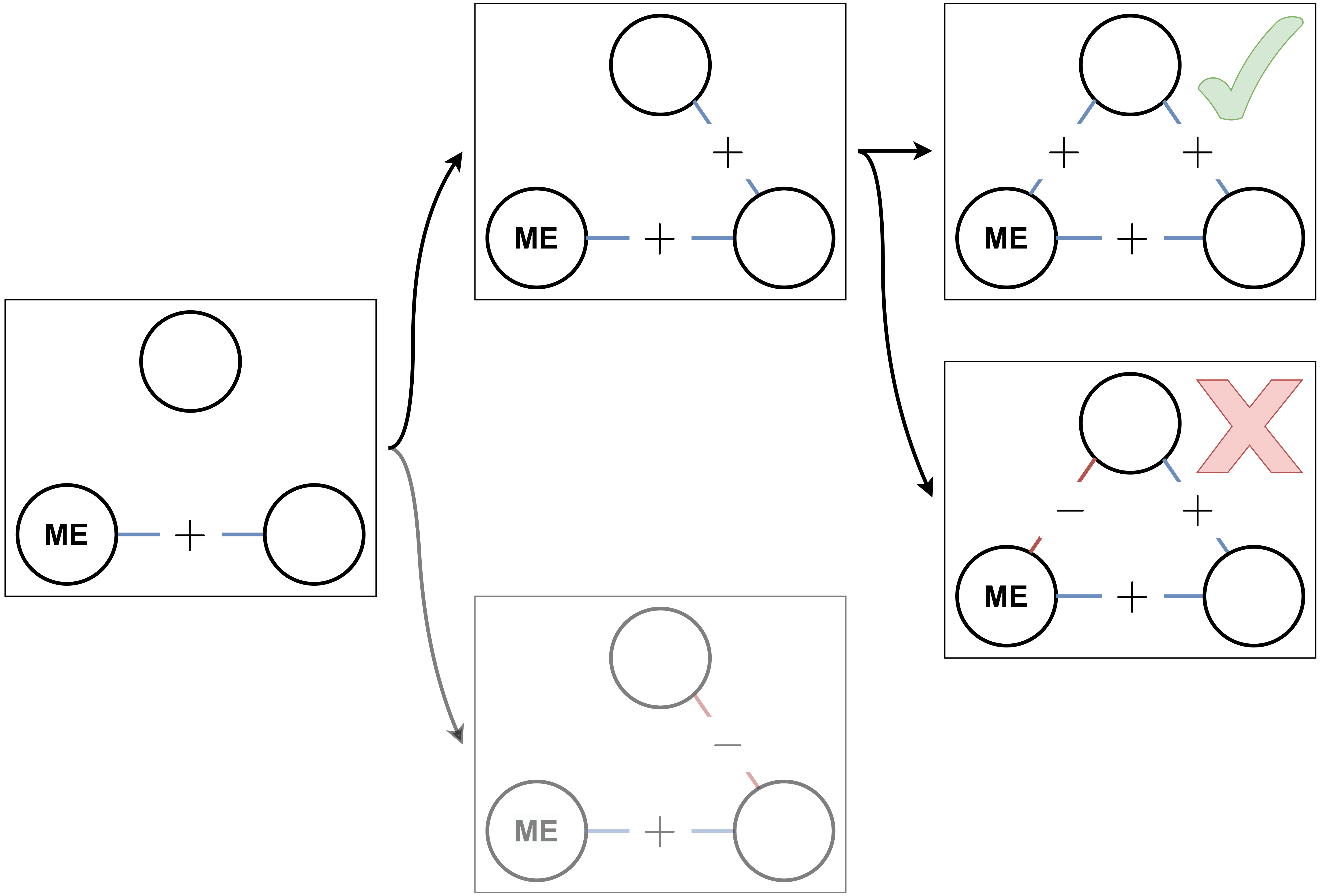
Network Theories

What's structural balance theory?



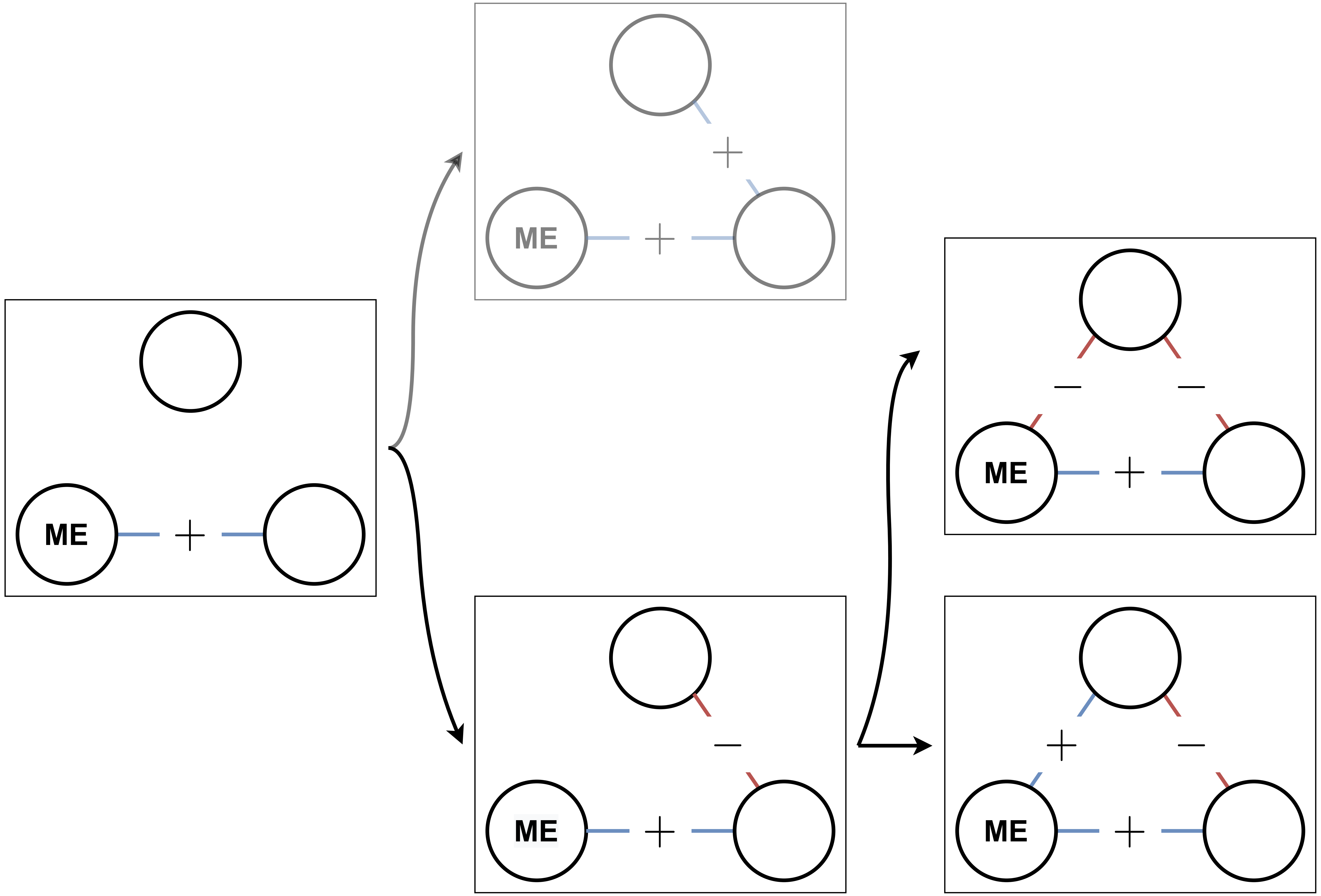
Network Theories

What's structural balance theory?



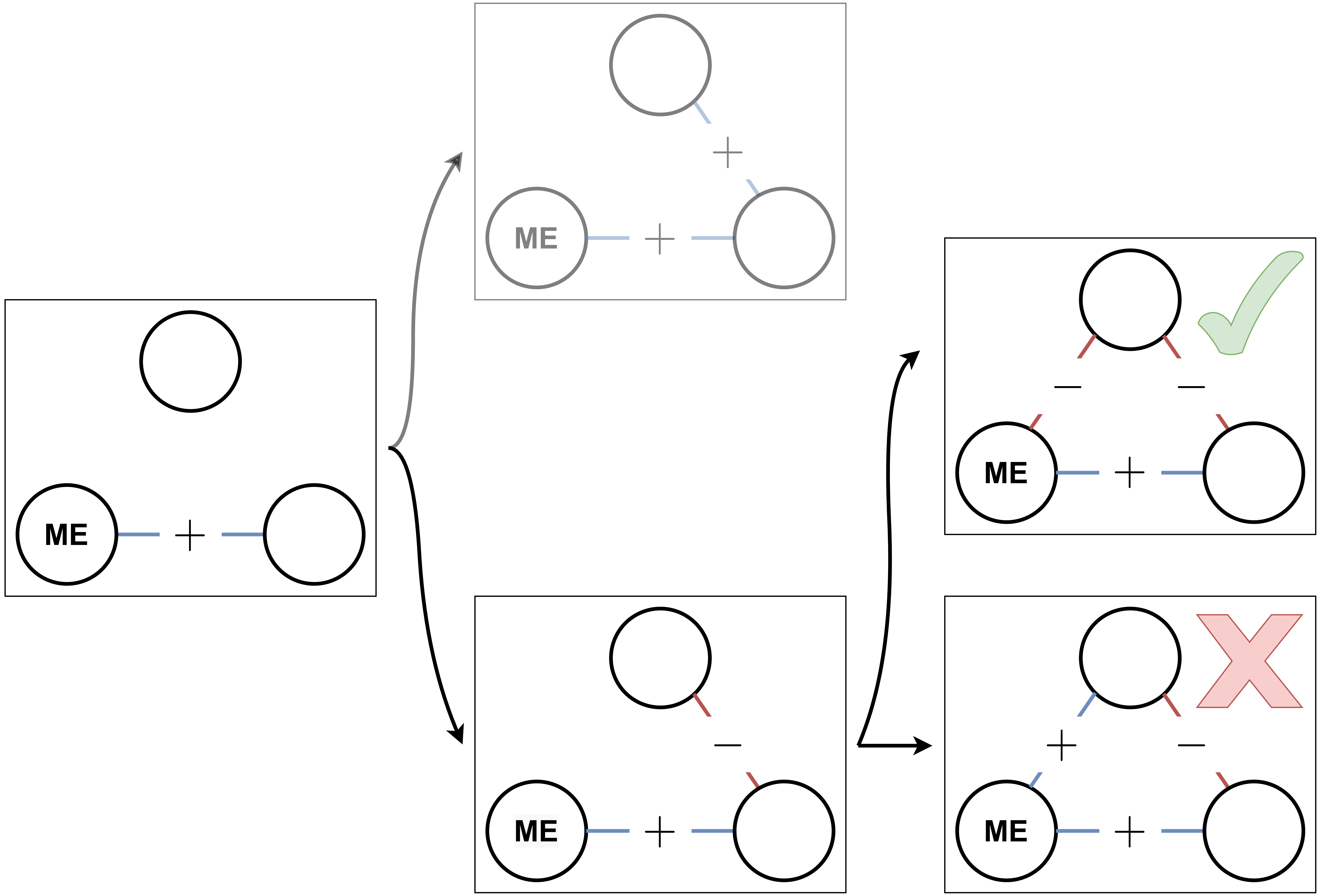
Network Theories

What’s structural balance theory?



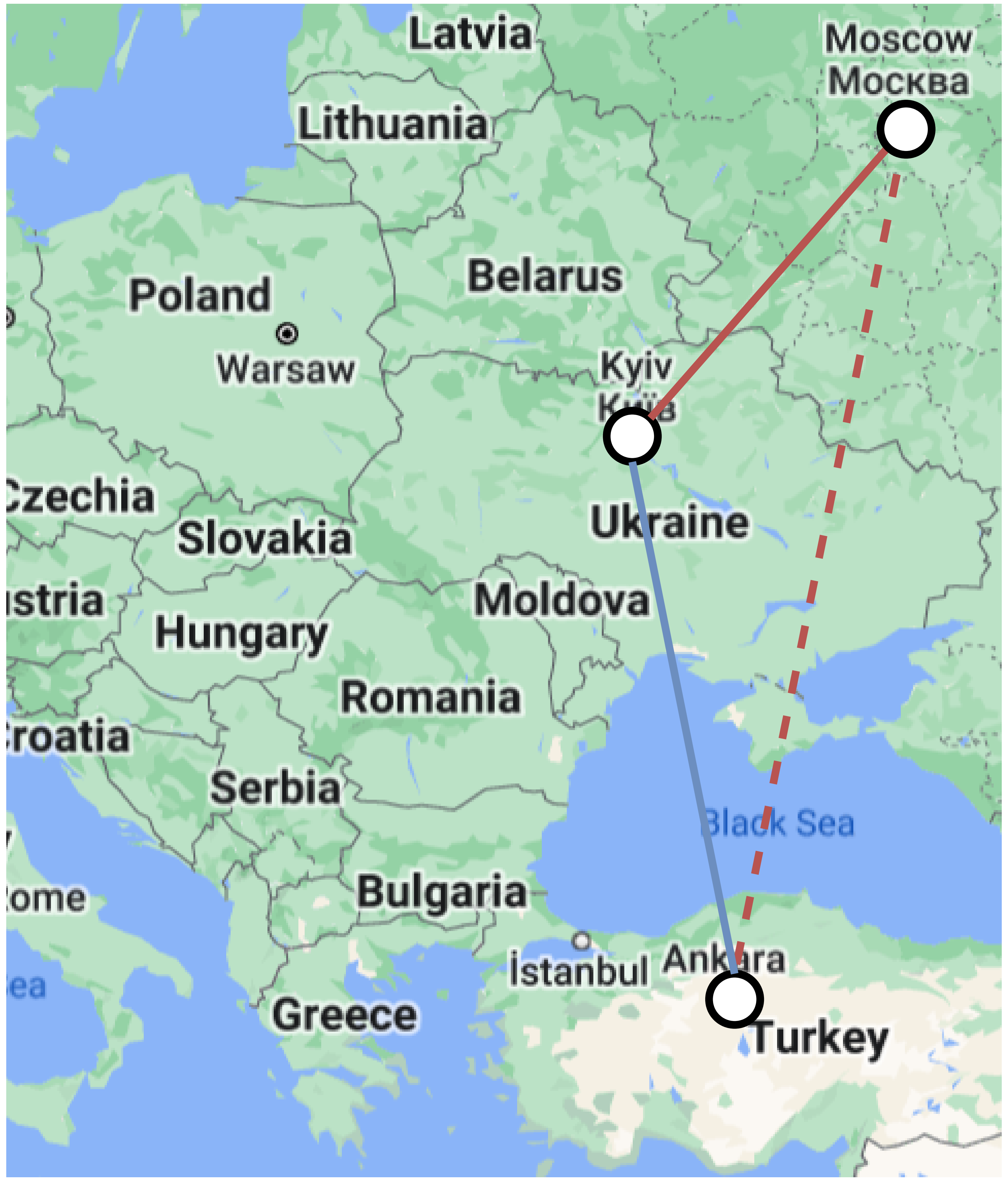
Network Theories

What’s structural balance theory?



Network Theories

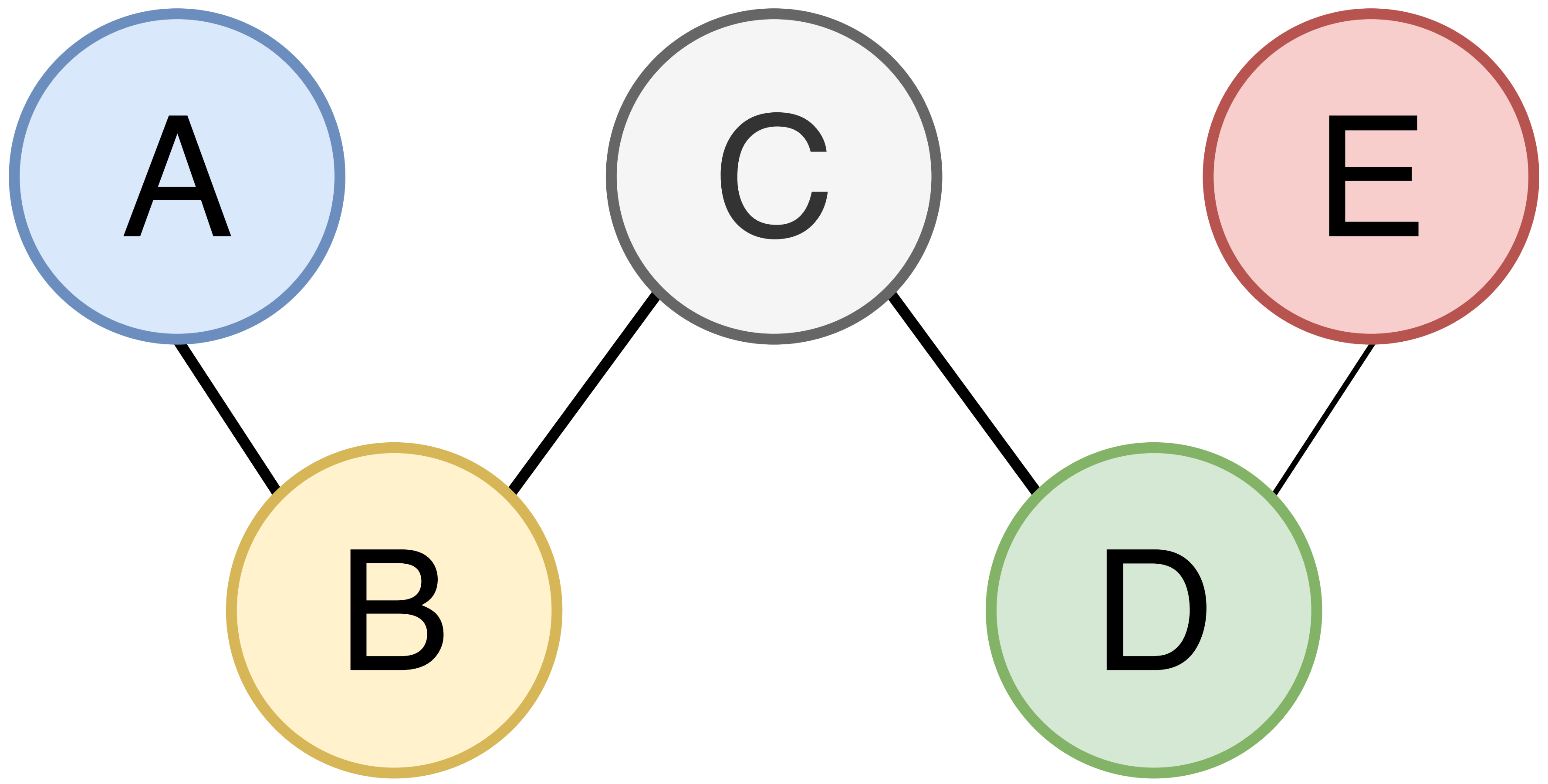
What's structural balance theory?



Aim Network Models

How do we want to study networks?

1. We want to define a probability distribution over graphs
2. Tackle problem that conditional independence assumptions are violated
3. Do all this with one network (“$n = 1$”)



Graph $\mathscr{G}$

$$
\begin{array}{c c}
 & \begin{array}{ccccc} \text{A} & \text{B} & \text{C} & \text{D} & \text{E} \end{array} \\
\begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \\ \text{D} \\ \text{E} \end{array} &
\begin{pmatrix}
- & 1 & 0 & 0 & 0 \\
1 & - & 1 & 0 & 0 \\
0 & 1 & - & 1 & 0 \\
0 & 0 & 1 & - & 1 \\
0 & 0 & 0 & 1 & -
\end{pmatrix}
\end{array}
$$

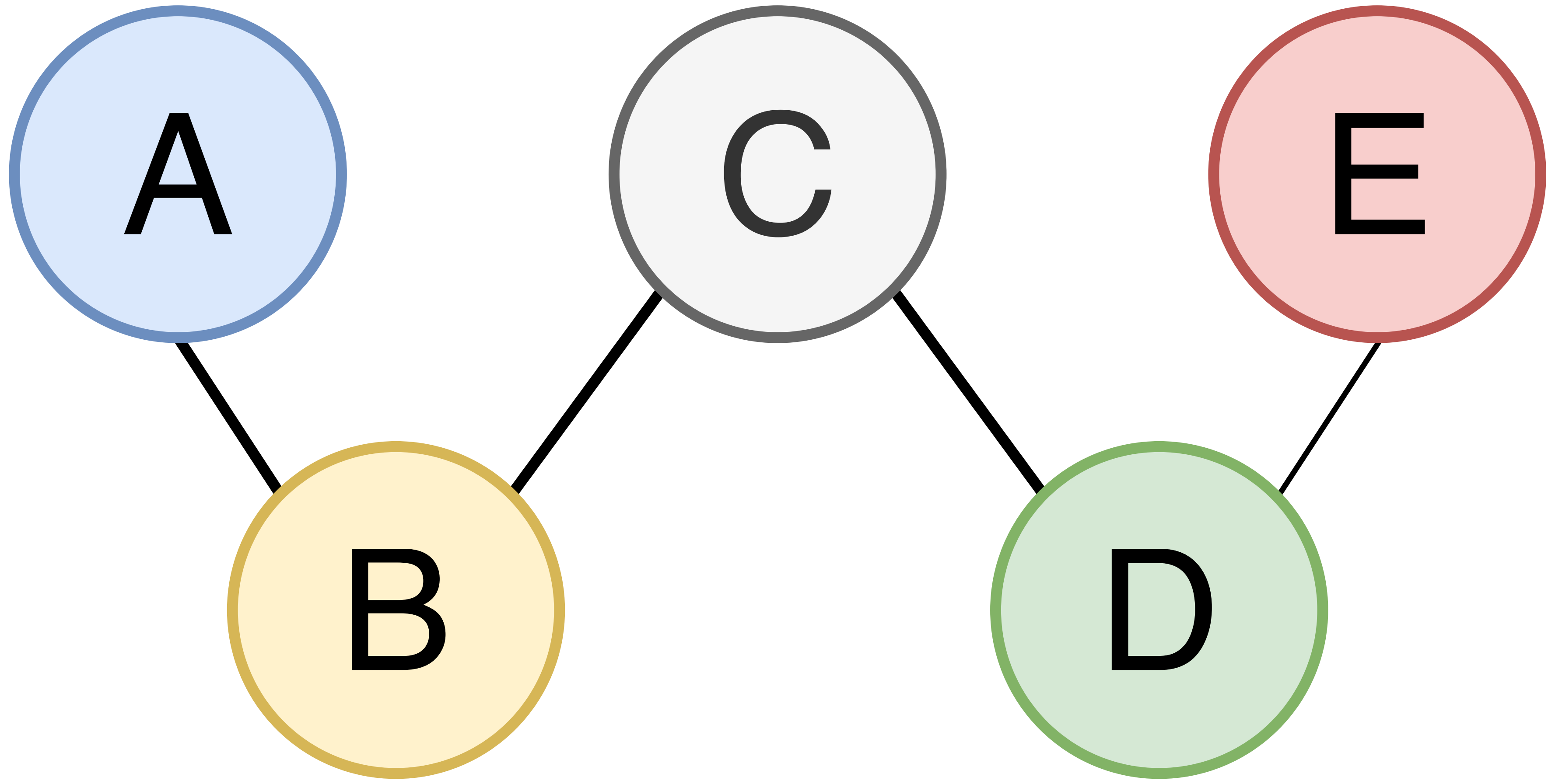
Adjacency Matrix y

$$
y_{ij} = \begin{cases} 1, \text{ if } (i,j) \in \mathscr{E} \\ 0, \text{ else} \end{cases}
$$

Aim Network Models

How do we want to study networks?

1. We want to define a probability distribution over graphs
2. Tackle problem that conditional independence assumptions are violated
3. Do all this with one network ("$n = 1$")



Graph $\mathscr{G}$

$$
\begin{array}{c c}
 & \begin{array}{ccccc} \text{A} & \text{B} & \text{C} & \text{D} & \text{E} \end{array} \\
\begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \\ \text{D} \\ \text{E} \end{array} & \begin{pmatrix} - & 1 & 0 & 0 & 0 \\ 1 & - & 1 & 0 & 0 \\ 0 & 1 & - & 1 & 0 \\ 0 & 0 & 1 & - & 1 \\ 0 & 0 & 0 & 1 & - \end{pmatrix}
\end{array}
$$

Adjacency Matrix y

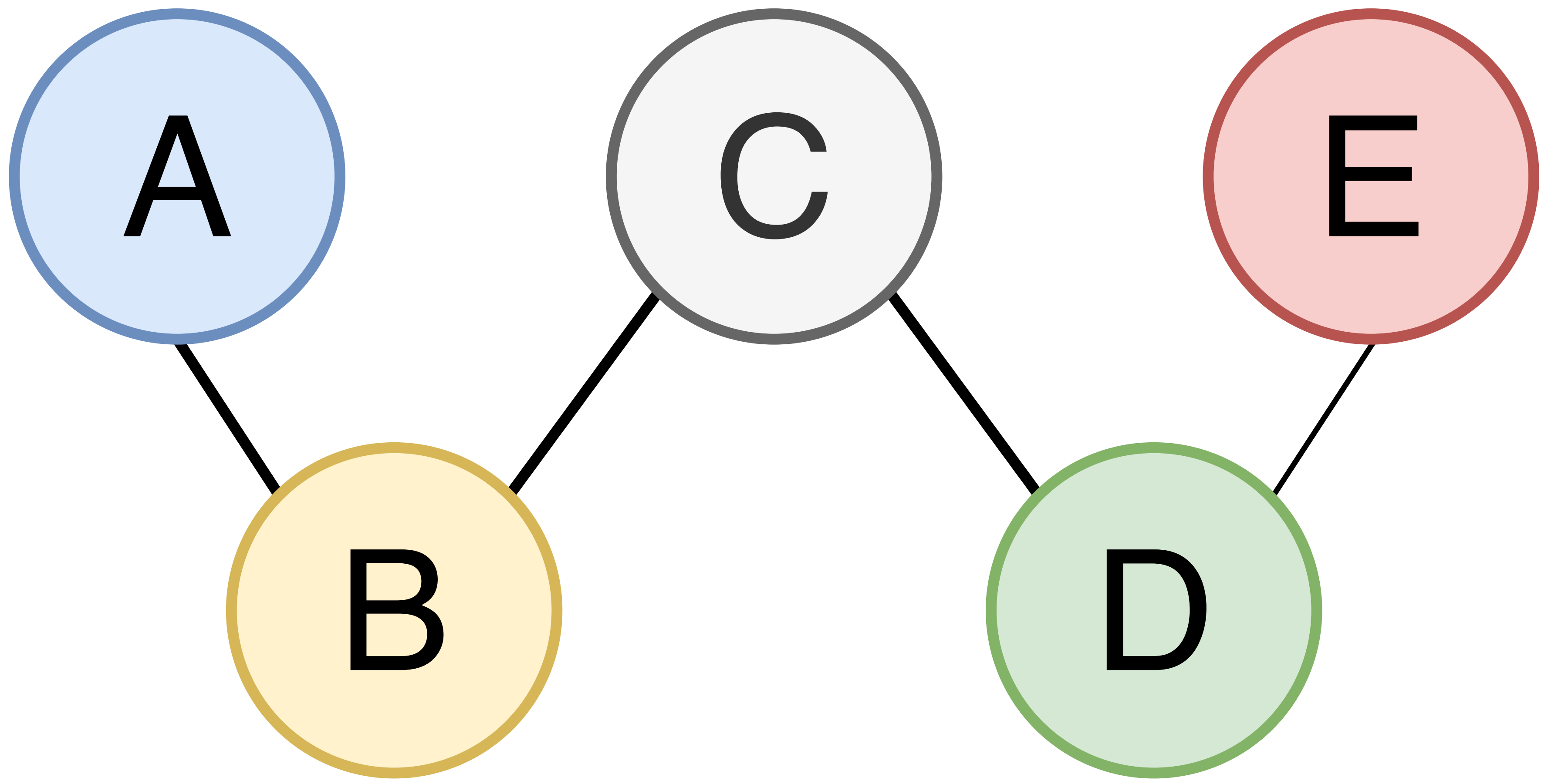
Set of observable adjacency matrices $\mathscr{Y}$

Number of actors N

Aim Network Models

How do we want to study networks?

1. We want to define a probability distribution over graphs
2. Tackle problem that conditional independence assumptions are violated
3. Do all this with one network ("$n = 1$")



Graph $\mathscr{G}$

$$
\begin{array}{c|ccccc}
 & A & B & C & D & E \\
\hline
A & - & 1 & 0 & 0 & 0 \\
B & 1 & - & 1 & 0 & 0 \\
C & 0 & 1 & - & 1 & 0 \\
D & 0 & 0 & 1 & - & 1 \\
E & 0 & 0 & 0 & 1 & -
\end{array}
$$

Adjacency Matrix y

Y Random variable

y Observed random variable

Exponential Random Graph Models

General Agenda

1. Propose a class of realistic statistical models for social networks
2. Estimate the parameters to identify the model with observed data
3. Understand the uncertainty associated with the estimated parameters
4. Test competing explanations for structural effects

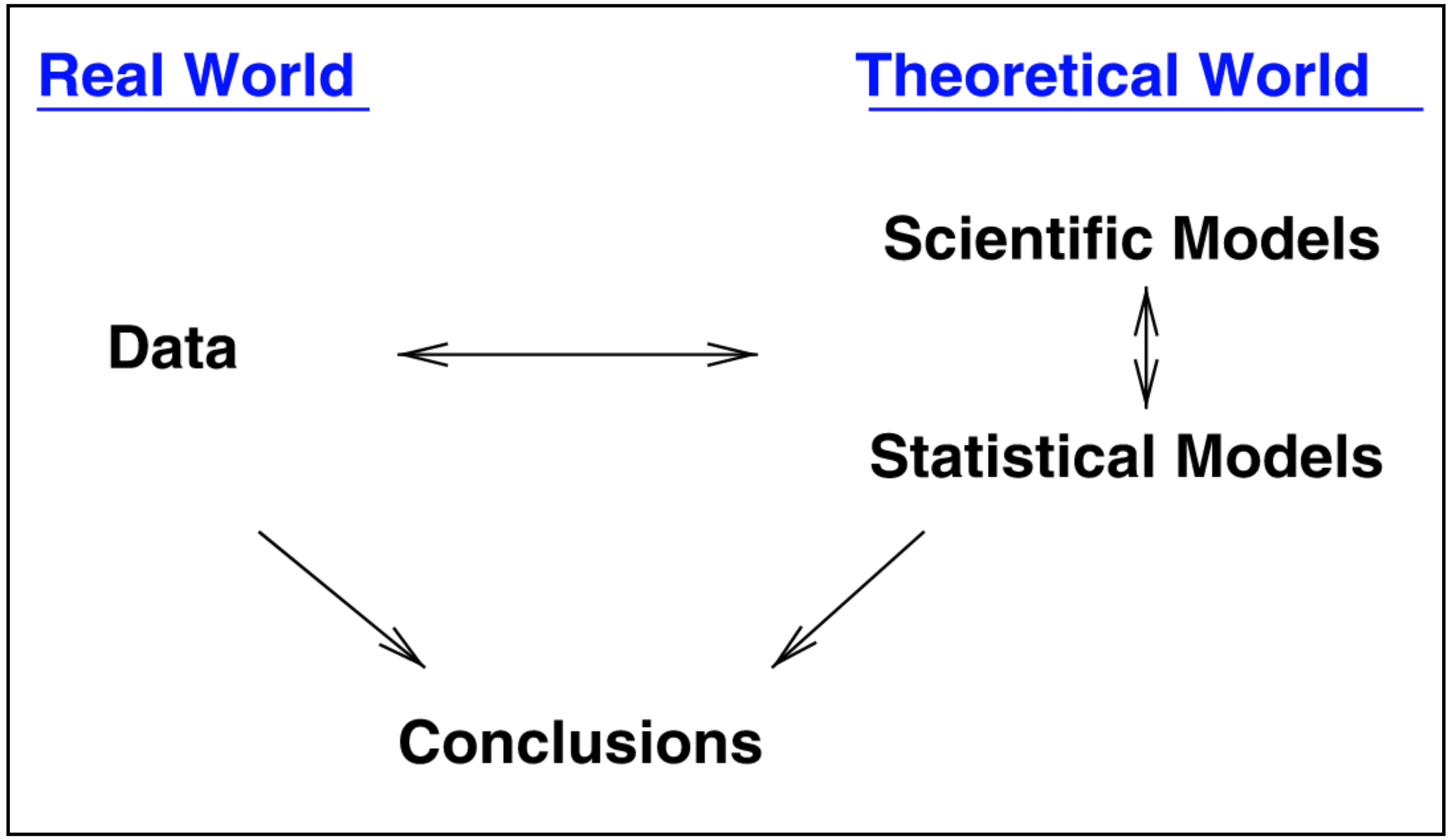
Solution: Random graph model

Define $\underbrace{\mathbb{P}_\theta(Y = y)}_{\text{Probability to observe } y \in \mathcal{Y}}$ such that $\mathbb{P}_\theta(Y = \underbrace{y_{\text{obs}}}_{\text{Observed network}}) = \max_{\tilde{y} \in \mathcal{Y}} \mathbb{P}_\theta\left(Y = \tilde{y}\right)$

General Agenda

1. Propose a class of realistic statistical models for social networks
2. Estimate the parameters to identify the model with observed data
3. Understand the uncertainty associated with the estimated parameters
4. Test competing explanations for structural effects

Solution: Random graph model



General Agenda

1. Propose a class of realistic statistical models for social networks
2. Estimate the parameters to identify the model with observed data
3. Understand the uncertainty associated with the estimated parameters
4. Test competing explanations for structural effects

Solution: Random graph model

Define $\underbrace{\mathbb{P}_\theta(Y = y)}_{\text{Probability to observe } y \in \mathcal{Y}}$ such that

$$\mathbb{P}_\theta\left(Y = \underbrace{\begin{pmatrix} - & 1 & 0 & 0 & 0 \\ 1 & - & 1 & 0 & 0 \\ 0 & 1 & - & 1 & 0 \\ 0 & 0 & 1 & - & 1 \\ 0 & 0 & 0 & 1 & - \end{pmatrix}}_{\text{Observed network}} \right) = \max_{\tilde{y} \in \mathcal{Y}} \mathbb{P}_\theta\left(Y = \tilde{y}\right)$$

General Agenda

1. Propose a class of realistic statistical models for social networks
2. Estimate the parameters to identify the model with observed data
3. Understand the uncertainty associated with the estimated parameters
4. Test competing explanations for structural effects

Solution: Random graph model

Random graph model

A random graph model is given by two components:

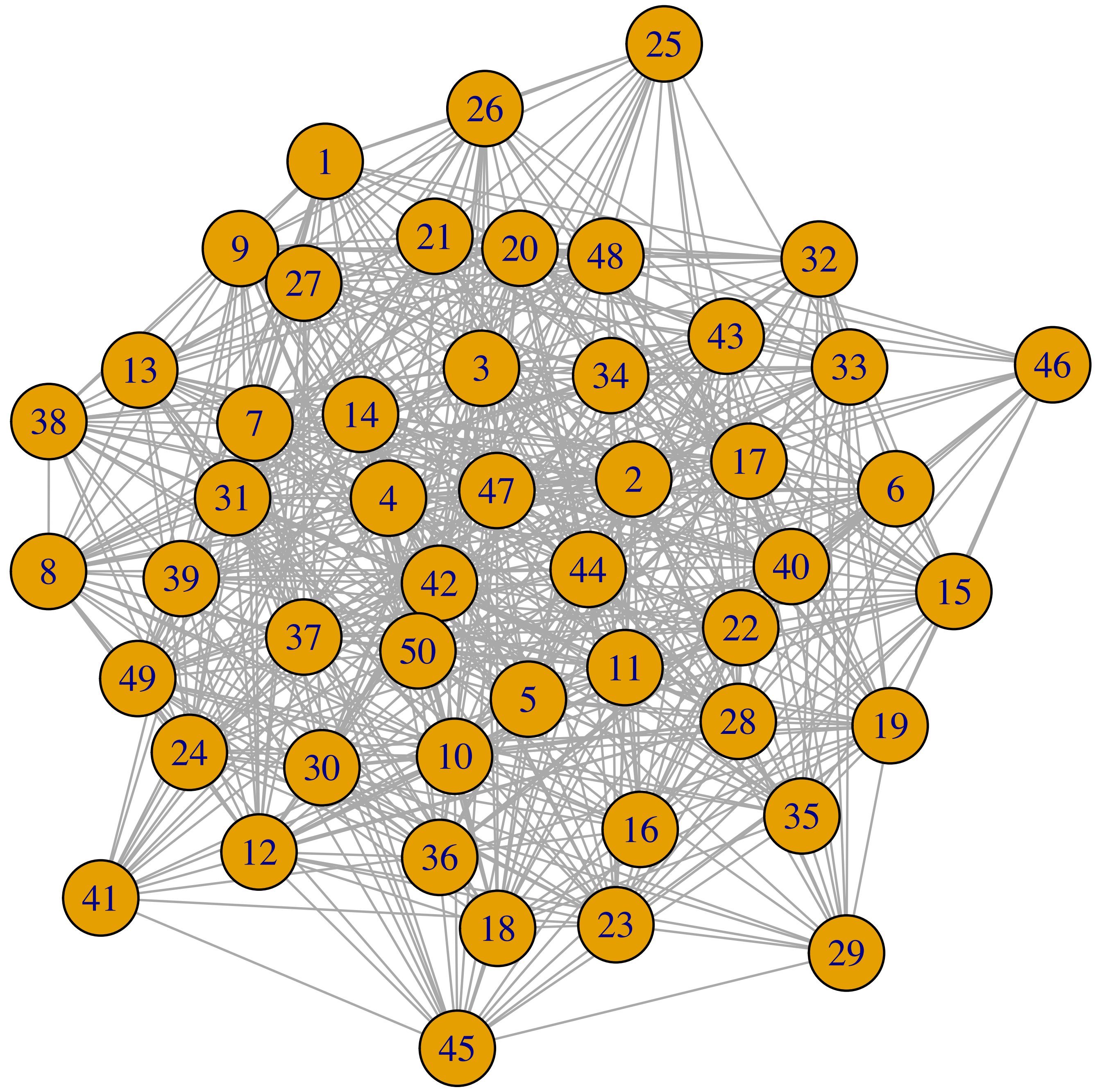
1. *Definition of a set of possible networks or graphs*
    - *How could a network look like or what could happen?*
    - *The actors are fixed and the edges random*
2. *Definition of a probability distribution on this set*
    - *Which networks are more/less likely to be observed?*
    - *The observed network should be the most likely*

ER Model

What's the most basic model for networks you can think of?

⇒ All graphs are equally likely

⇒ All edges are independent and follow a Bernoulli distribution with $\pi = 0.5$



$\pi = 0.5$

$$\mathbb{P}_\theta(Y = y) = \frac{1}{|\mathcal{Y}|} = \frac{1}{2^{\binom{n}{2}}}$$

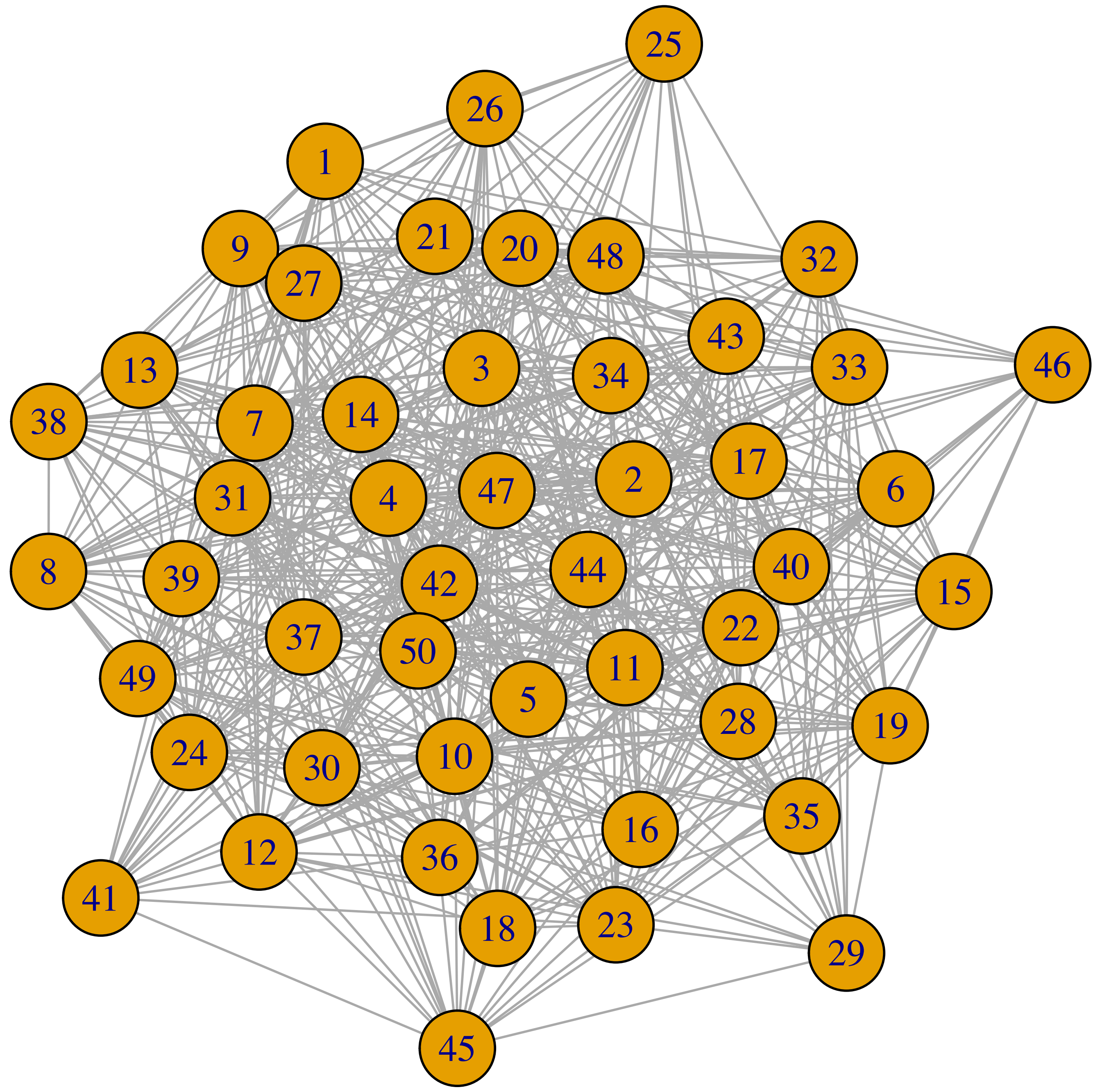
ER Model

What's the most basic model for networks you can think of?

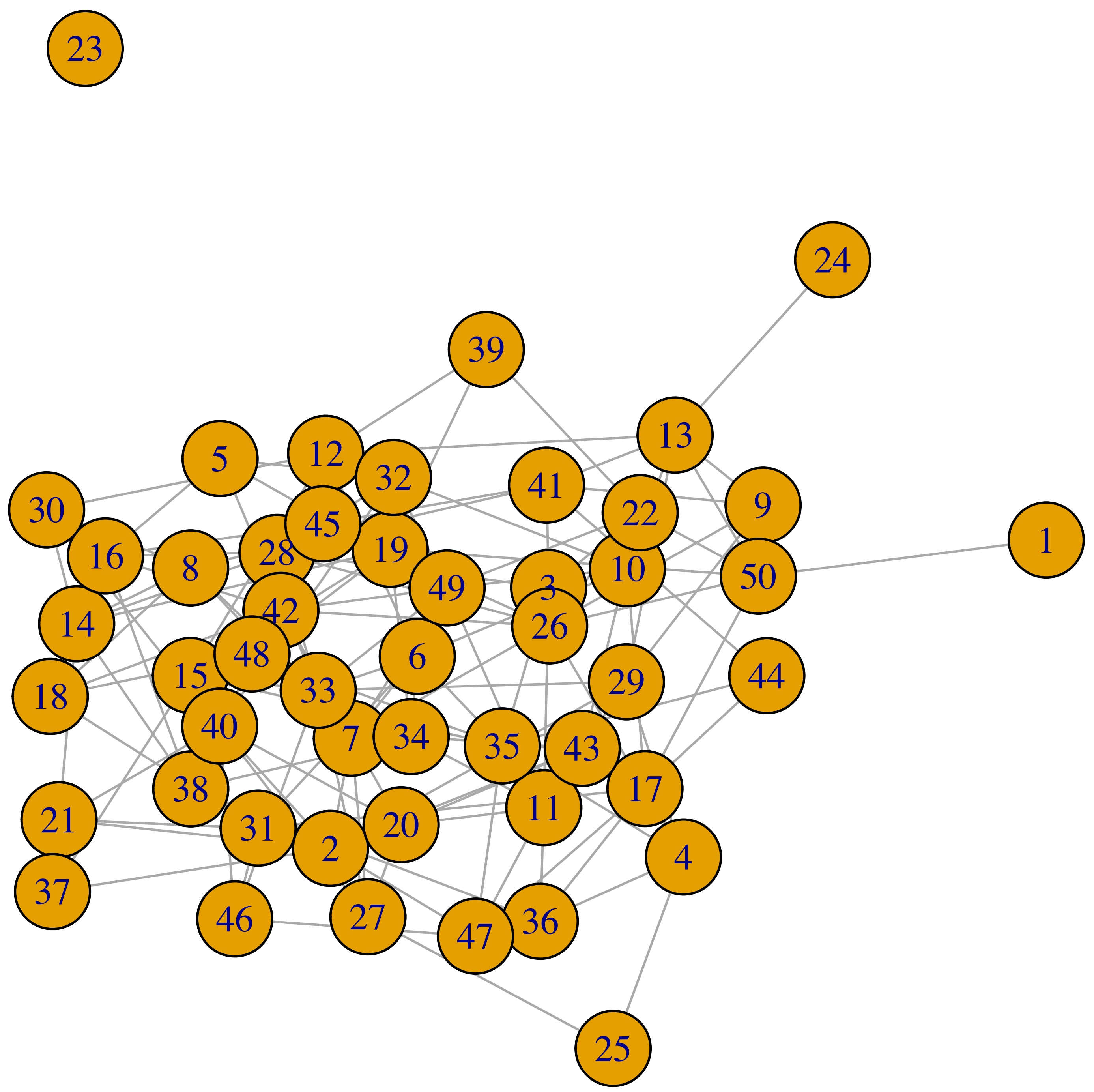
⇒ All graphs are equally likely

⇒ All edges are independent and follow a Bernoulli distribution with $\pi = 0.5$

⇒ For Erdös-Renyi models π can be set arbitrary and be estimated from data



$\pi = 0.5$



$\pi = 0.1$

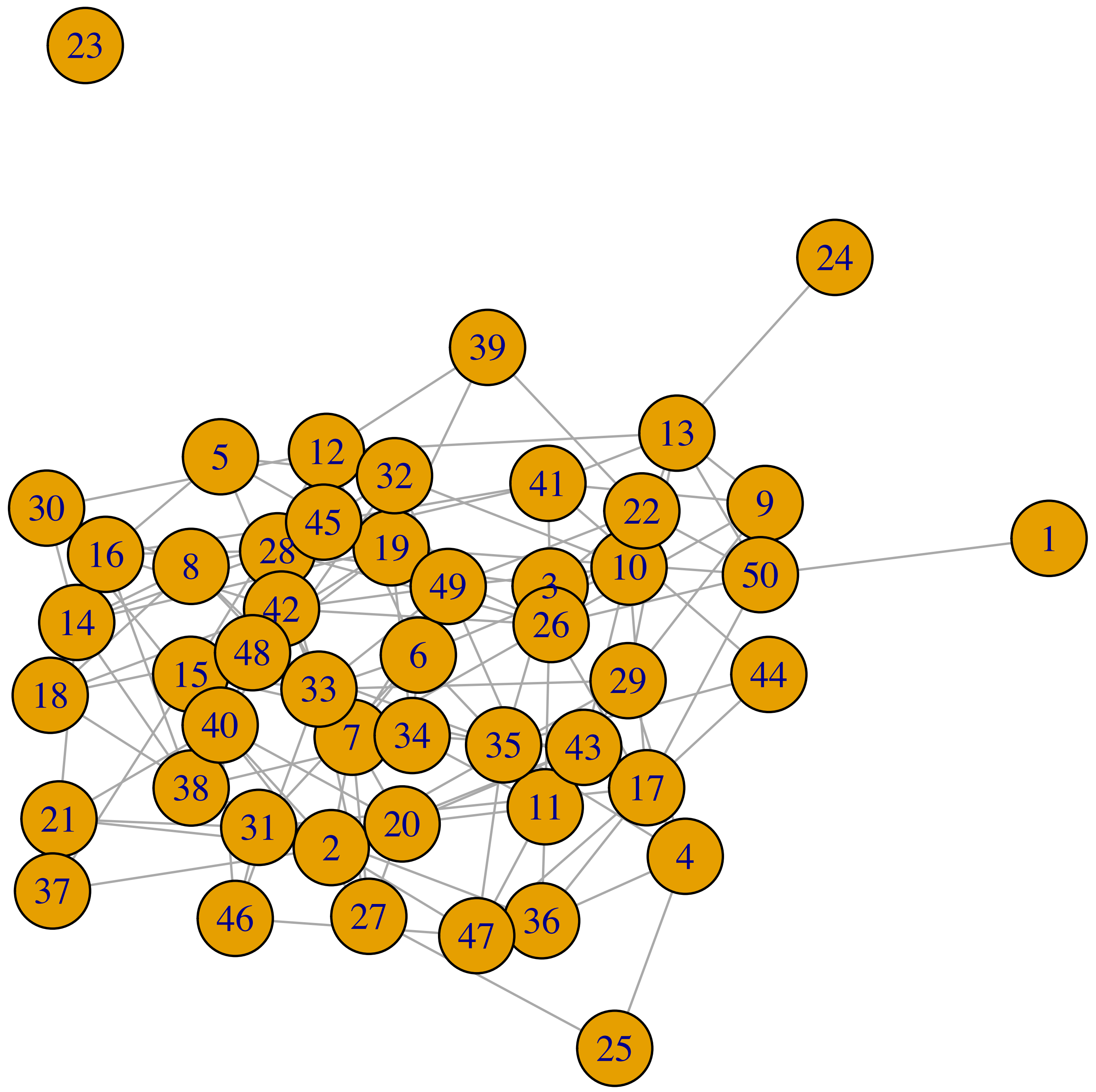
ER Model

What's the most basic model for networks you can think of?

$\Rightarrow$ All graphs are equally likely

$\Rightarrow$ All edges are independent and follow a Bernoulli distribution with $\pi = 0.5$

$\Rightarrow$ For Erdös-Renyi models π can be set arbitrary and be estimated from data



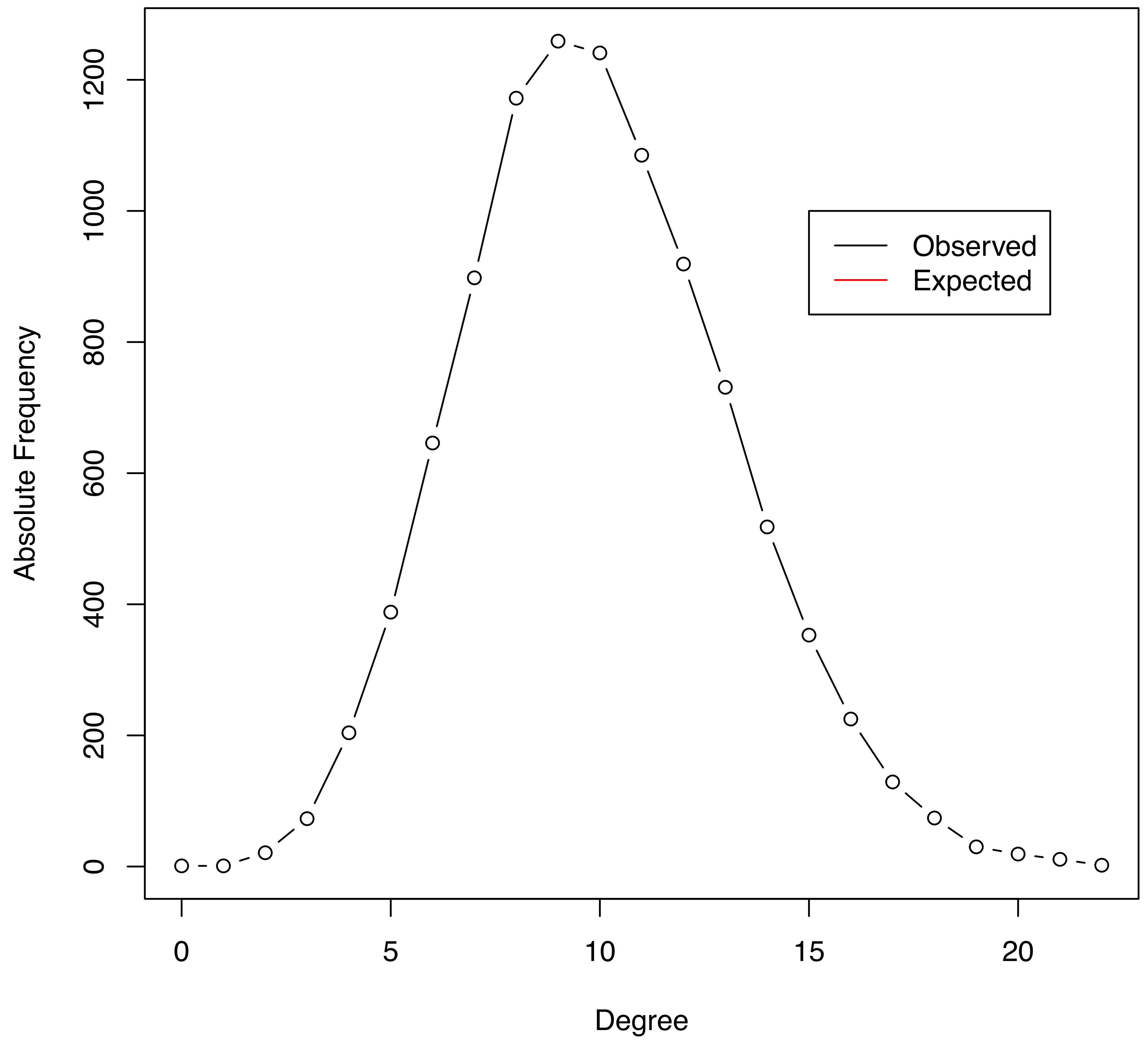
$\pi = 0.1$

$$Y_{ij} \sim \text{Bin}(n = 1, p = \pi)$$

$$\mathbb{P}\pi(Y = y) = \prod_{i<j} \pi^{y_{ij}} (1 - \pi)^{1 - y_{ij}}$$

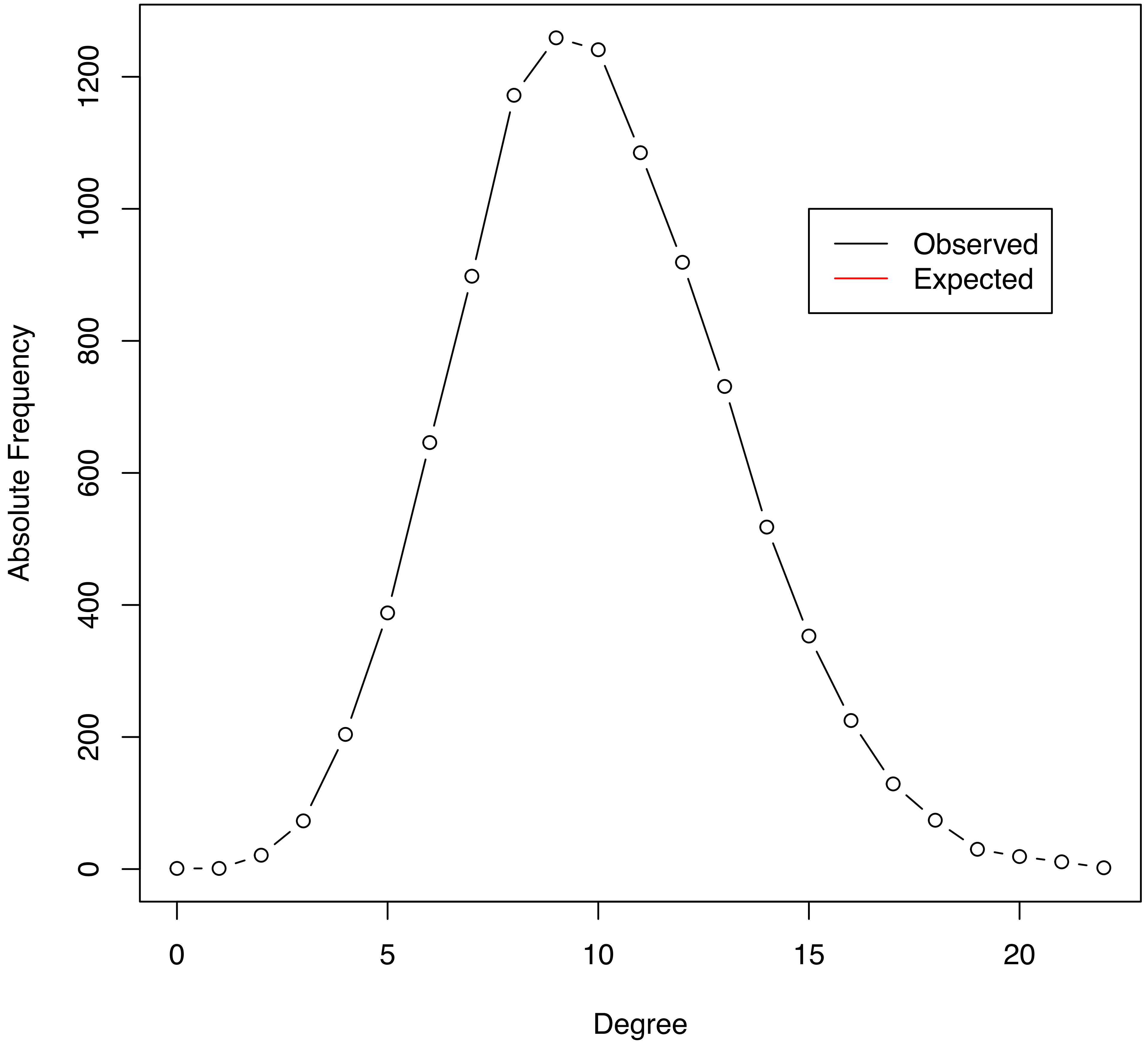
ER Model

$$\mathbb{P}_\pi(Deg(X_i) = k) = \mathbb{P}_\pi\left(\sum_{j \neq i} Y_{ij} = k\right)$$



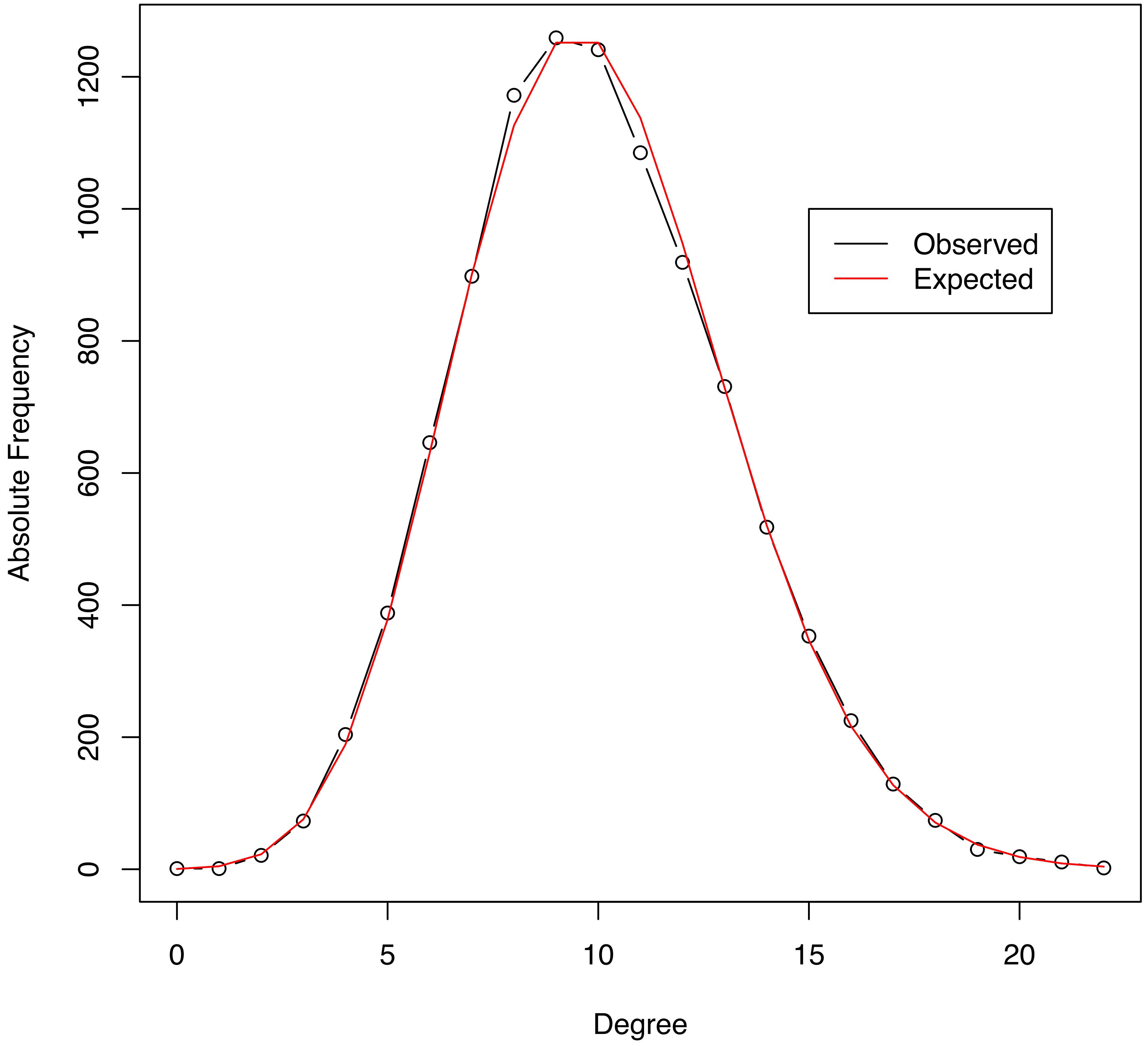
ER Model

$$\mathbb{P}_\pi(Deg(X_i) = k) = \mathbb{P}_\pi\left(\sum_{j \neq i} Y_{ij} = k\right) = \binom{k}{n-1} \pi^k (1-\pi)^{n-1-k}$$



ER Model

$$\mathbb{P}_\pi(Deg(X_i) = k) = \mathbb{P}_\pi\left(\sum_{j\neq i} Y_{ij} = k\right) = \binom{k}{n-1}\pi^k(1-\pi)^{n-1-k}$$



ER Model

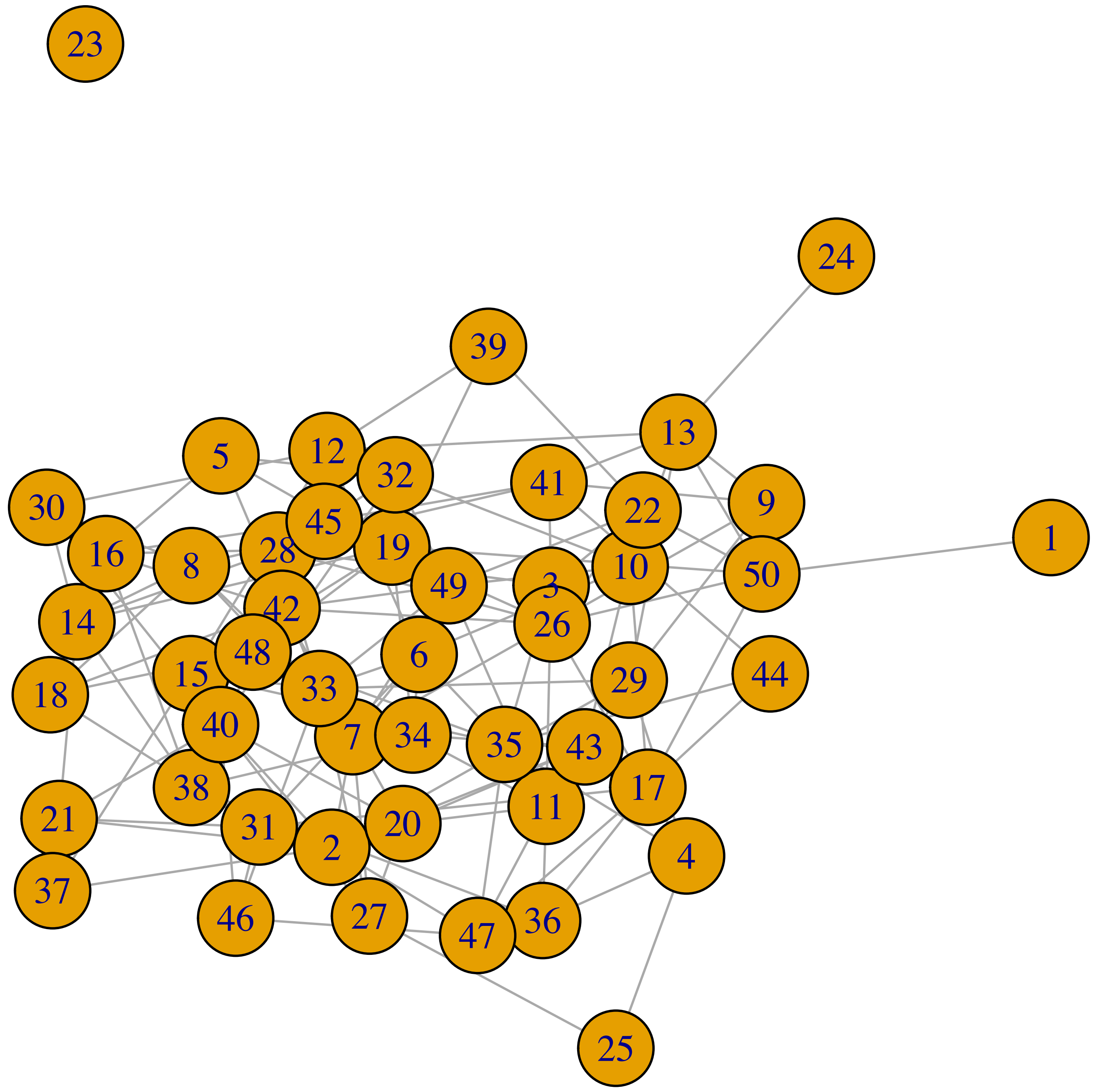
What's the most basic model for networks you can think of?

⇒ All graphs are equally likely

⇒ All edges are independent and follow a Bernoulli distribution with $\pi = 0.5$

⇒ For Erdös-Renyi models π can be set arbitrary and be estimated from data

Does every node behave the same way?



$\pi = 0.1$

$$Y_{ij} \sim \text{Bin}(n = 1, p = \pi)$$

$$\mathbb{P}\pi(Y = y) = \prod_{i<j} \pi^{y_{ij}}(1 - \pi)^{1-y_{ij}}$$

ER Model with Covariates

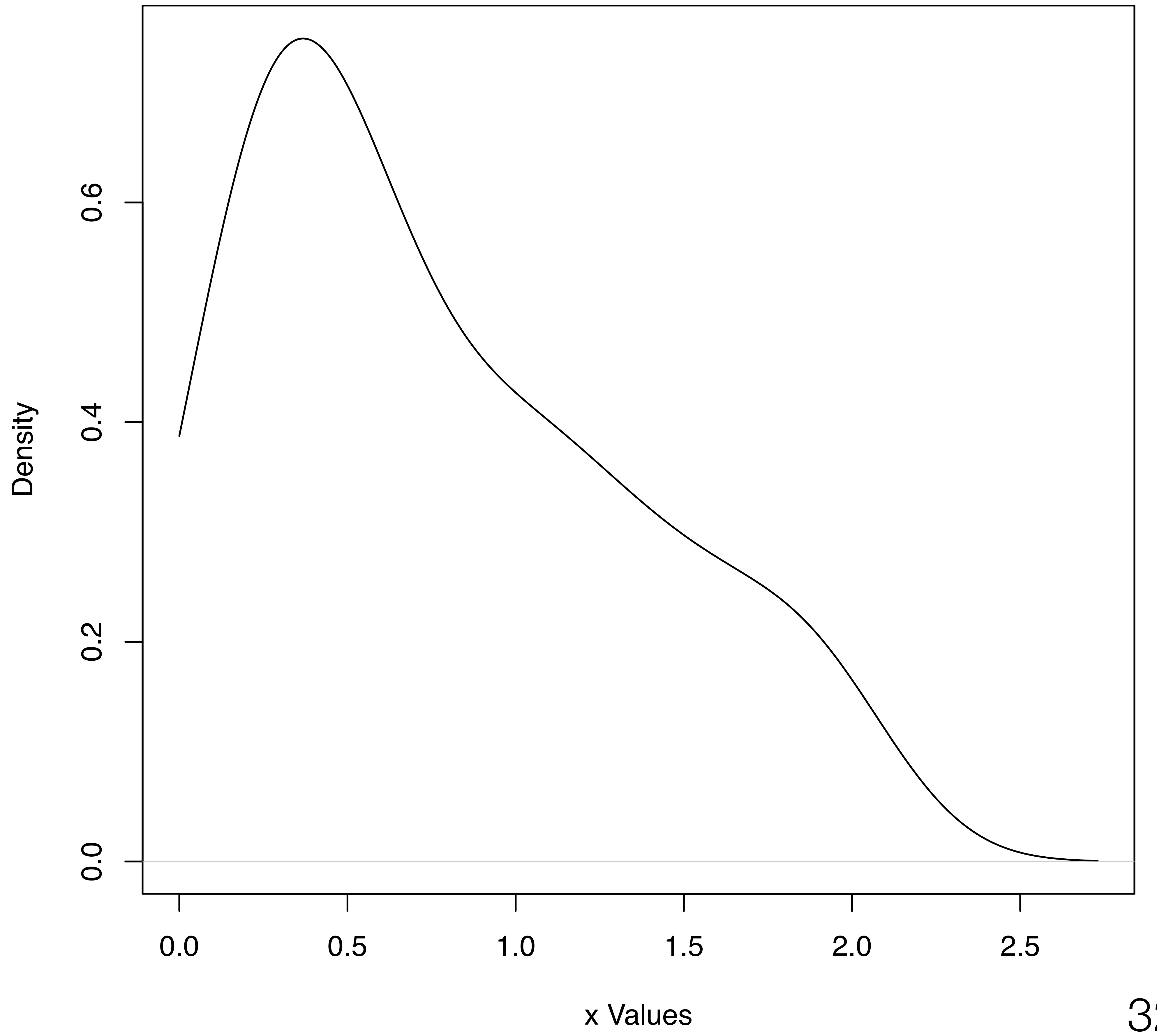
$$\mathbb{P}_\theta(Y = y) = \prod_{i<j} \left(\frac{\exp\{\theta^\top x_{ij}\}}{1 + \exp\{\theta^\top x_{ij}\}} \right)^{y_{ij}} \left(\frac{1}{1 + \exp\{\theta^\top x_{ij}\}} \right)^{1-y_{ij}}$$

Let's add covariates $x_{ij,q} = |x_i - x_j|$!

$\Rightarrow \pi_{ij}$ now changes with different values of x_{ij}

This is the likelihood of a logistic regression!

- $\theta_q > 0$: Higher values of $x_{ij,q}$ make $Y_{ij} = 1$ more likely
- $\theta_q < 0$: Higher values of $x_{ij,q}$ make $Y_{ij} = 1$ less likely



ER Model with Covariates

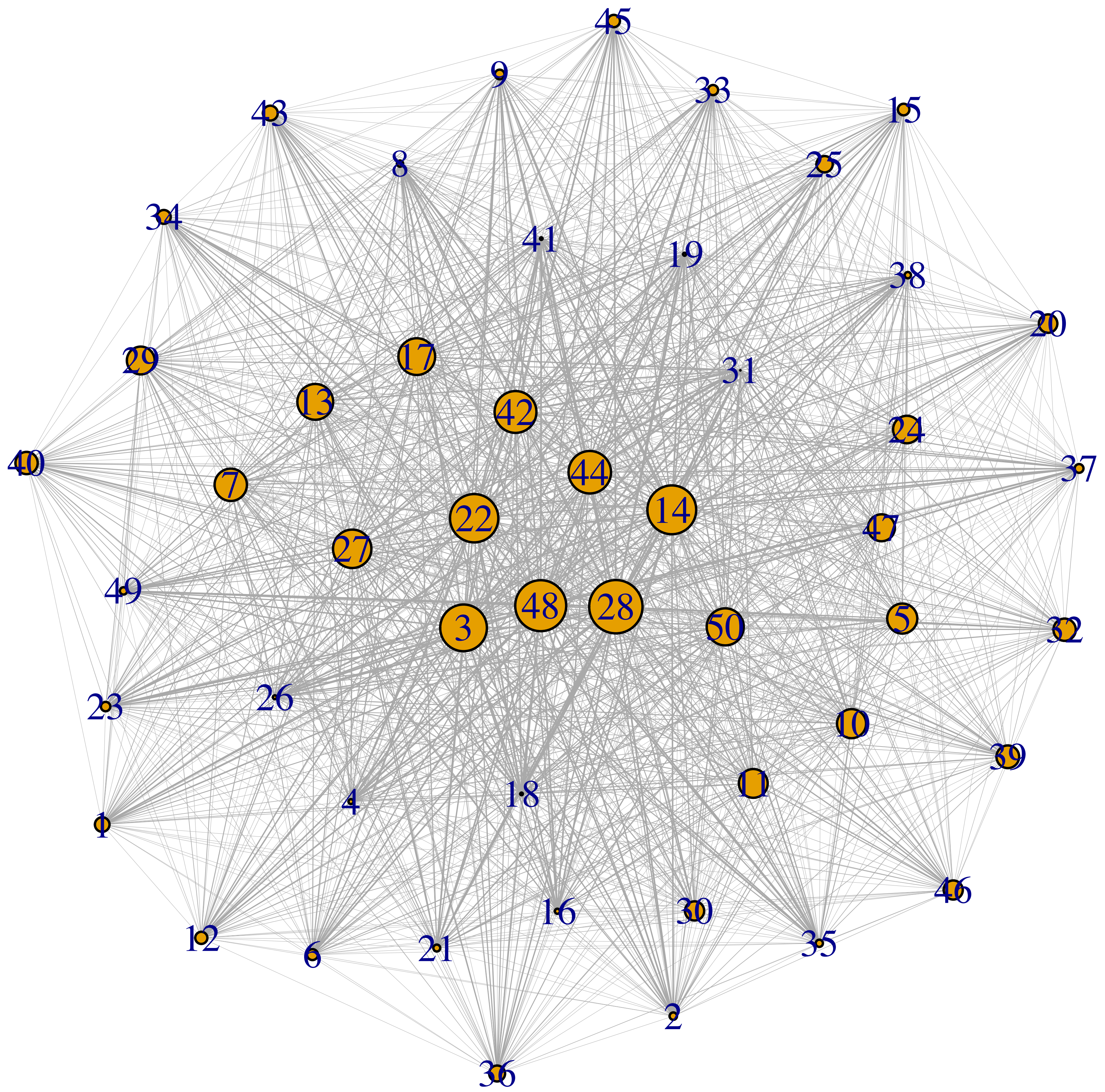
$$\mathbb{P}_\theta(Y = y) = \prod_{i<j} \left(\frac{\exp\{\theta^\top x_{ij}\}}{1 + \exp\{\theta^\top x_{ij}\}} \right)^{y_{ij}} \left(\frac{1}{1 + \exp\{\theta^\top x_{ij}\}} \right)^{1-y_{ij}}$$

Let's add covariates $x_{ij,q} = |x_i - x_j|$!

$\Rightarrow$ π_{ij} now changes with different values of x_{ij}

This is the likelihood of a logistic regression!

- $\theta_q > 0$: Higher values of $x_{ij,q}$ make $Y_{ij} = 1$ more likely
- $\theta_q < 0$: Higher values of $x_{ij,q}$ make $Y_{ij} = 1$ less likely



ER Model with Covariates

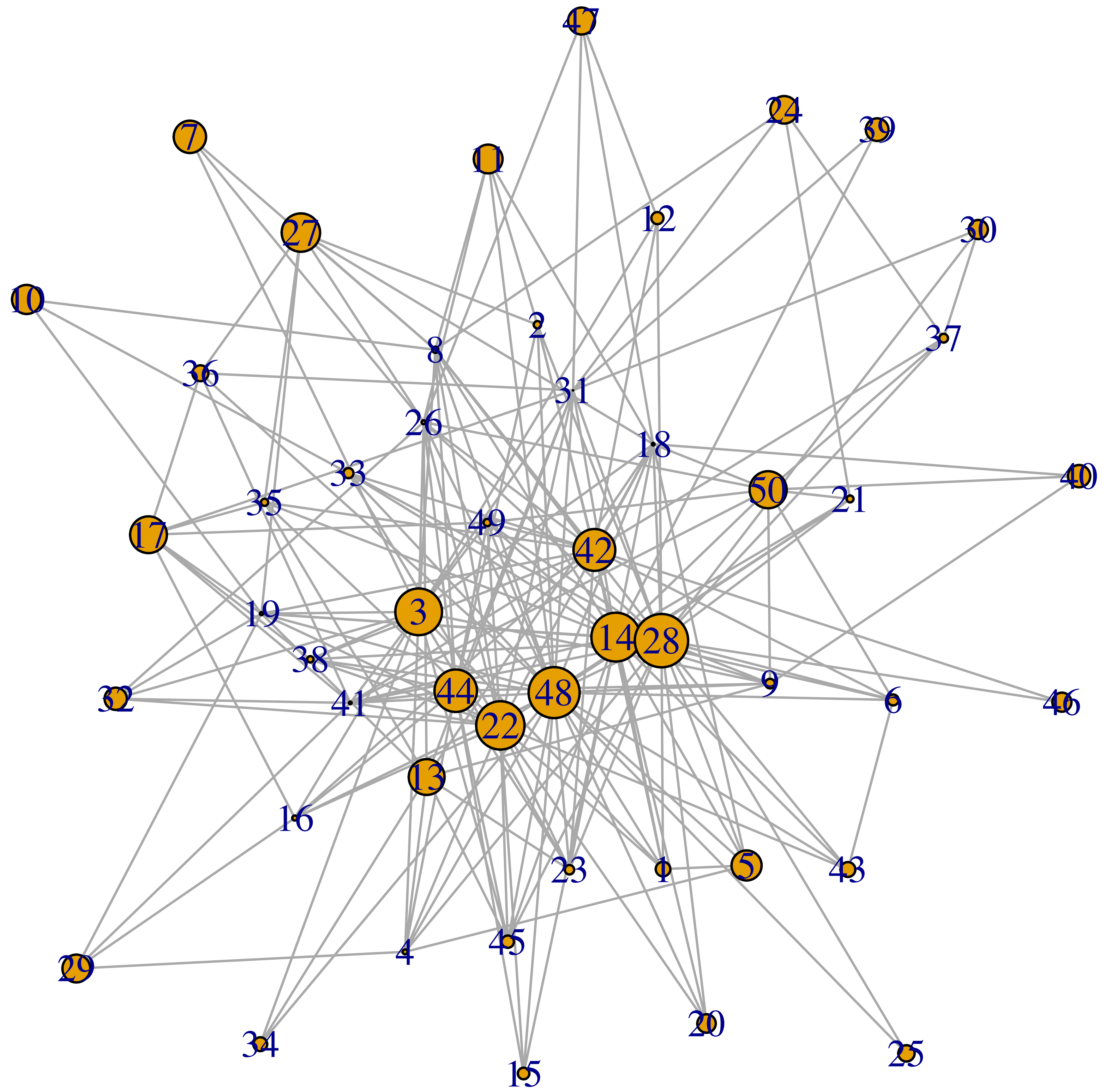
$$\mathbb{P}_\theta(Y = y) = \prod_{i<j} \left(\frac{\exp\{\theta^\top x_{ij}\}}{1 + \exp\{\theta^\top x_{ij}\}} \right)^{y_{ij}} \left(\frac{1}{1 + \exp\{\theta^\top x_{ij}\}} \right)^{1-y_{ij}}$$

Let's add covariates $x_{ij,q} = |x_i - x_j|$!

$\Rightarrow$ π_{ij} now changes with different values of x_{ij}

This is the likelihood of a logistic regression!

- $\theta_q > 0$: Higher values of $x_{ij,q}$ make $Y_{ij} = 1$ more likely
- $\theta_q < 0$: Higher values of $x_{ij,q}$ make $Y_{ij} = 1$ less likely



ER Model with Covariates

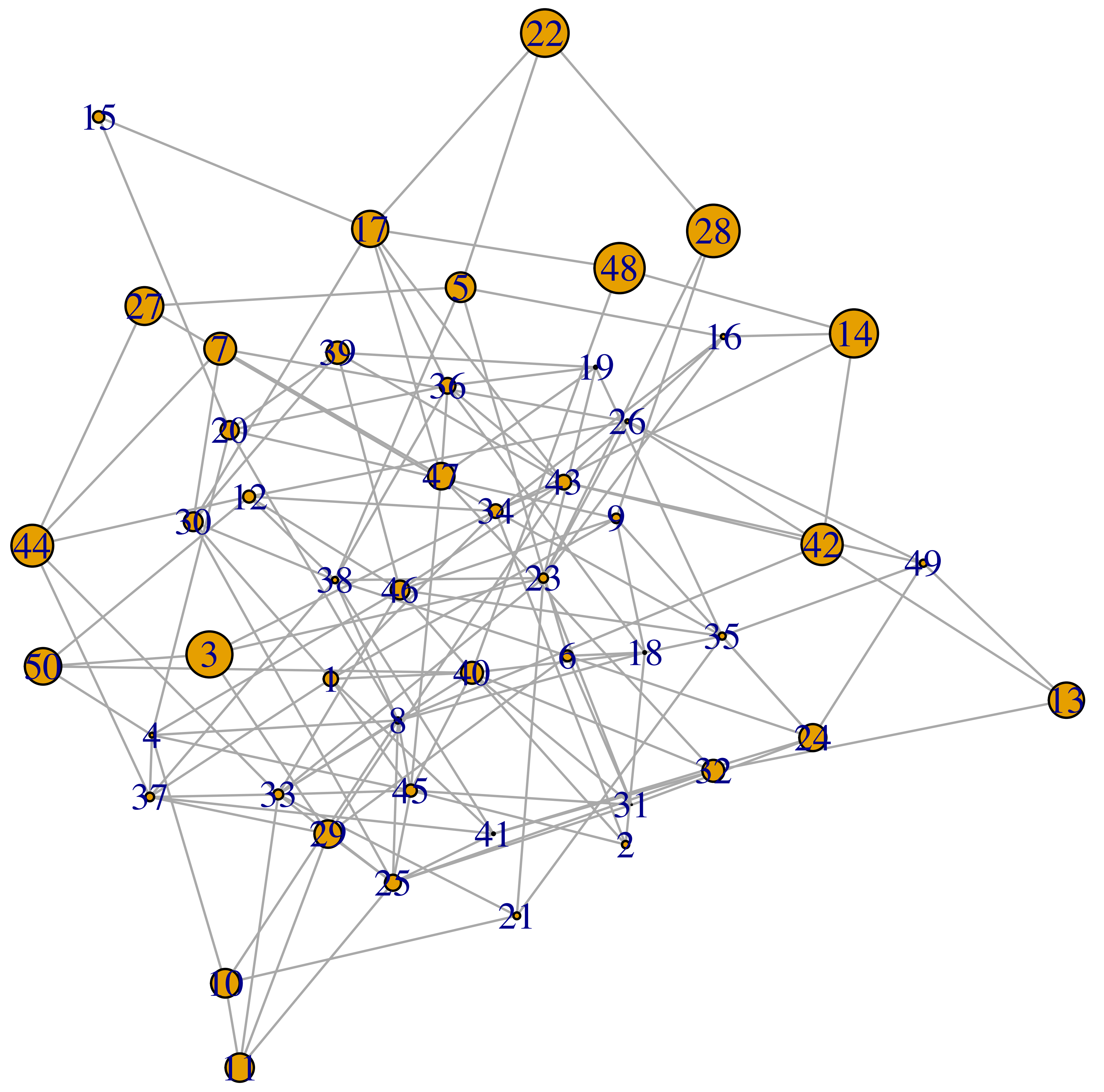
$$\mathbb{P}_\theta(Y = y) = \prod_{i<j} \left(\frac{\exp\{\theta^\top x_{ij}\}}{1 + \exp\{\theta^\top x_{ij}\}} \right)^{y_{ij}} \left(\frac{1}{1 + \exp\{\theta^\top x_{ij}\}} \right)^{1-y_{ij}}$$

Let's add covariates $x_{ij,q} = |x_i - x_j|$!

$\Rightarrow \pi_{ij}$ now changes with different values of x_{ij}

This is the likelihood of a logistic regression!

- $\theta_q > 0$: Higher values of $x_{ij,q}$ make $Y_{ij} = 1$ more likely
- $\theta_q < 0$: Higher values of $x_{ij,q}$ make $Y_{ij} = 1$ less likely



ER Model with Covariates

$$\mathbb{P}_\theta(Y = y) = \prod_{i<j} \left(\frac{\exp\{\theta^\top x_{ij}\}}{1 + \exp\{\theta^\top x_{ij}\}} \right)^{y_{ij}} \left(\frac{1}{1 + \exp\{\theta^\top x_{ij}\}} \right)^{1-y_{ij}}$$

Let's add covariates $x_{ij,q} = |x_i - x_j|$!

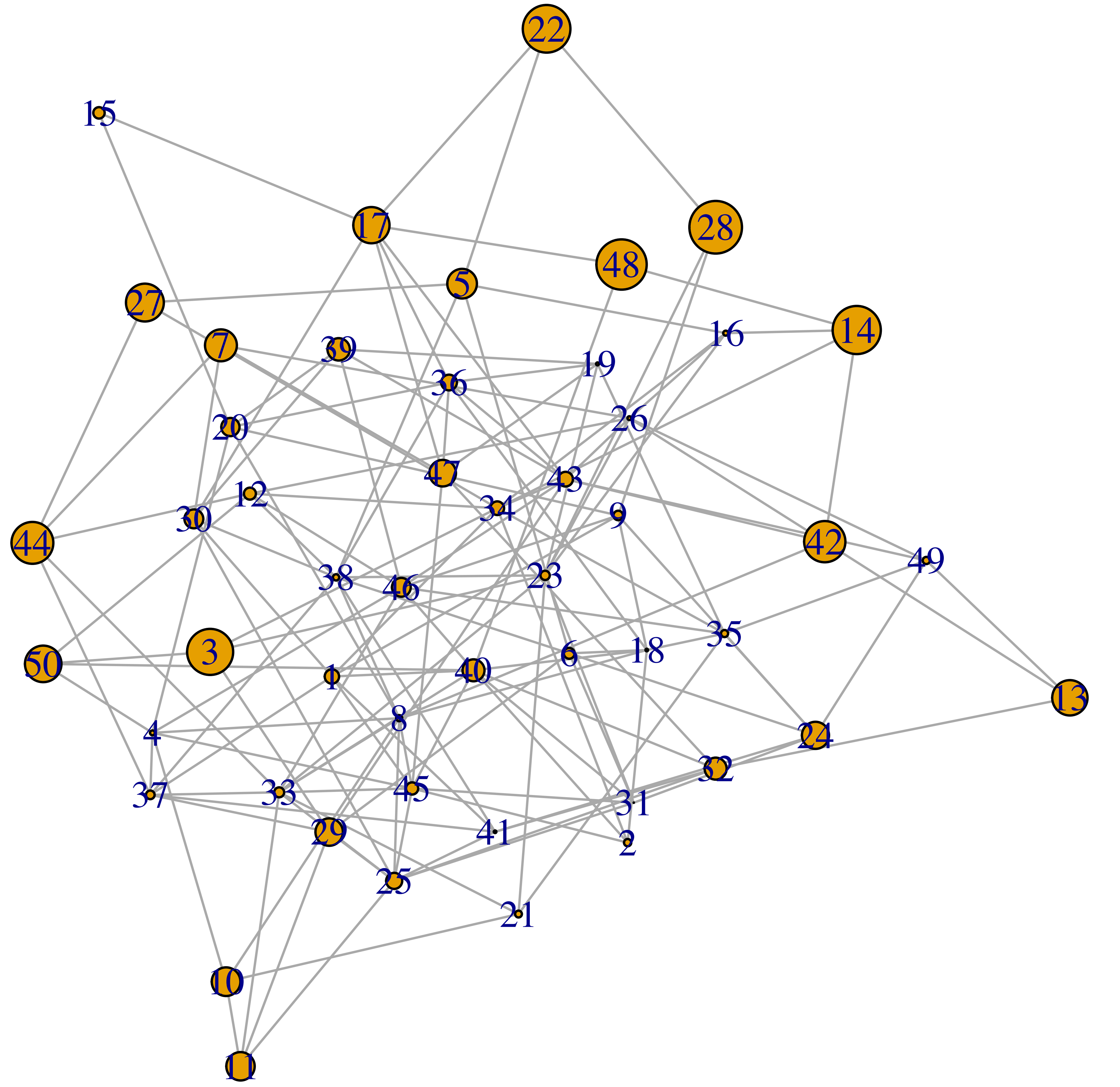
$\Rightarrow \pi_{ij}$ now changes with different values of x_{ij}

This is the likelihood of a logistic regression!

- $\theta_q > 0$: Higher values of $x_{ij,q}$ make $Y_{ij} = 1$ more likely
- $\theta_q < 0$: Higher values of $x_{ij,q}$ make $Y_{ij} = 1$ less likely

What would change for directed networks?

What if some edges are not possible?



ER Model with Covariates

$$\mathbb{P}_\theta(Y=y) = \frac{\exp\{\theta^\top s(y)\}}{\kappa(\theta)} \text{ with } s(y) = (s_1(y), \ldots, s_Q(y)) \text{ and } s_q(y) = \sum_{i<j} y_{ij} x_{ij,q}$$

Let's add covariates $x_{ij,q} = |x_i - x_j|$!

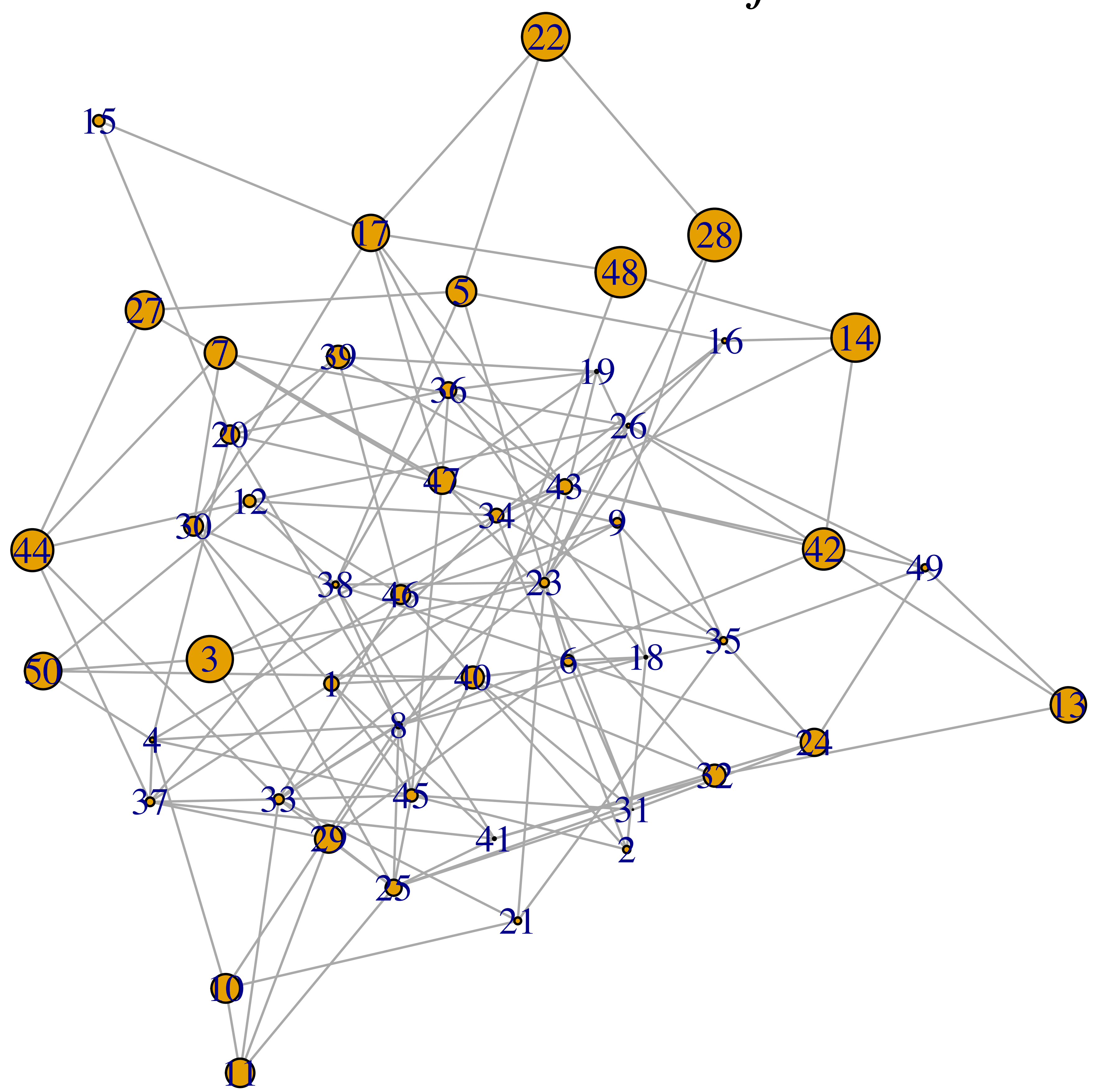
$\Rightarrow \pi_{ij}$ now changes with different values of x_{ij}

This is the likelihood of a logistic regression!

- $\theta_q > 0$: Higher values of $x_{ij,q}$ make $Y_{ij} = 1$ more likely
- $\theta_q < 0$: Higher values of $x_{ij,q}$ make $Y_{ij} = 1$ less likely

What would change for directed networks?

What if some edges are not possible?



Intermezzo

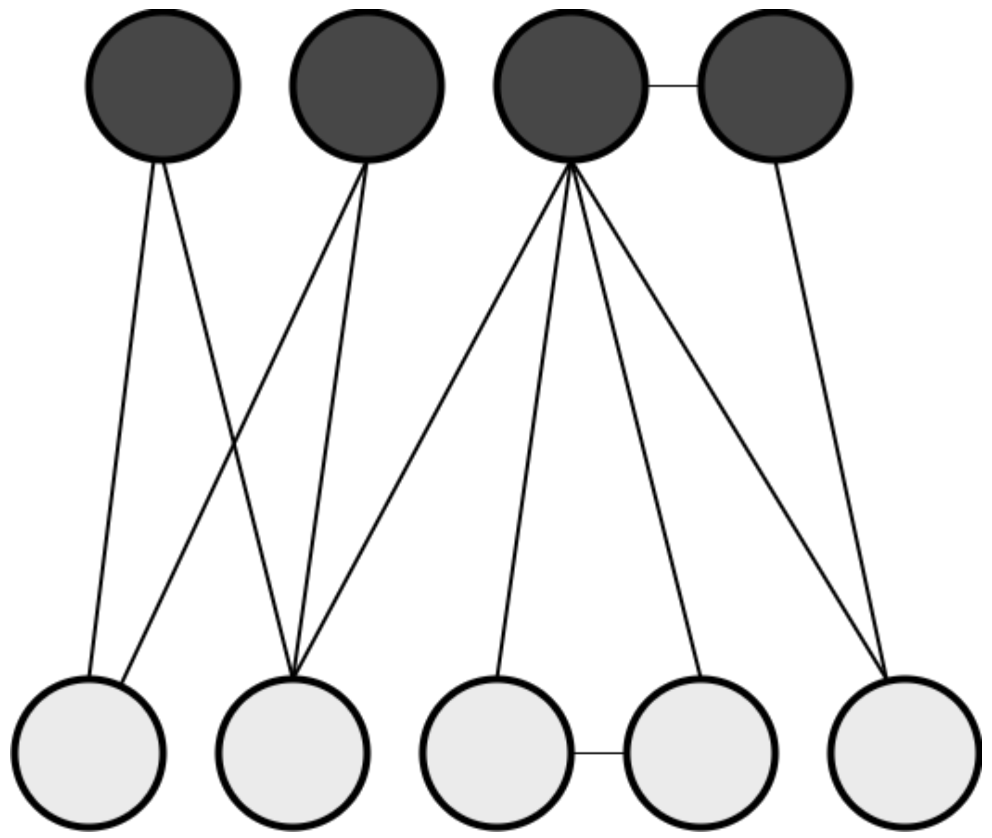
We can represent:

- Dense or sparse graphs with Bernoulli degree distributions
- Differential densities regarding covariates

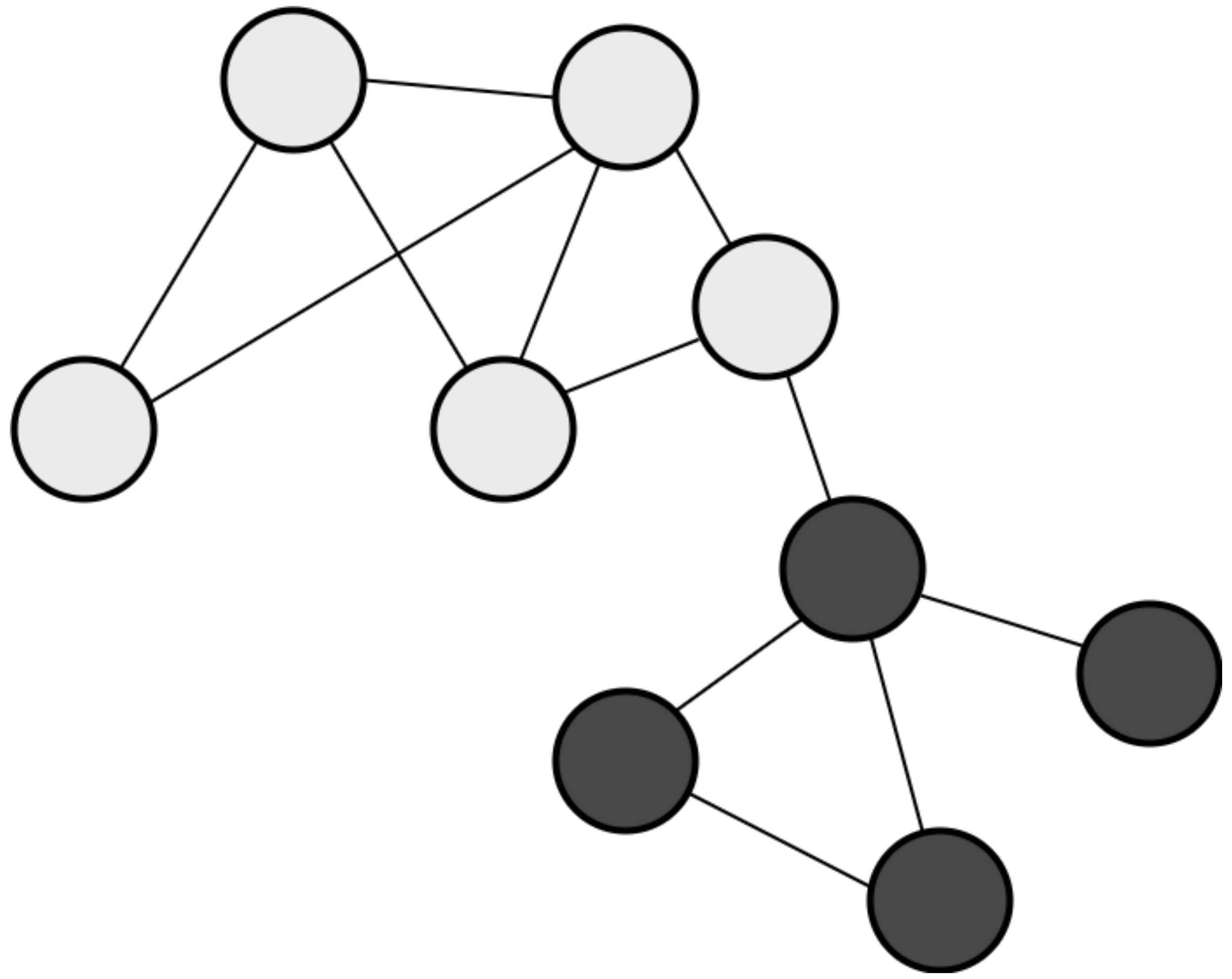
  ⇒ Homophily and heterophily
- How can we generalize this to “arbitrary” patterns?
- Via the sufficient statistics $s(y)$

What does this allow us?

- Test structural hypothesis and compare alternative structural mechanisms
- Capture dependencies between edges in a network
- Aggregate local network patterns to global statistics



Heterophily

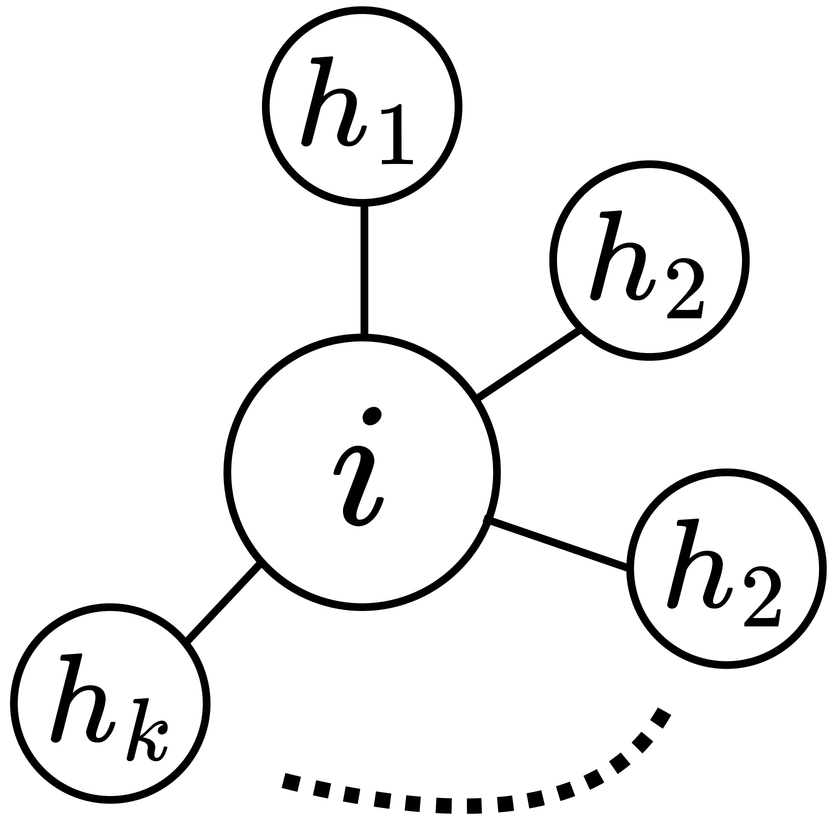


Homophily

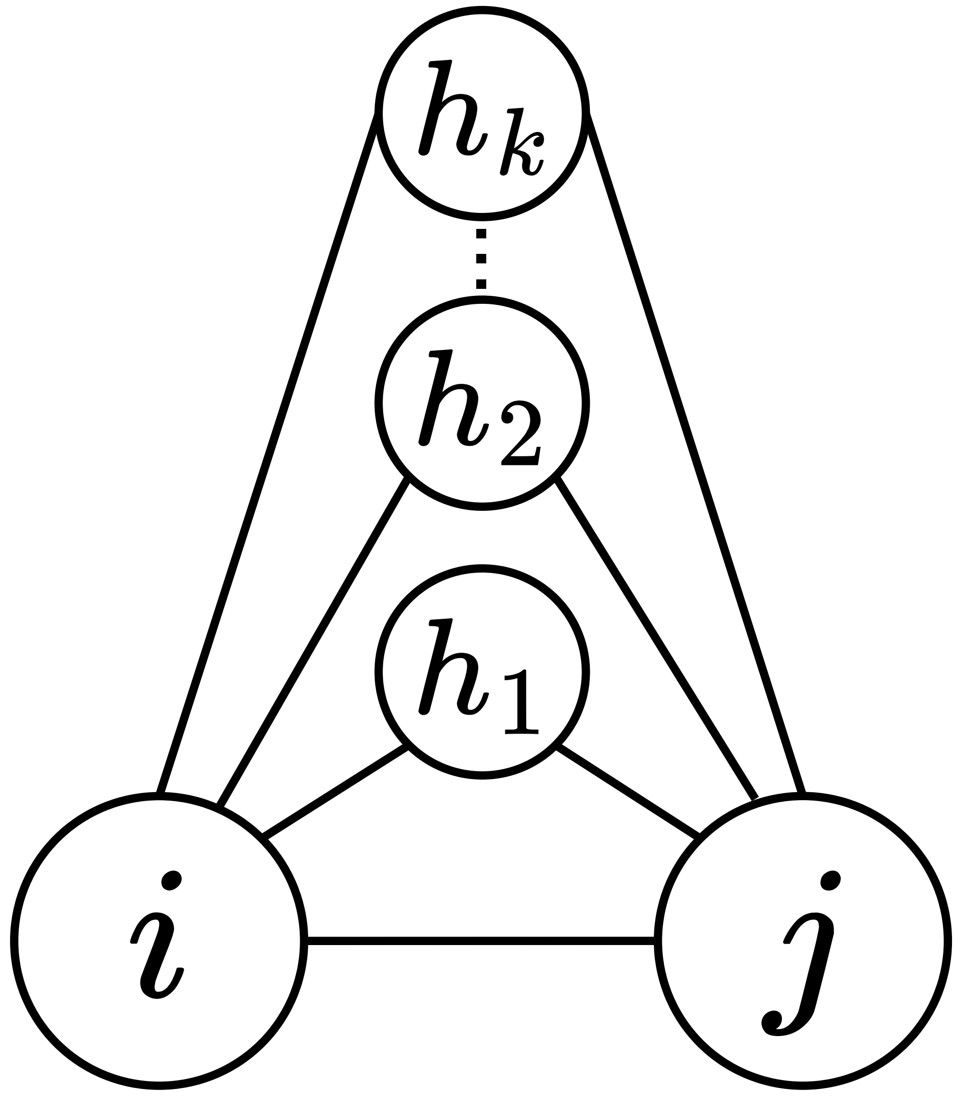
ERGM Formulation

$$\mathbb{P}_\theta(Y = y) = \frac{\exp\{\theta^\top s(y)\}}{\kappa(\theta)}$$

- $\theta \in \mathbb{R}^p$ are parameters to be estimated
- $s : \mathcal{Y} \to \mathbb{R}^p$ is a function calculating the vector of sufficient statistics for any network in $\mathcal{Y}$
- $\kappa(\theta)$ is a normalizing constant



Actors with Degree k

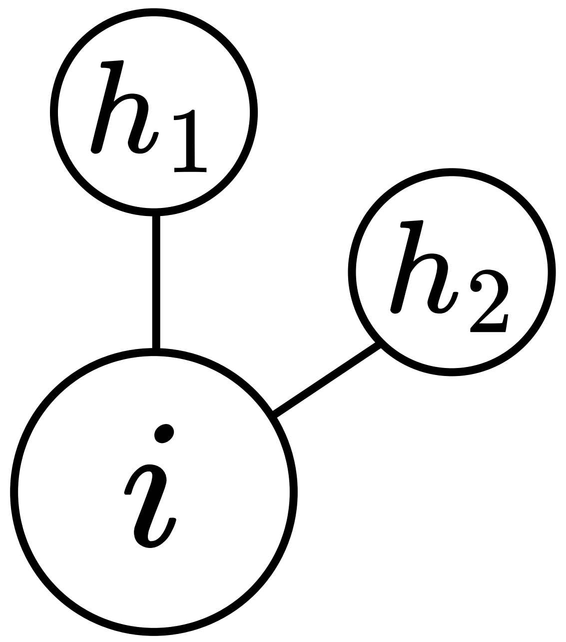


k Edgewise Shared Partners

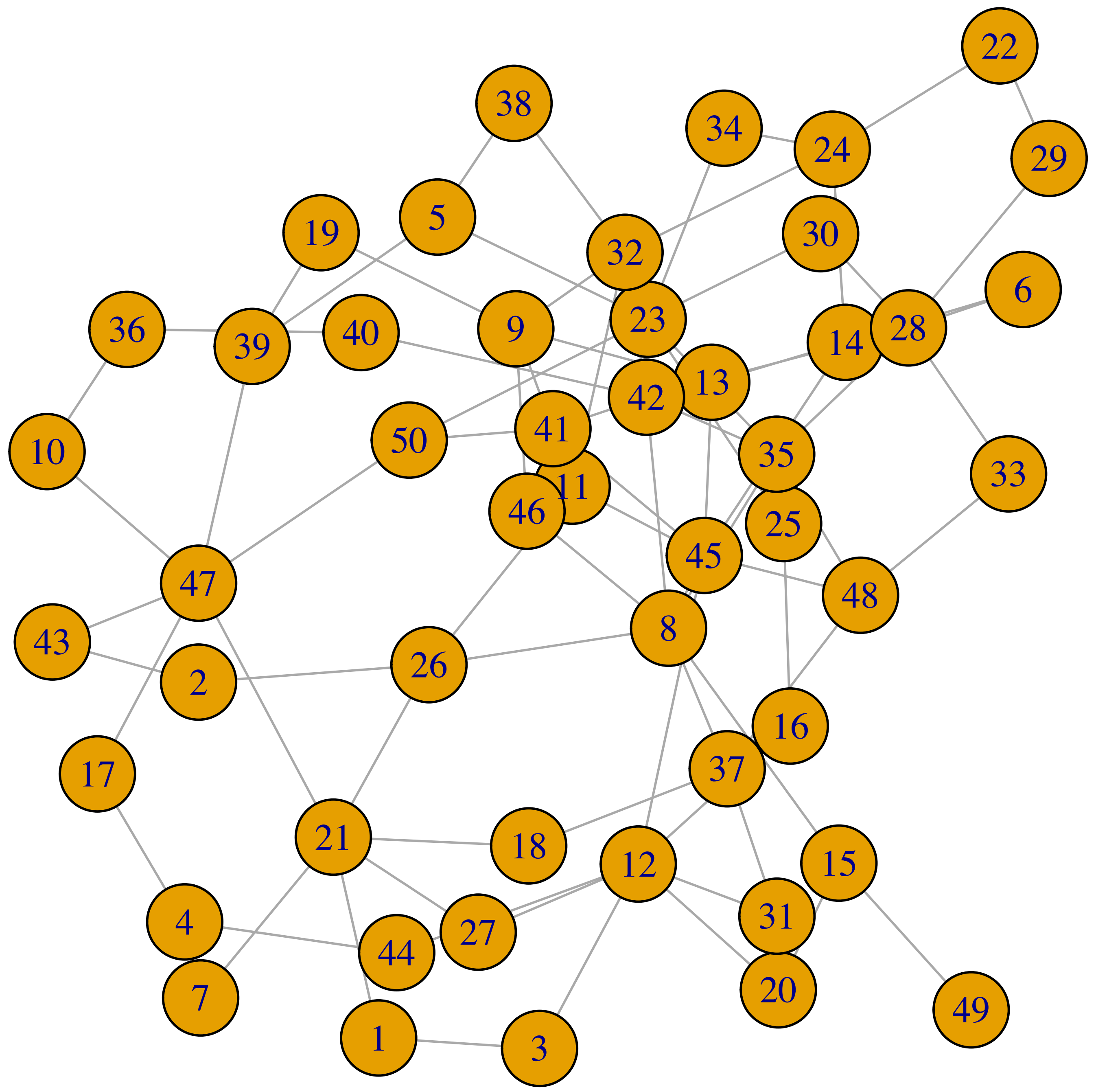
ERGM Formulation

$$\mathbb{P}_\theta(Y = y) = \frac{\exp\{\theta^\top s(y)\}}{\kappa(\theta)}$$

- $\theta \in \mathbb{R}^p$ are parameters to be estimated
- $s : \mathcal{Y} \to \mathbb{R}^p$ is a function calculating the vector of sufficient statistics for any network in $\mathcal{Y}$
- $\kappa(\theta)$ is a normalizing constant



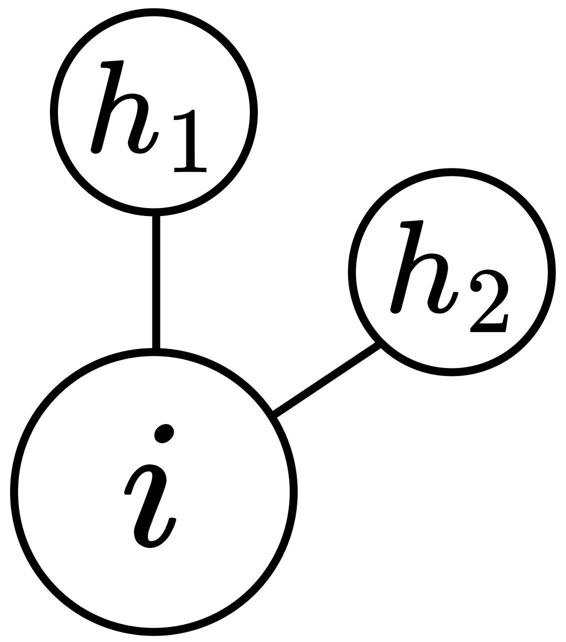
Actors with Degree 2



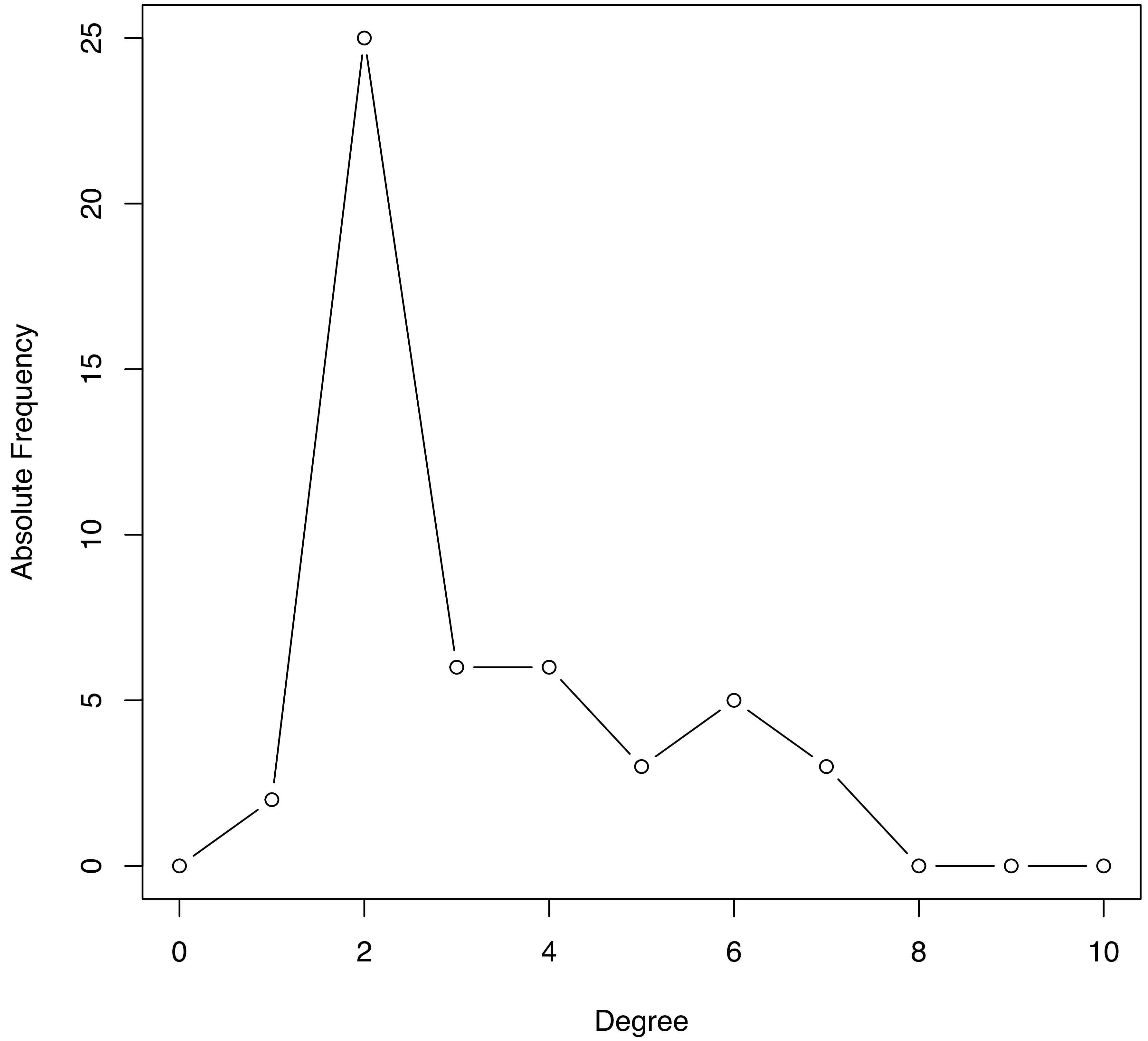
ERGM Formulation

$$\mathbb{P}_\theta(Y = y) = \frac{\exp\{\theta^\top s(y)\}}{\kappa(\theta)}$$

- $\theta \in \mathbb{R}^p$ are parameters to be estimated
- $s : \mathcal{Y} \to \mathbb{R}^p$ is a function calculating the vector of sufficient statistics for any network in $\mathcal{Y}$
- $\kappa(\theta)$ is a normalizing constant



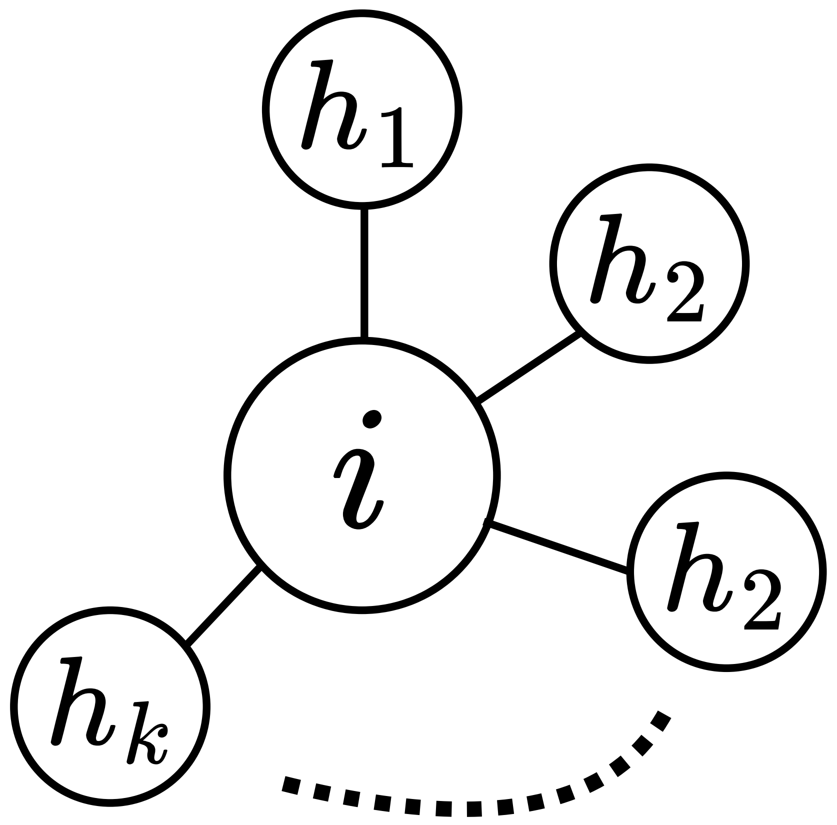
Actors with Degree 2



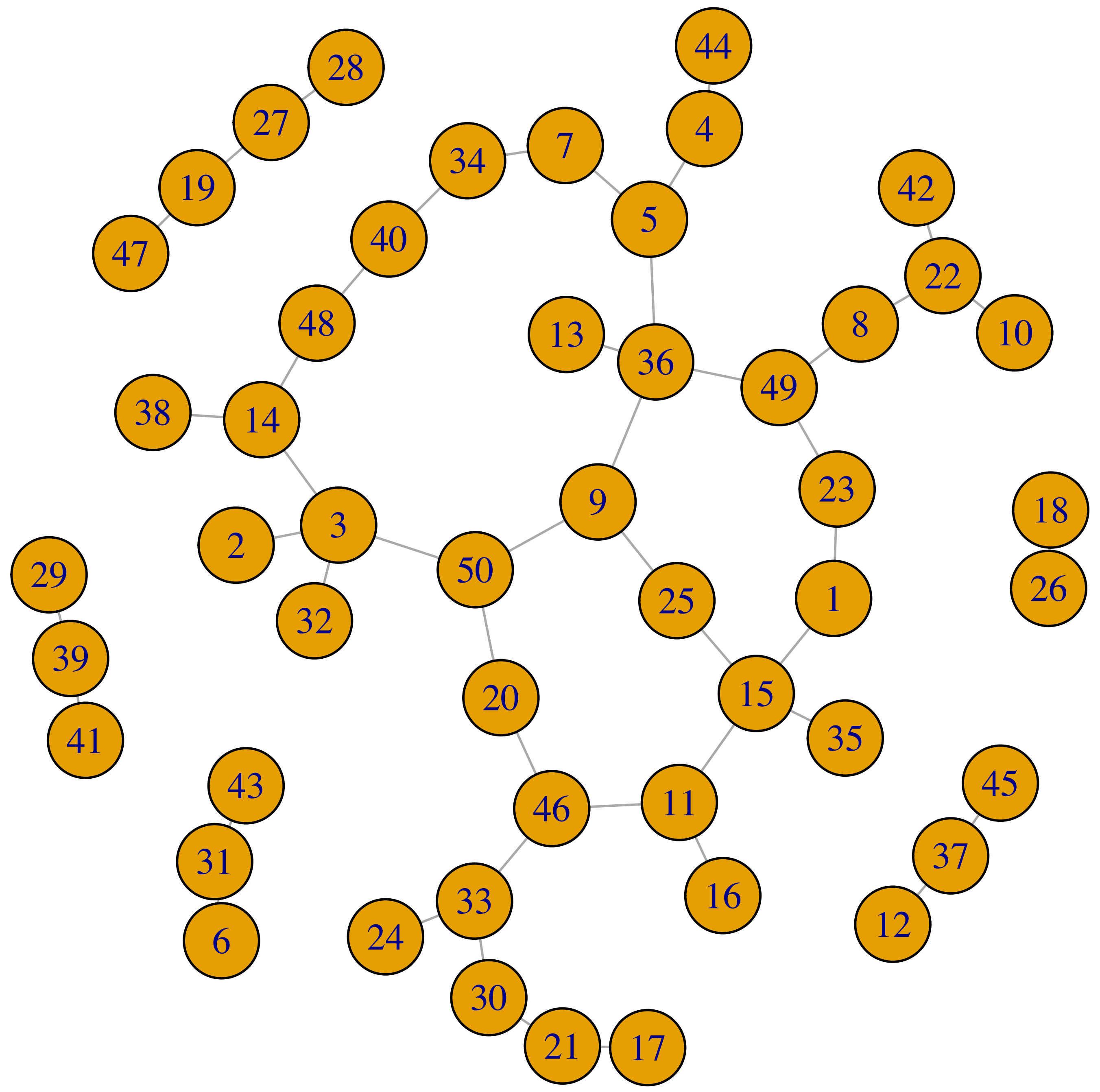
ERGM Formulation

$$\mathbb{P}_\theta(Y = y) = \frac{\exp\{\theta^\top s(y)\}}{\kappa(\theta)}$$

- $\theta \in \mathbb{R}^p$ are parameters to be estimated
- $s : \mathcal{Y} \to \mathbb{R}^p$ is a function calculating the vector of sufficient statistics for any network in $\mathcal{Y}$
- $\kappa(\theta)$ is a normalizing constant



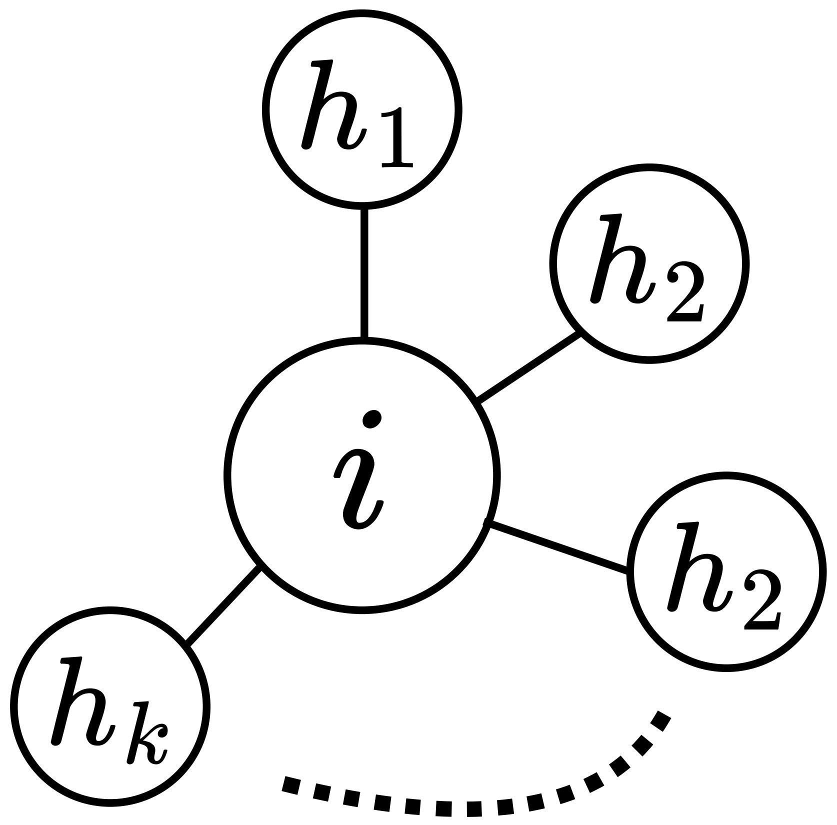
Actors with Degree 1 to 4



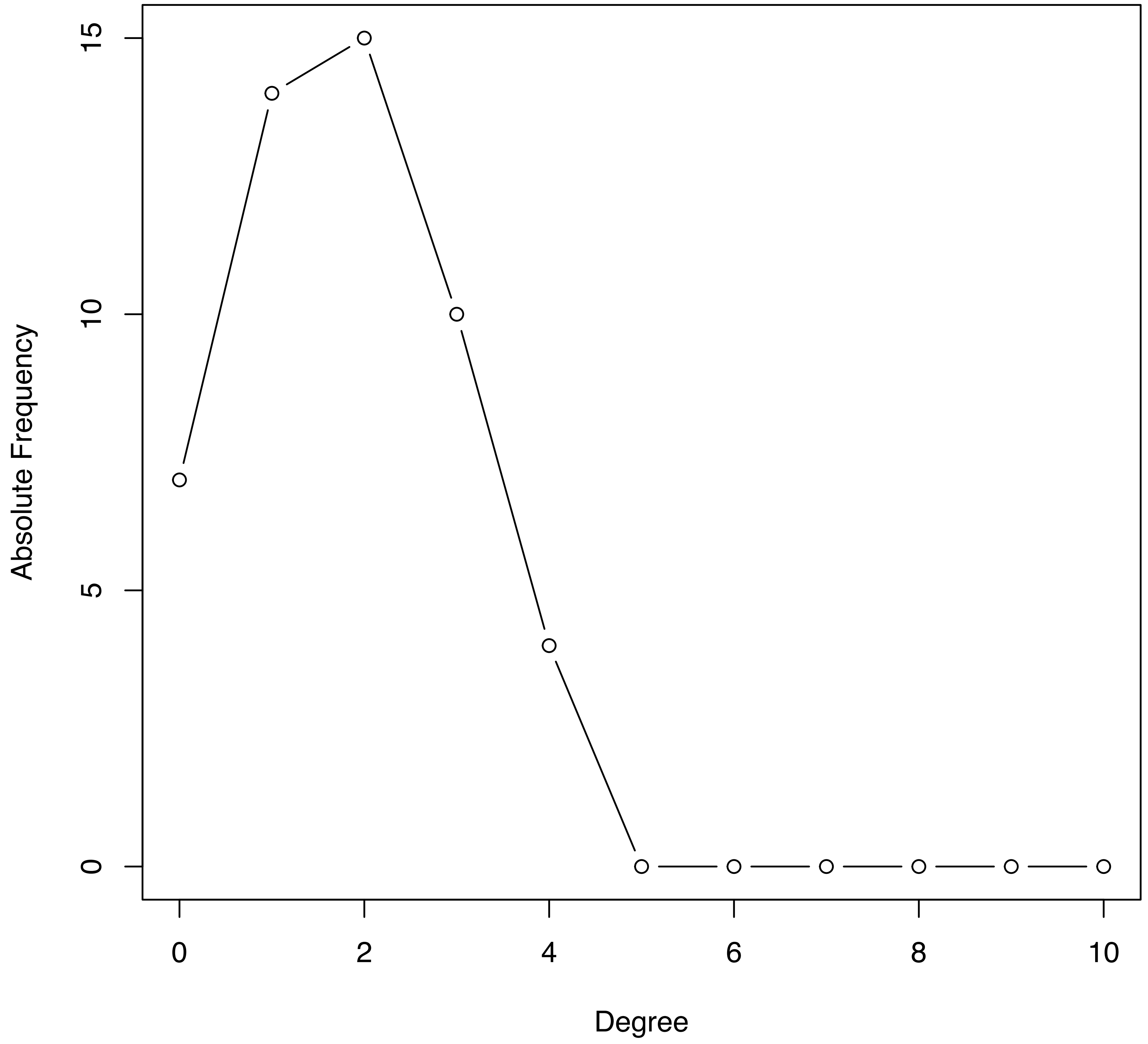
ERGM Formulation

$$\mathbb{P}_\theta(Y = y) = \frac{\exp\{\theta^\top s(y)\}}{\kappa(\theta)}$$

- $\theta \in \mathbb{R}^p$ are parameters to be estimated
- $s : \mathcal{Y} \rightarrow \mathbb{R}^p$ is a function calculating the vector of sufficient statistics for any network in $\mathcal{Y}$
- $\kappa(\theta)$ is a normalizing constant



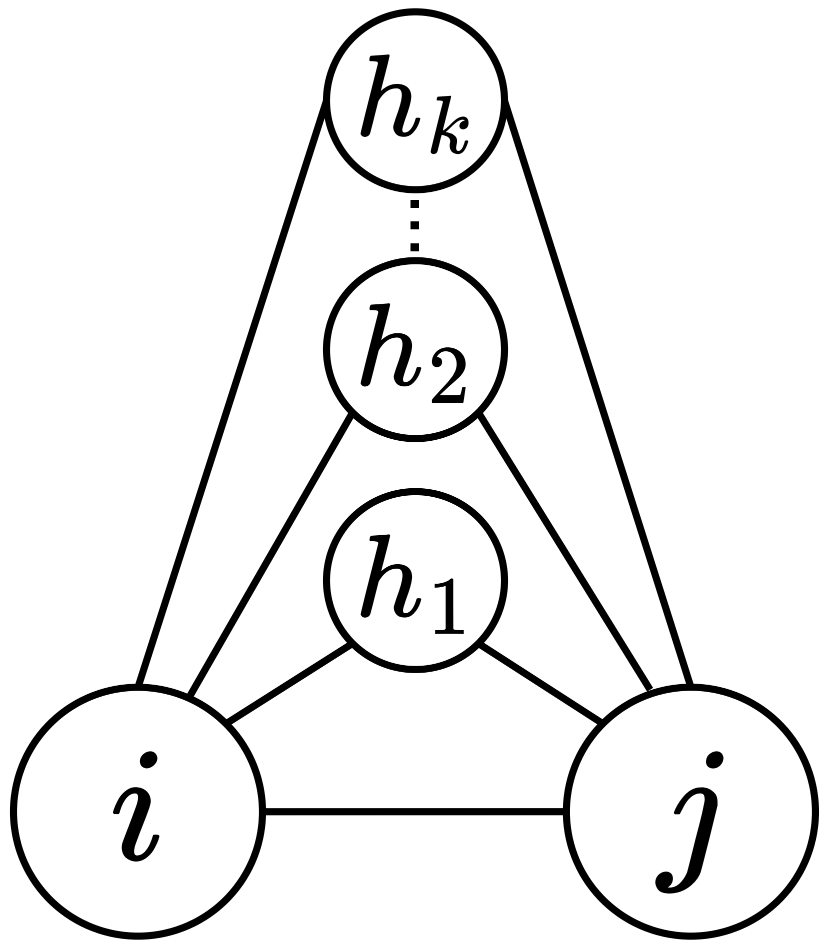
Actors with Degree 1 to 4



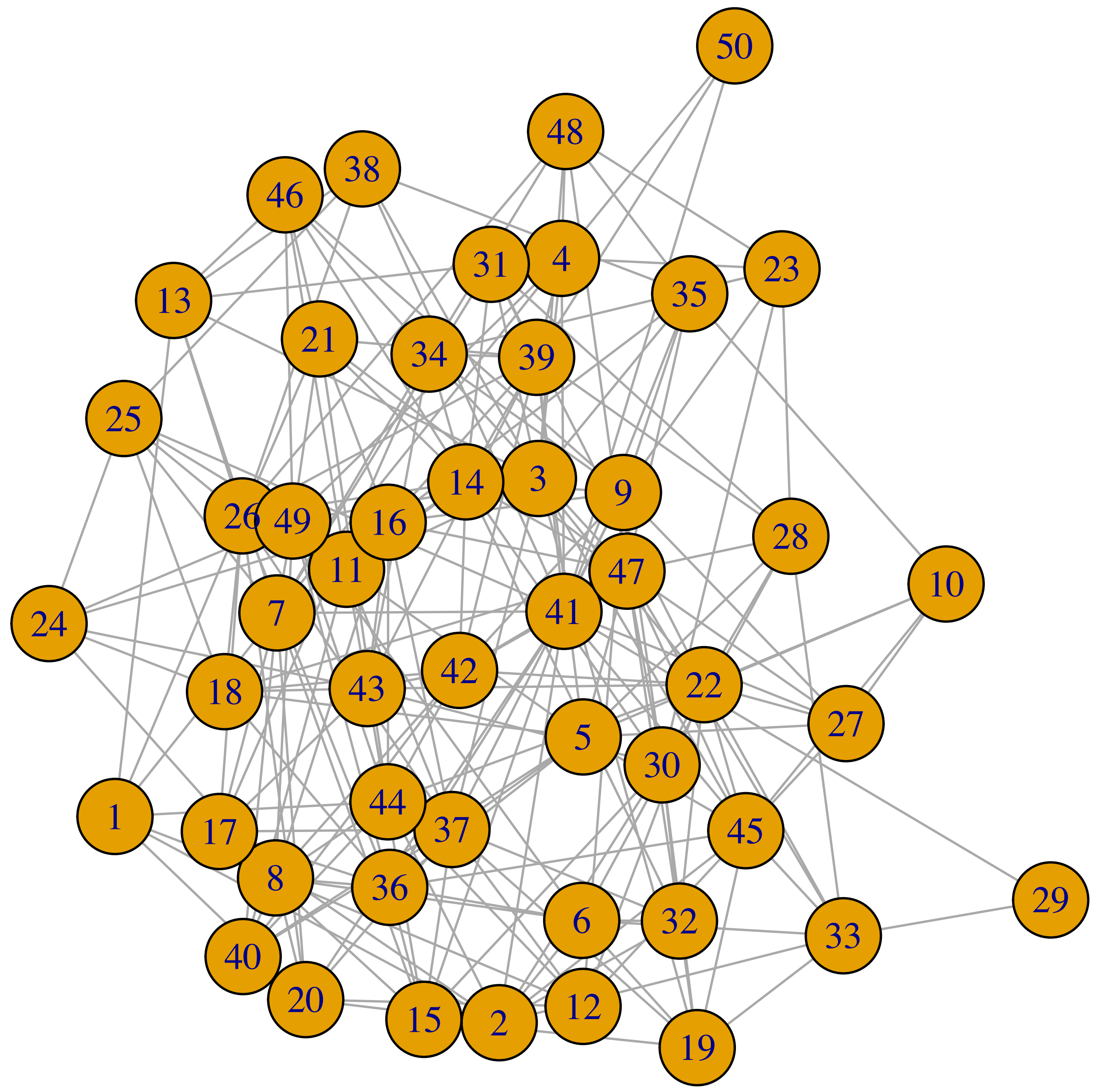
ERGM Formulation

$$\mathbb{P}_\theta(Y = y) = \frac{\exp\{\theta^\top s(y)\}}{\kappa(\theta)}$$

- $\theta \in \mathbb{R}^p$ are parameters to be estimated
- $s : \mathcal{Y} \to \mathbb{R}^p$ is a function calculating the vector of sufficient statistics for any network in $\mathcal{Y}$
- $\kappa(\theta)$ is a normalizing constant



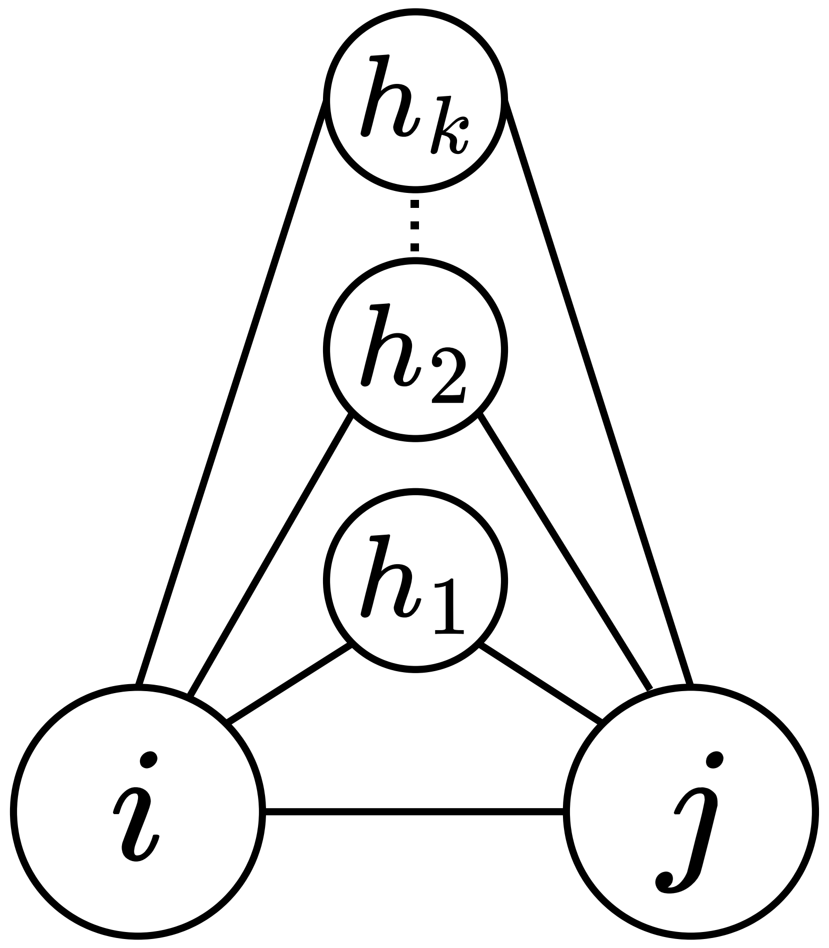
2 Edgewise Shared Partners



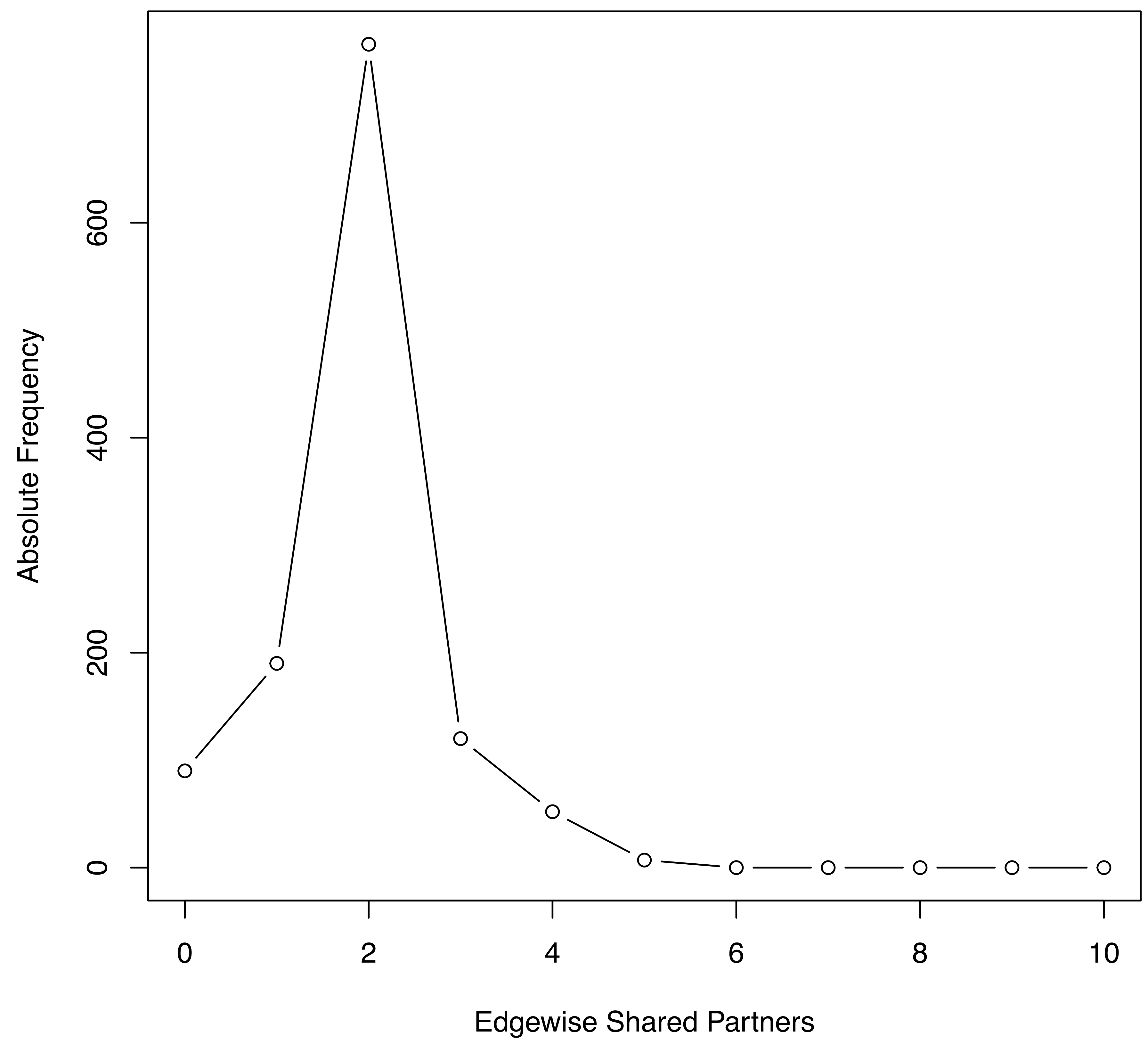
ERGM Formulation

$$\mathbb{P}_\theta(Y = y) = \frac{\exp\{\theta^\top s(y)\}}{\kappa(\theta)}$$

- $\theta \in \mathbb{R}^p$ are parameters to be estimated
- $s : \mathcal{Y} \to \mathbb{R}^p$ is a function calculating the vector of sufficient statistics for any network in $\mathcal{Y}$
- $\kappa(\theta)$ is a normalizing constant



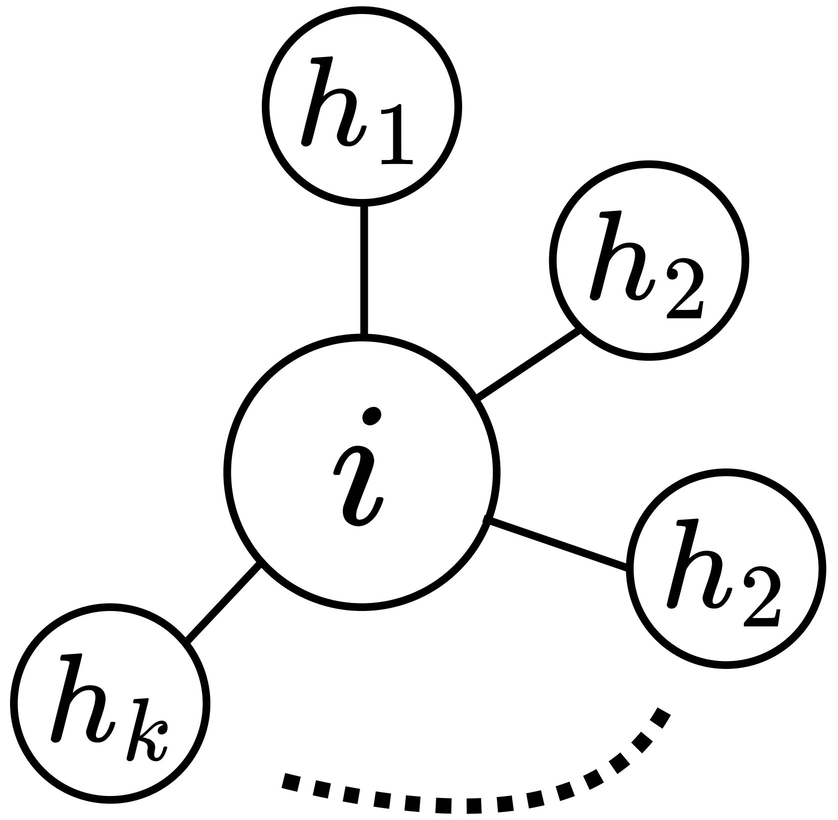
2 Edgewise Shared Partners



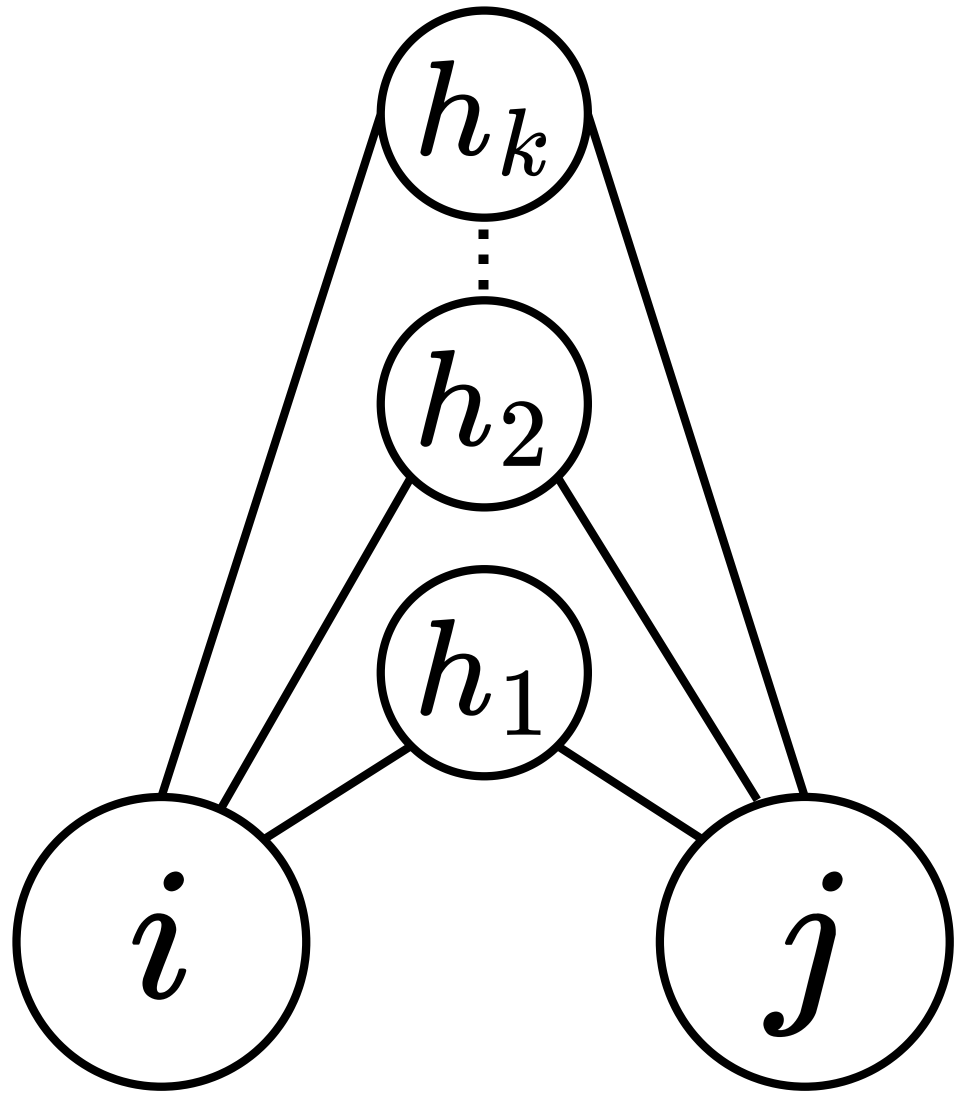
ERGM Formulation

$$\mathbb{P}_\theta(Y = y) = \frac{\exp\{\theta^\top s(y)\}}{\kappa(\theta)}$$

- $\theta \in \mathbb{R}^p$ are parameters to be estimated
- $s : \mathcal{Y} \to \mathbb{R}^p$ is a function calculating the vector of sufficient statistics for any network in $\mathcal{Y}$
- $\kappa(\theta)$ is a normalizing constant



Actors with Degree k

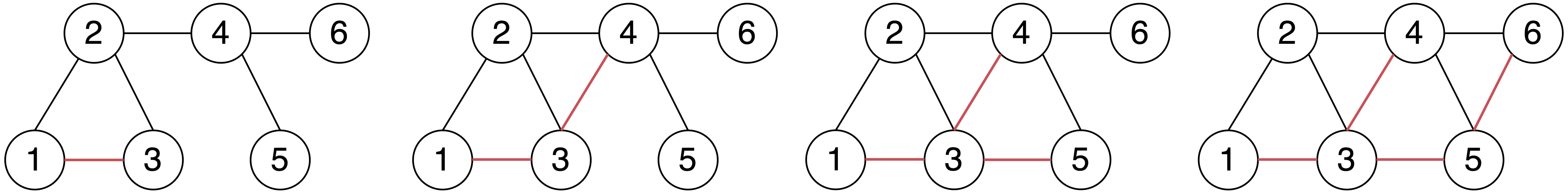


k Dyadwise Shared Partners

Global ERGM Interpretation

$$\mathbb{P}_\theta(Y = y) = \frac{\exp\{\theta^\top s(y)\}}{\kappa(\theta)} = \frac{\exp\{\sum_{q=1}^{Q} \theta_q s_q(y)\}}{\kappa(\theta)}$$

- $\theta_q > 0$: networks with increasing values of $s_q(y)$ are also increasingly more likely
- $\theta_q > 0$: networks with decreasing values of $s_q(y)$ are also increasingly more likely
- Take $\theta_{\text{Triangle}} = 0.5$ and $\theta_{\text{Edges}} = 0$

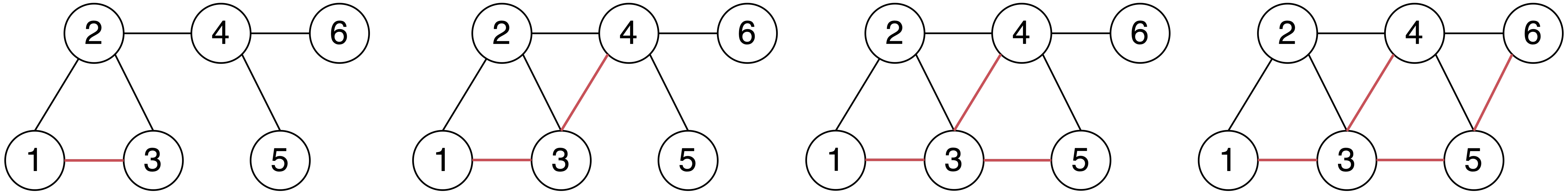


Triangles = Network Id	1	2	3	4
Edges	6	7	8	9
$\mathbb{P}_\theta(y^{Id})/\mathbb{P}_\theta(y^1)$	1	1.649	2.718	4.482

Global ERGM Interpretation

$$\mathbb{P}_\theta(Y = y) = \frac{\exp\{\theta^\top s(y)\}}{\kappa(\theta)} = \frac{\exp\{\sum_{q=1}^{Q} \theta_q s_q(y)\}}{\kappa(\theta)}$$

- $\theta_q > 0$: networks with increasing values of $s_q(y)$ are also increasingly more likely
- $\theta_q > 0$: networks with decreasing values of $s_q(y)$ are also increasingly more likely
- Take $\theta_{\text{Triangle}} = 0.5$ and $\theta_{\text{Edges}} = 0$



Triangles = Network Id	1	2	3	4
Edges	6	7	8	9
$\mathbb{P}_\theta(y^{Id})/\mathbb{P}_\theta(y^1)$	1	1.649	2.718	4.482

exp{0.5·1} = 1.649

Local ERGM Interpretation

$$\mathbb{P}_\theta(Y_{ij} = 1 \,|\, \text{Rest}) = \frac{\exp\{\theta^\top s(y_{ij}^1)\}}{1 + \exp\{\theta^\top s(y_{ij}^1)\}}$$

- Conditional distribution is a logistic regression model
- y_{ij}^1 is defined as y with the y_{ij} set to 1
- Tie-level interpretation akin to logistic regression in terms of conditional log-odds of y_{ij} to be 1 rather than 0

$$\log\left(\frac{\mathbb{P}_\theta(Y_{ij} = 1 \,|\, \text{Rest})}{\mathbb{P}_\theta(Y_{ij} = 0 \,|\, \text{Rest})}\right) = \theta^\top \underbrace{\left((s(y_{ij}^1) - s(y_{ij}^1)\right)}_{\text{Change statistic}}$$

  - if switching the value of y_{ij} from 0 to 1 raises only the qth entry of the change statistic by one, the conditional log-odds of Y_{ij} are changed by the additive factor θ_q

ML Estimation

$$\hat{\theta} = \text{argmax}_{\theta \in \mathbb{R}^Q} \frac{\exp\{\theta^\top s(y)\}}{\kappa(\theta)}$$

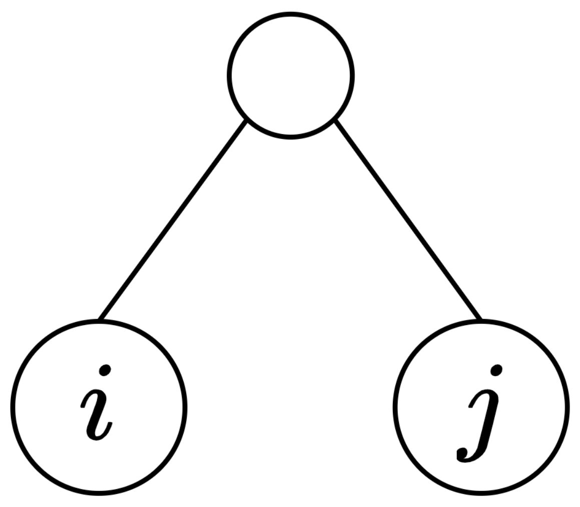
- How can we find $\hat{\theta}$?
- Problem: $\kappa(\theta) = \sum_{y \in \mathcal{Y}} \exp\{\theta^\top s(y)\}$ cannot be evaluated since $|\mathcal{Y}| = 2^{\binom{n}{2}}$
- Solution: For known θ_0 the logarithmic likelihood ratio of θ and θ_0 is:

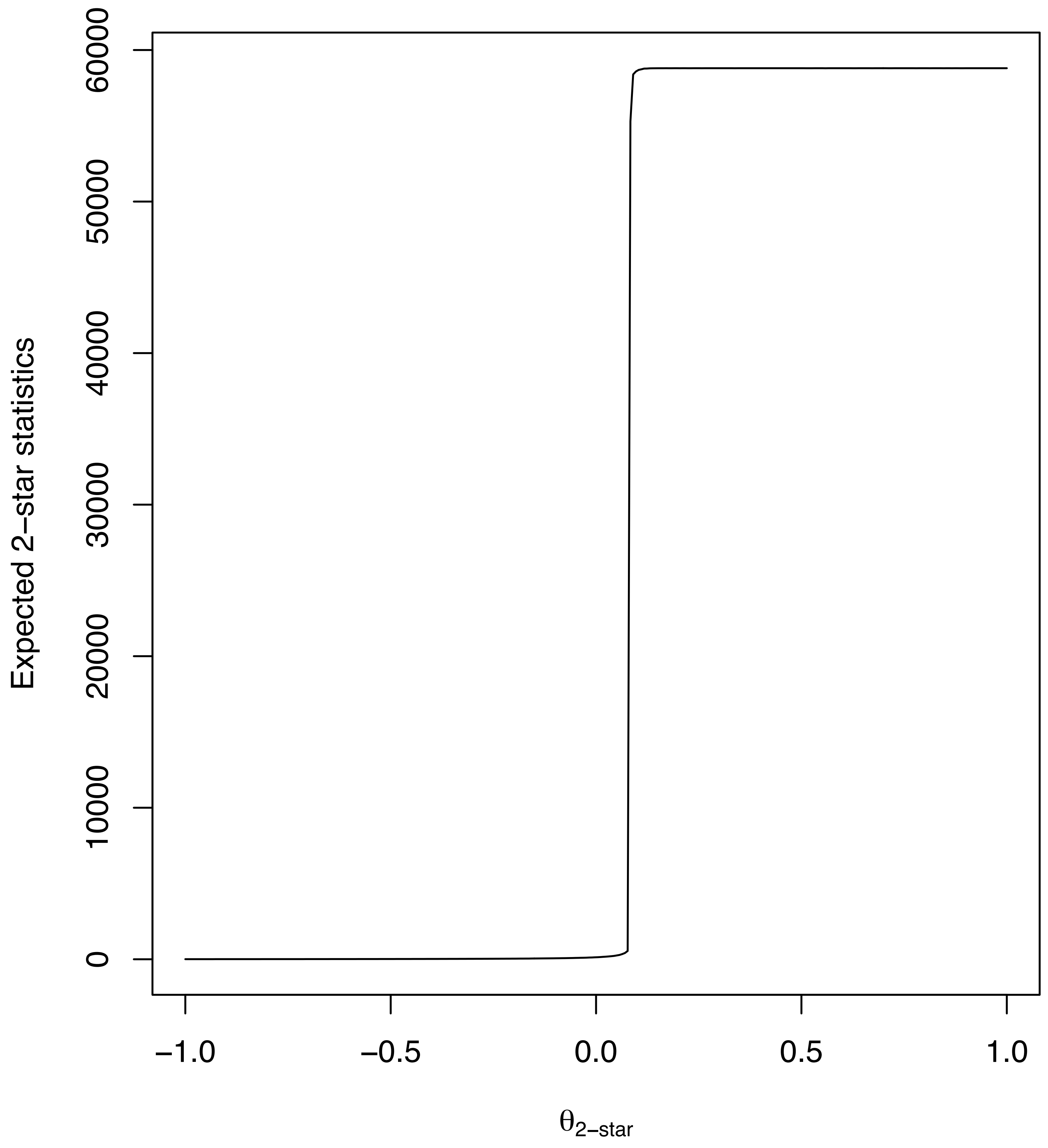
$$r(\theta, \theta_0) = (\theta - \theta_0)^\top s(y) - \log\left(\mathbb{E}_{\theta_0}\left(\exp\left\{(\theta - \theta_0)^\top s(Y)\right\}\right)\right)$$

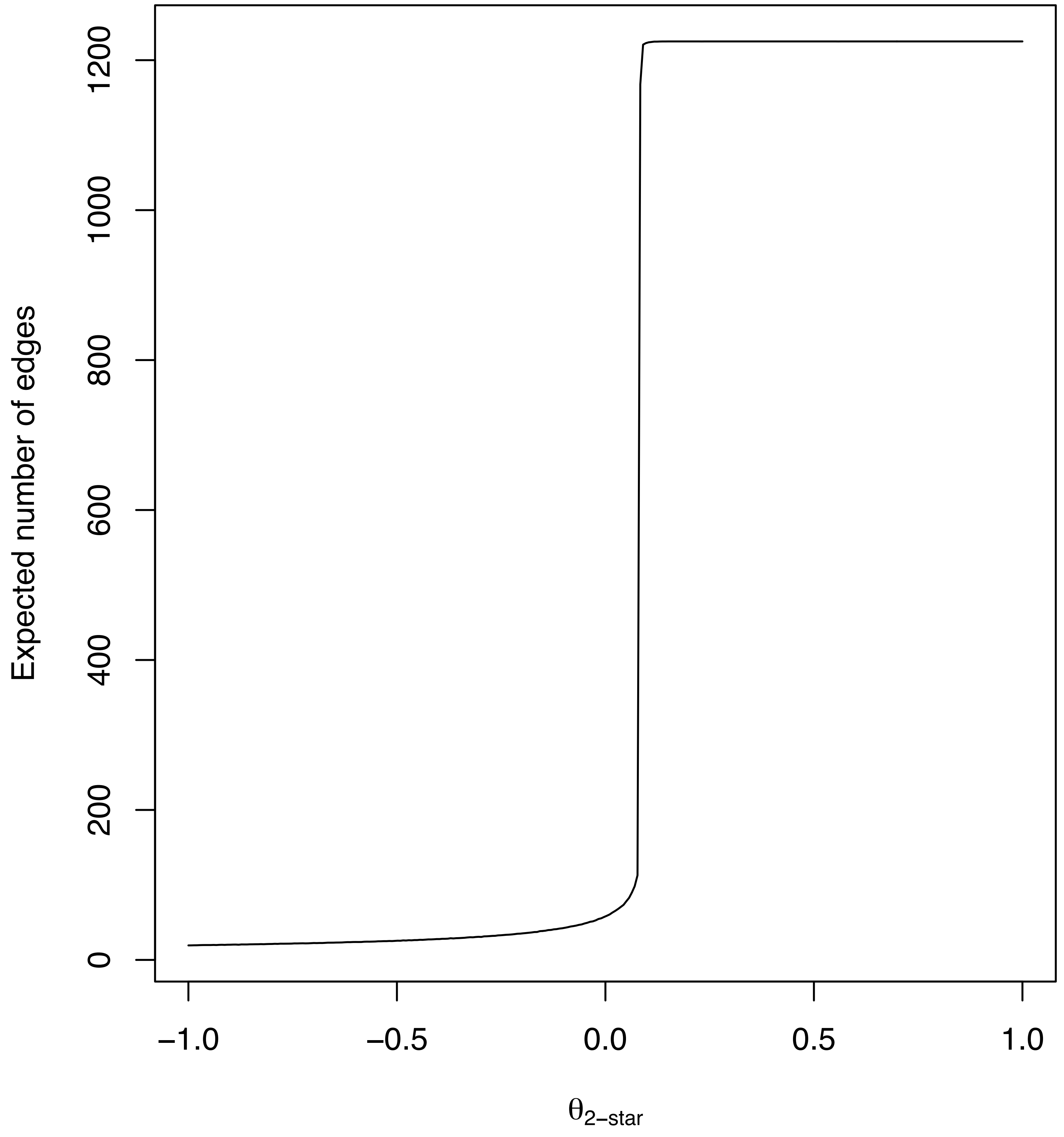
- We can approximate the expectation with MCMC samples
- How to sample from an ERGM?
  - We already derived the conditional distribution on the last slide
  - Get some proposal and the Gibbs sampling starts running!

Degeneracy Issues

- Problem 1: Some small changes to θ_q lead to big changes in $s_q(y)$
- Degeneracy or phase transition: Most probability is put on the full graph

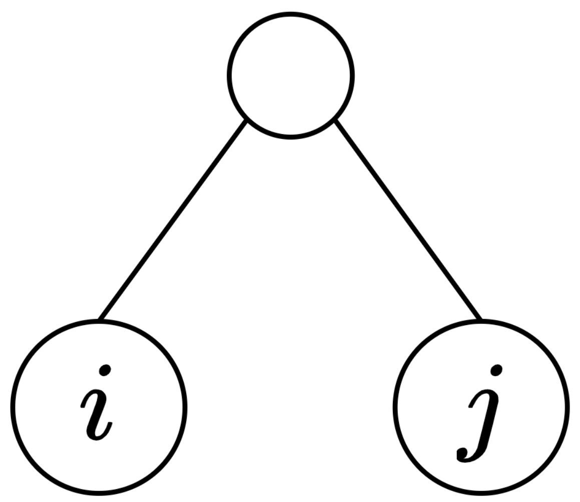




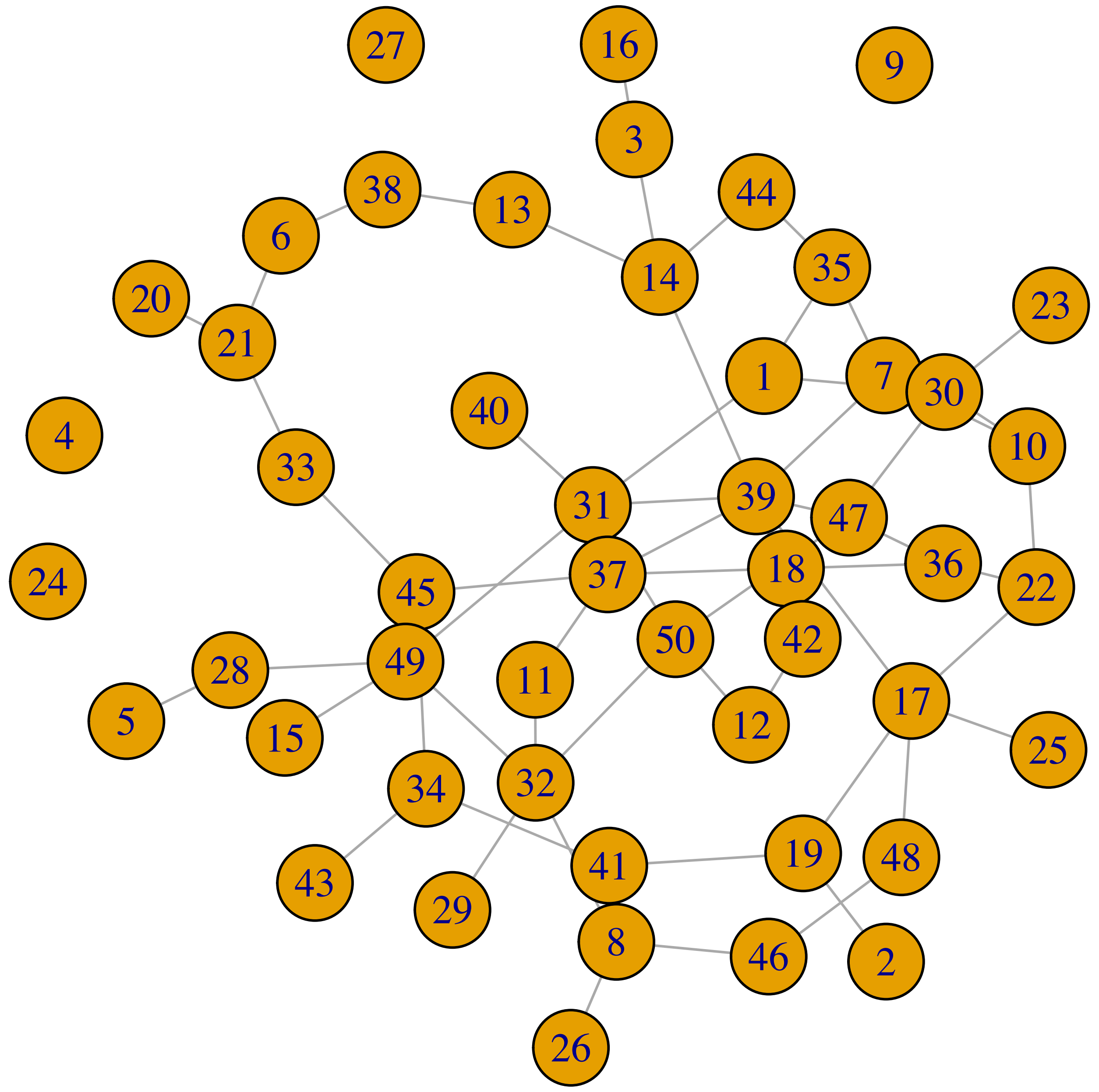


Degeneracy Issues

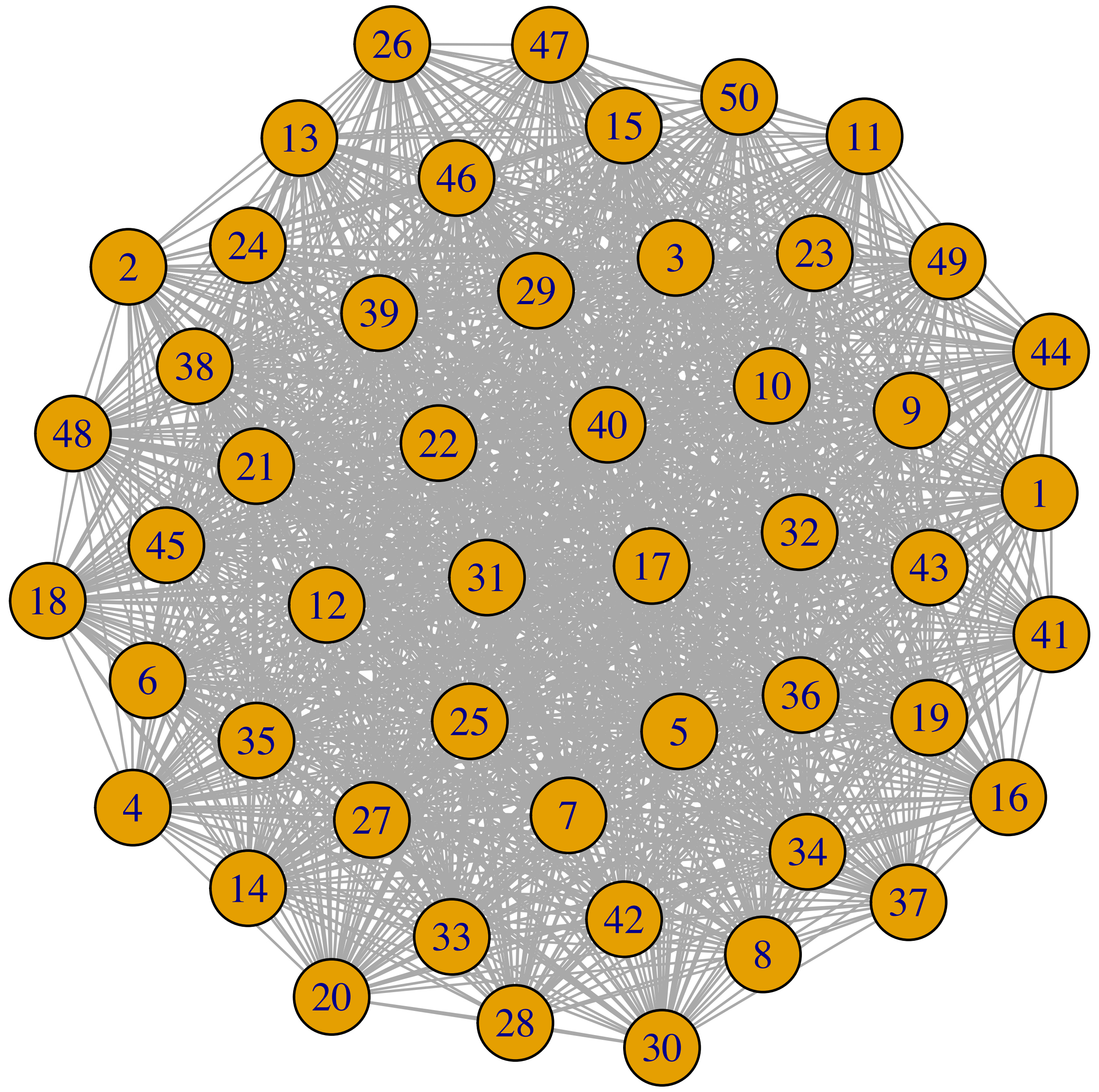
- Problem 1: Some small changes to θ_q lead to big changes in $s_q(y)$
- Degeneracy or phase transition: Most probability is put on the full graph



$\theta_{2\text{-star}} = 0.05$

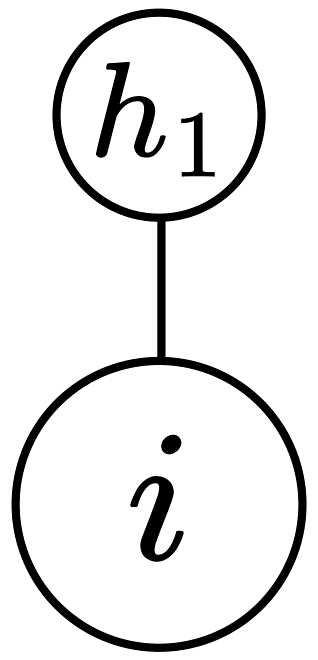


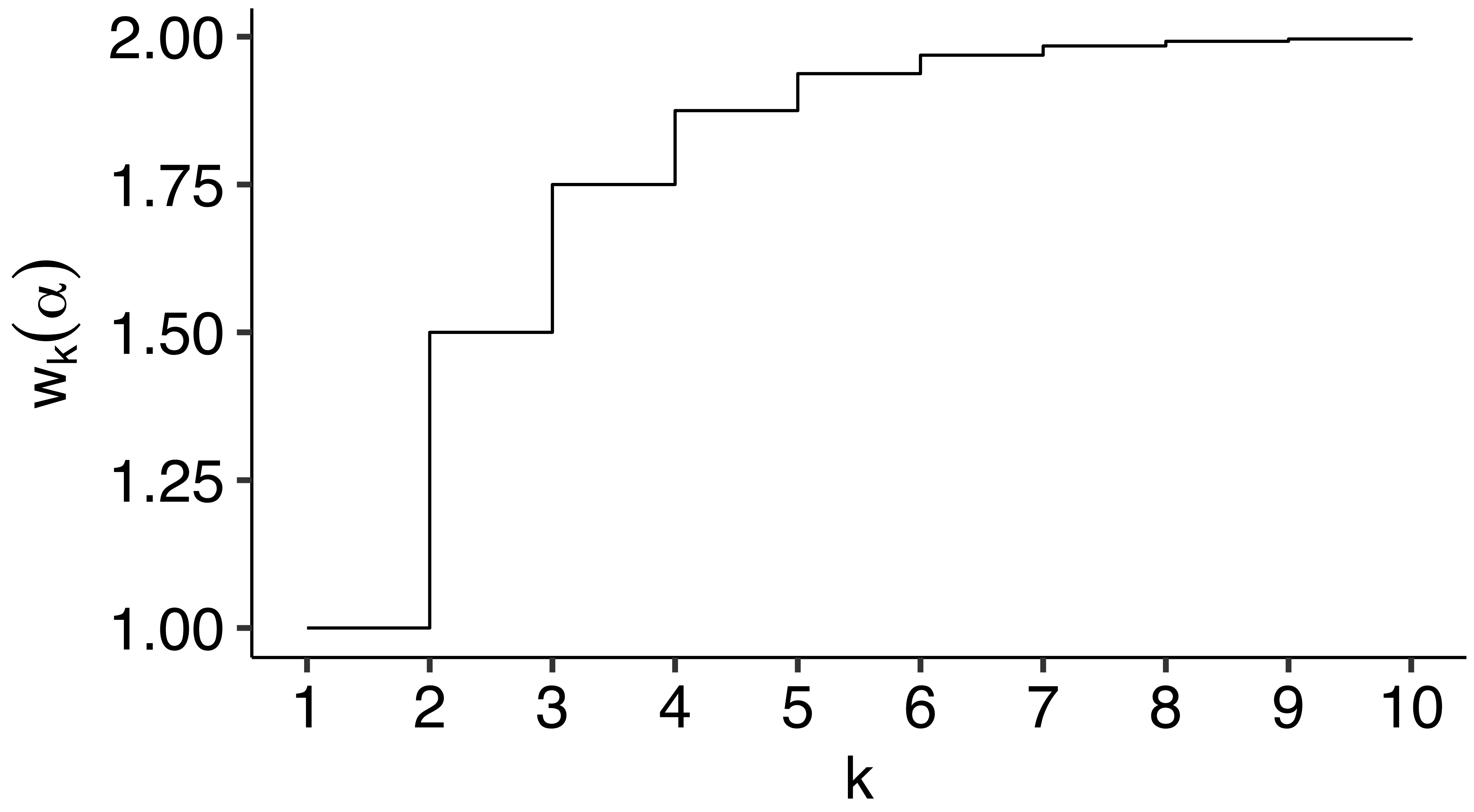
$\theta_{2\text{-star}} = 0.1$



Degeneracy Issues

- Problem 2: Which degree and triangular statistics to include?
  - There are n degree and shared partner statistics?
  - Incorporating all of them makes the model unstable
- Solution: Incorporate geometrically weighted statistics



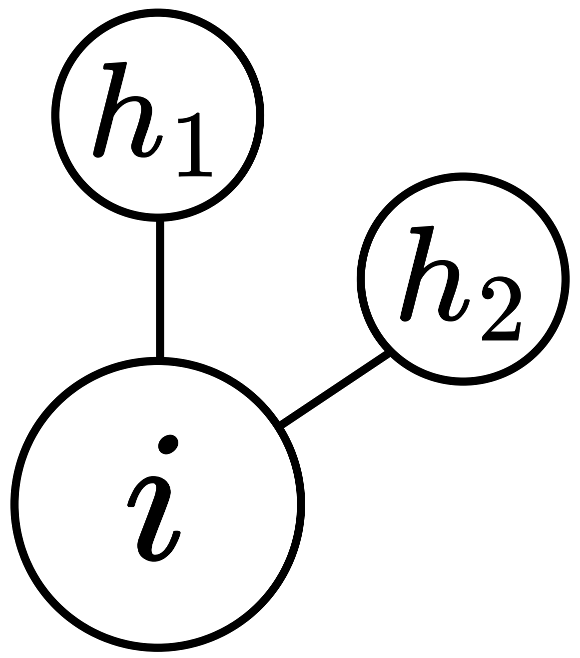


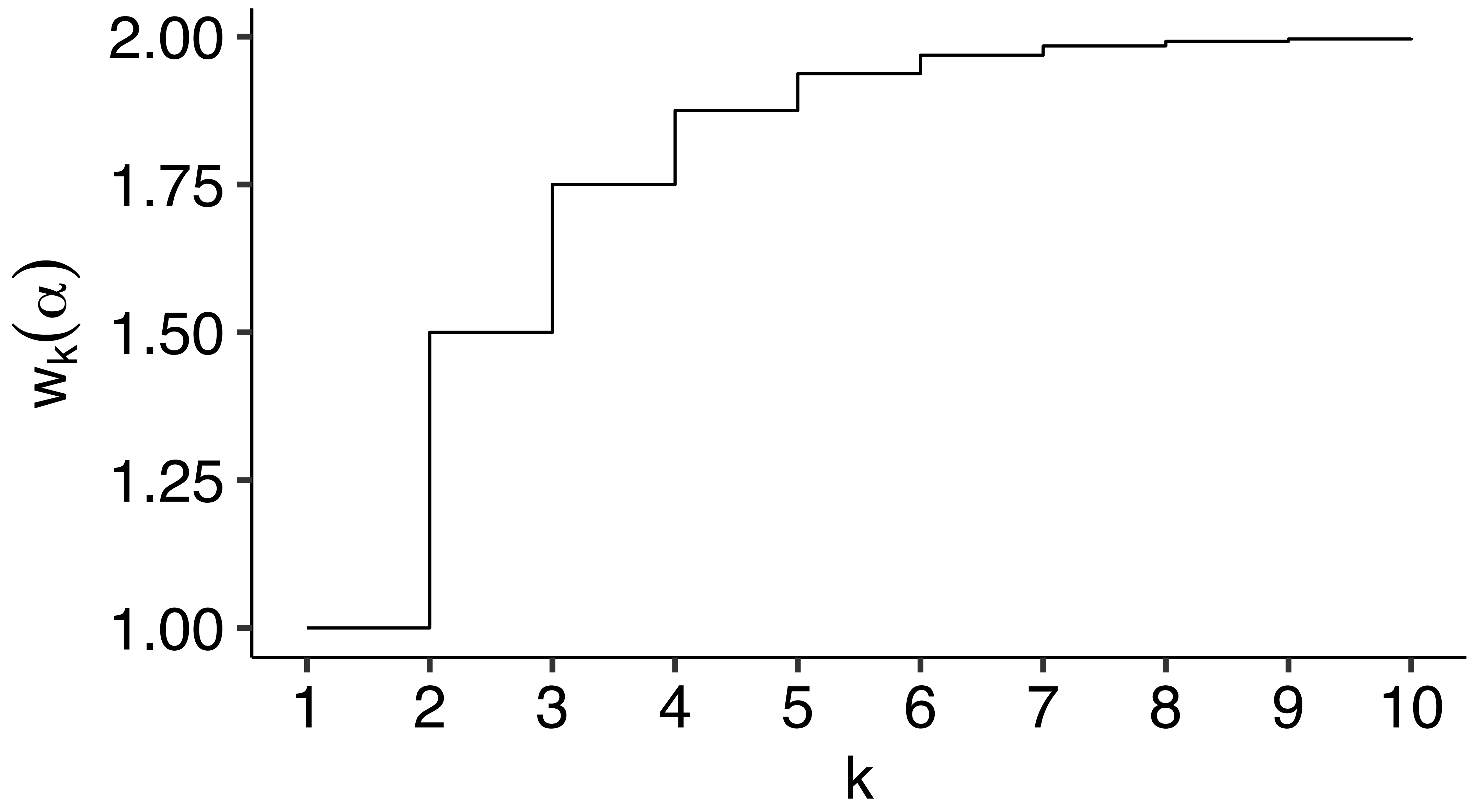
$$GWDEG(\mathbf{y}_t, \alpha) = \sum_{k=1}^{n-2} w_k(\alpha) DEG_k(\mathbf{y}_t)$$

$$w_k(\alpha) = \exp\{\alpha\}\left(1 - \left(1 - \exp\{-\alpha\}\right)^k\right)$$

Degeneracy Issues

- Problem 2: Which degree and triangular statistics to include?
  - There are n degree and shared partner statistics?
  - Incorporating all of them makes the model unstable
- Solution: Incorporate geometrically weighted statistics



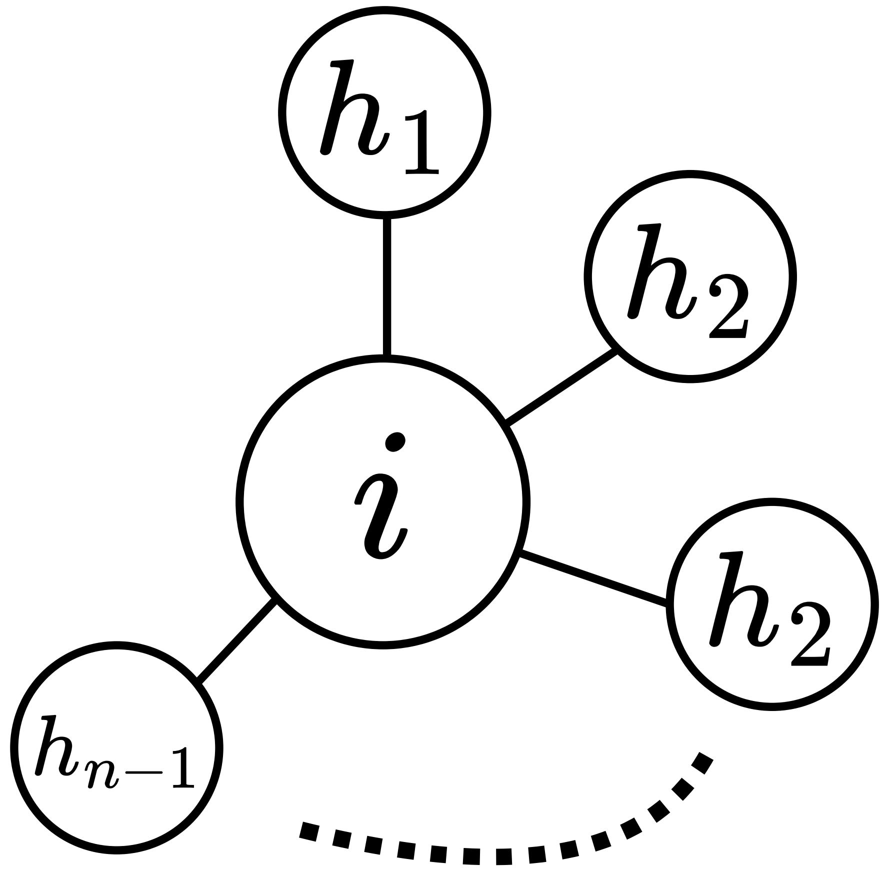


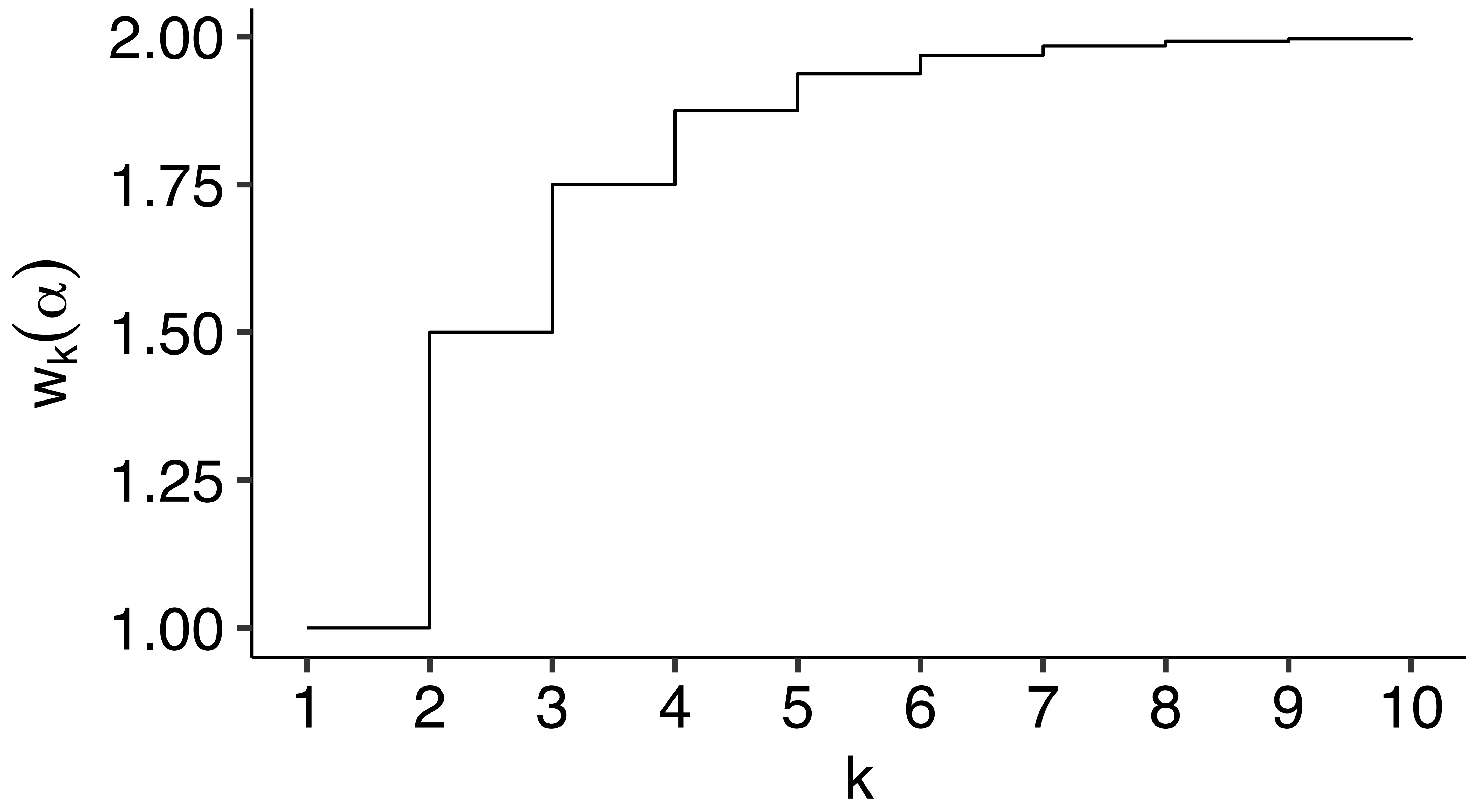
$$GWDEG(\mathbf{y}_t, \alpha) = \sum_{k=1}^{n-2} w_k(\alpha) DEG_k(\mathbf{y}_t)$$

$$w_k(\alpha) = \exp\{\alpha\}\left(1 - \left(1 - \exp\{-\alpha\}\right)^k\right)$$

Degeneracy Issues

- Problem 2: Which degree and triangular statistics to include?
  - There are n degree and shared partner statistics?
  - Incorporating all of them makes the model unstable
- Solution: Incorporate geometrically weighted statistics



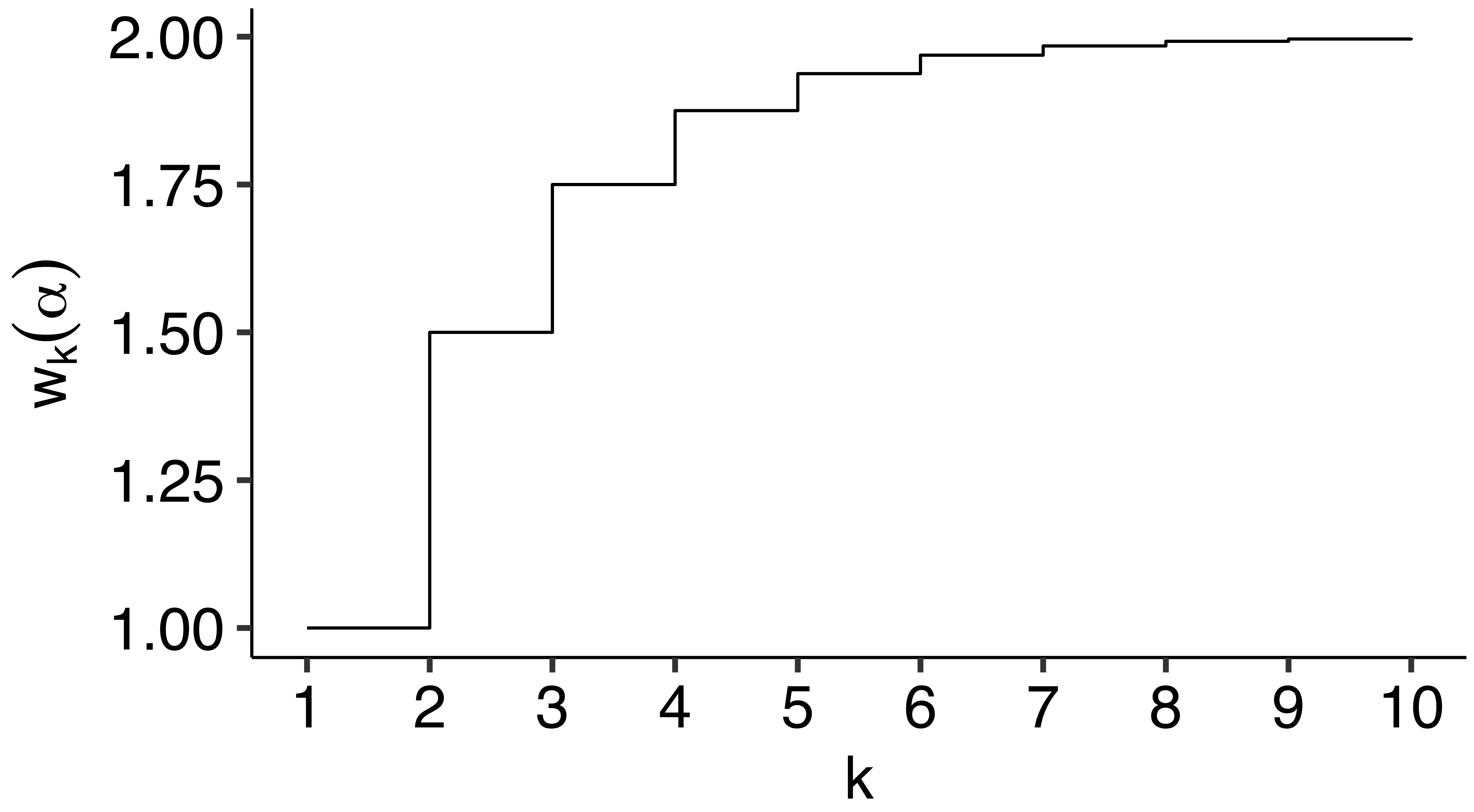


$$GWDEG(\mathbf{y}_t, \alpha) = \sum_{k=1}^{n-2} w_k(\alpha) DEG_k(\mathbf{y}_t)$$

$$w_k(\alpha) = \exp\{\alpha\}\left(1 - \left(1 - \exp\{-\alpha\}\right)^k\right)$$

Degeneracy Issues

- Interpretation for degree statistics
  - $\theta_{GWDEG} > 0$: An edge from low-degree actors is more likely than from high-degree actors ⇒ Decentralized Network
  - $\theta_{GWDEG} < 0$: An edge from high-degree actors is more likely than from low-degree actors ⇒ Centralized Network



$$GWDEG(\mathbf{y}_t, \alpha) = \sum_{k=1}^{n-2} w_k(\alpha) DEG_k(\mathbf{y}_t)$$

$$w_k(\alpha) = \exp\{\alpha\}\left(1 - \left(1 - \exp\{-\alpha\}\right)^k\right)$$

Application Example

- High School Friendships
- Included covariates:
    1. Edges: How many edges are in the network?
    2. Gw. Edegewise-shared Partner/ Degree: Is there clustering or some type of centralization in the network?
    3. Grade/Race/Sex: Do we observe homophily effects?

Code Snippet

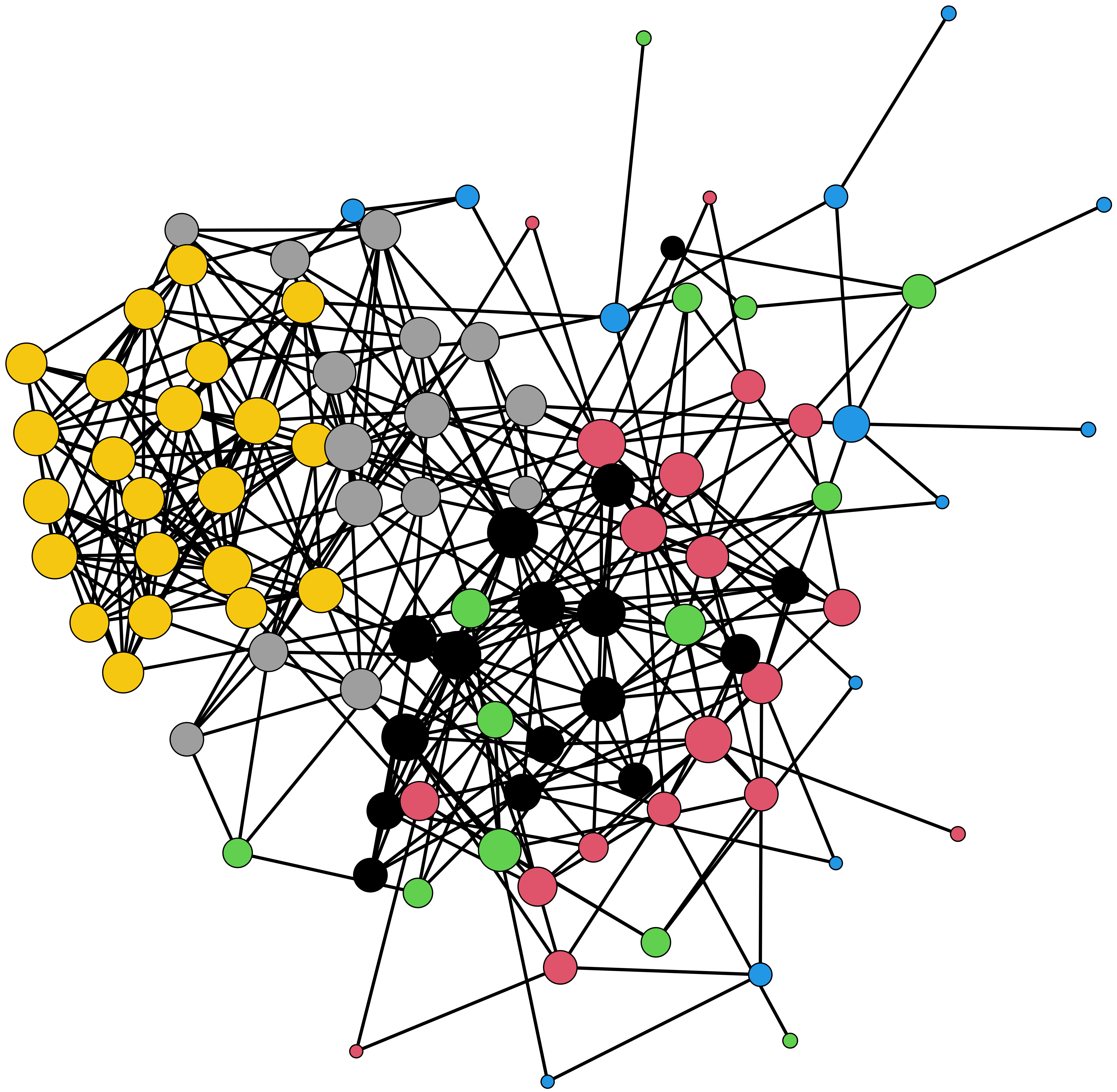
```
plot(friendship_network,
     mode = "kamadakawai",
     vertex.cex =friendship_network %v% "deg_log_log",
     vertex.col=friendship_network %v% "grade")
```

Code Snippet

```
> friendship_network
 Network attributes:
  vertices = 100
  directed = FALSE
  hyper = FALSE
  loops = FALSE
  multiple = FALSE
  bipartite = FALSE
  total edges= 348
    missing edges= 0
    non-missing edges= 348

 Vertex attribute names:
    grade race sex vertex.names

No edge attributes
```

Application Example

- High School Friendships
- Included covariates:
    1. Edges: How many edges are in the network?
    2. Gw. Edegewise-shared Partner/ Degree: Is there clustering or some type of centralization in the network?
    3. Grade/Race/Sex: Do we observe homophily effects?



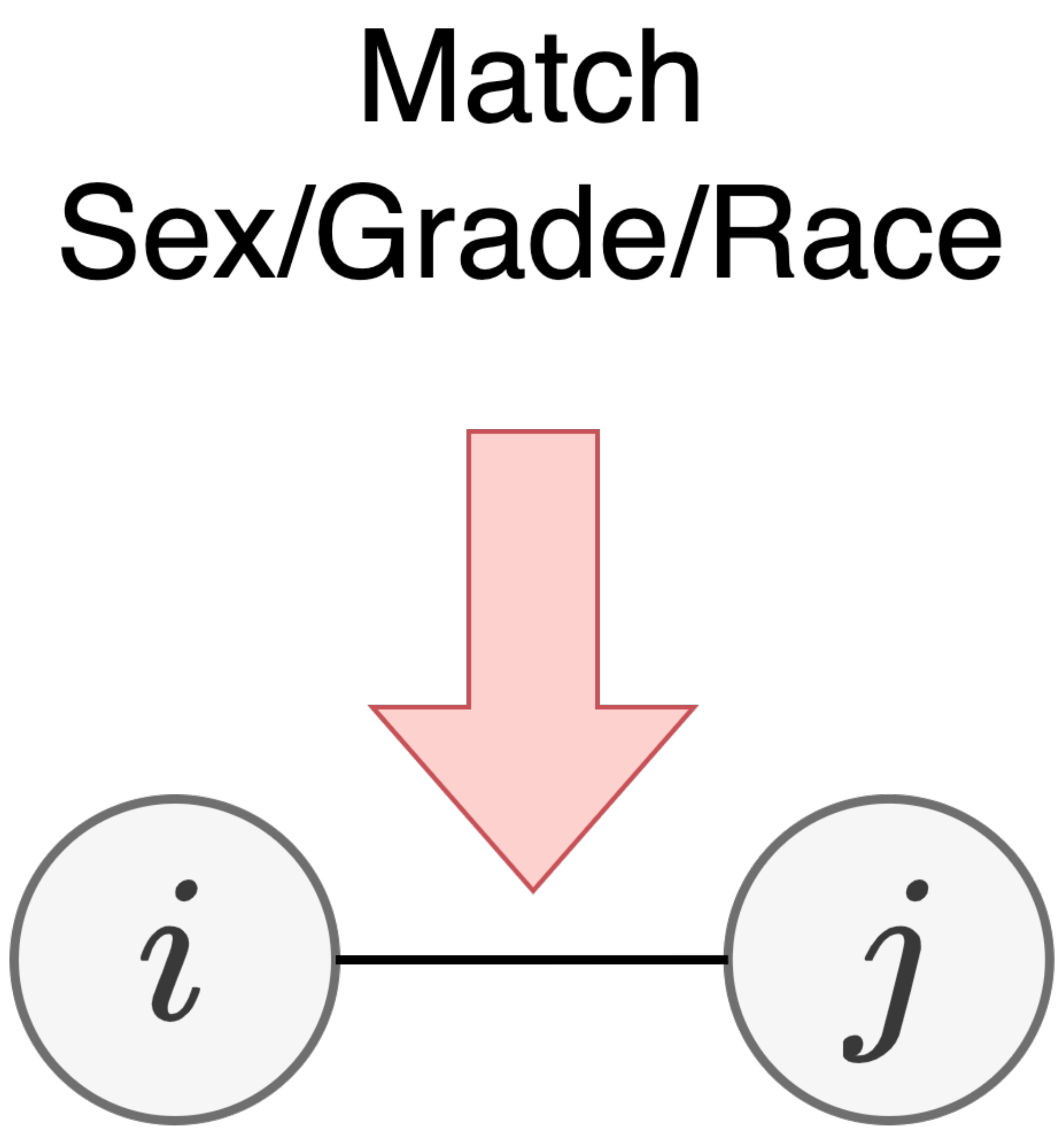
Code Snippet

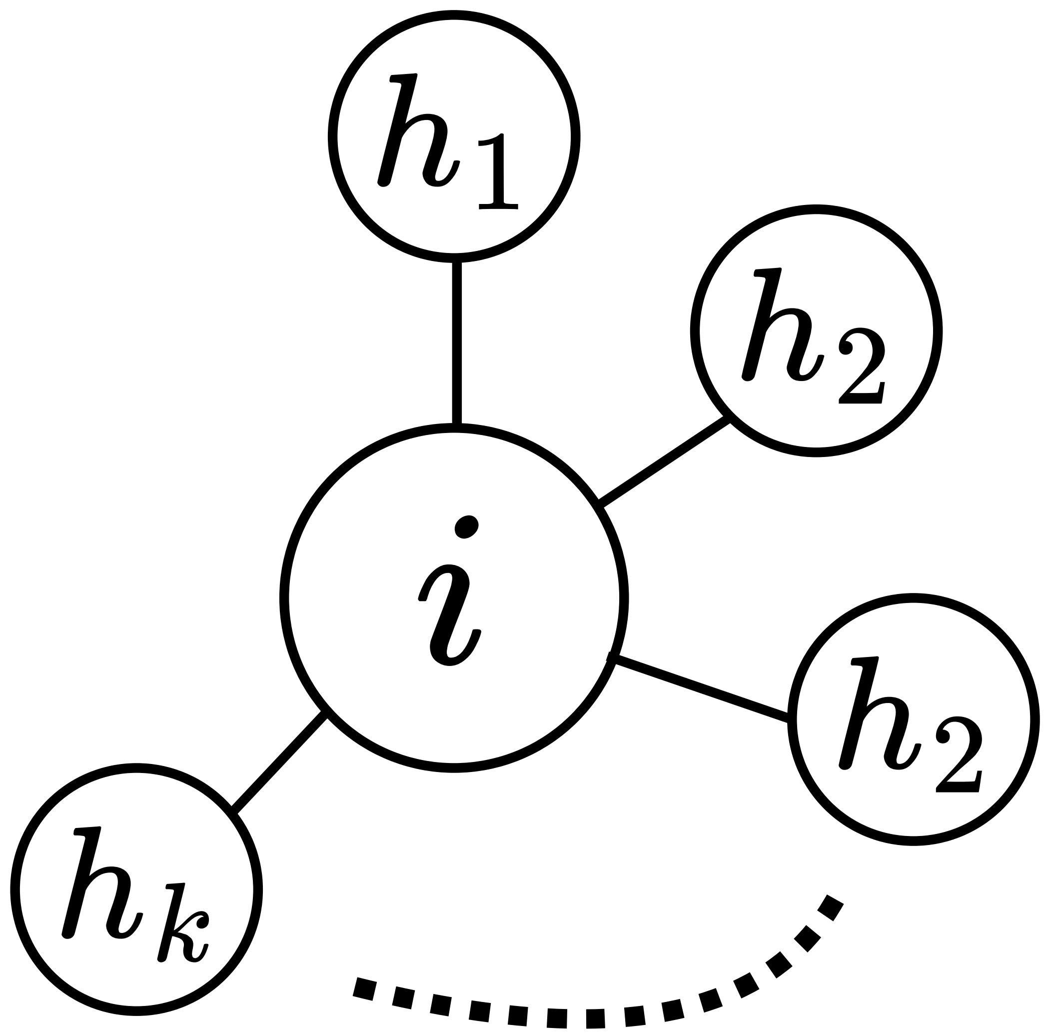
```
plot(friendship_network,
     mode = "kamadakawai",
     vertex.cex =friendship_network %v% "deg_log_log",
     vertex.col=friendship_network %v% "grade")
```

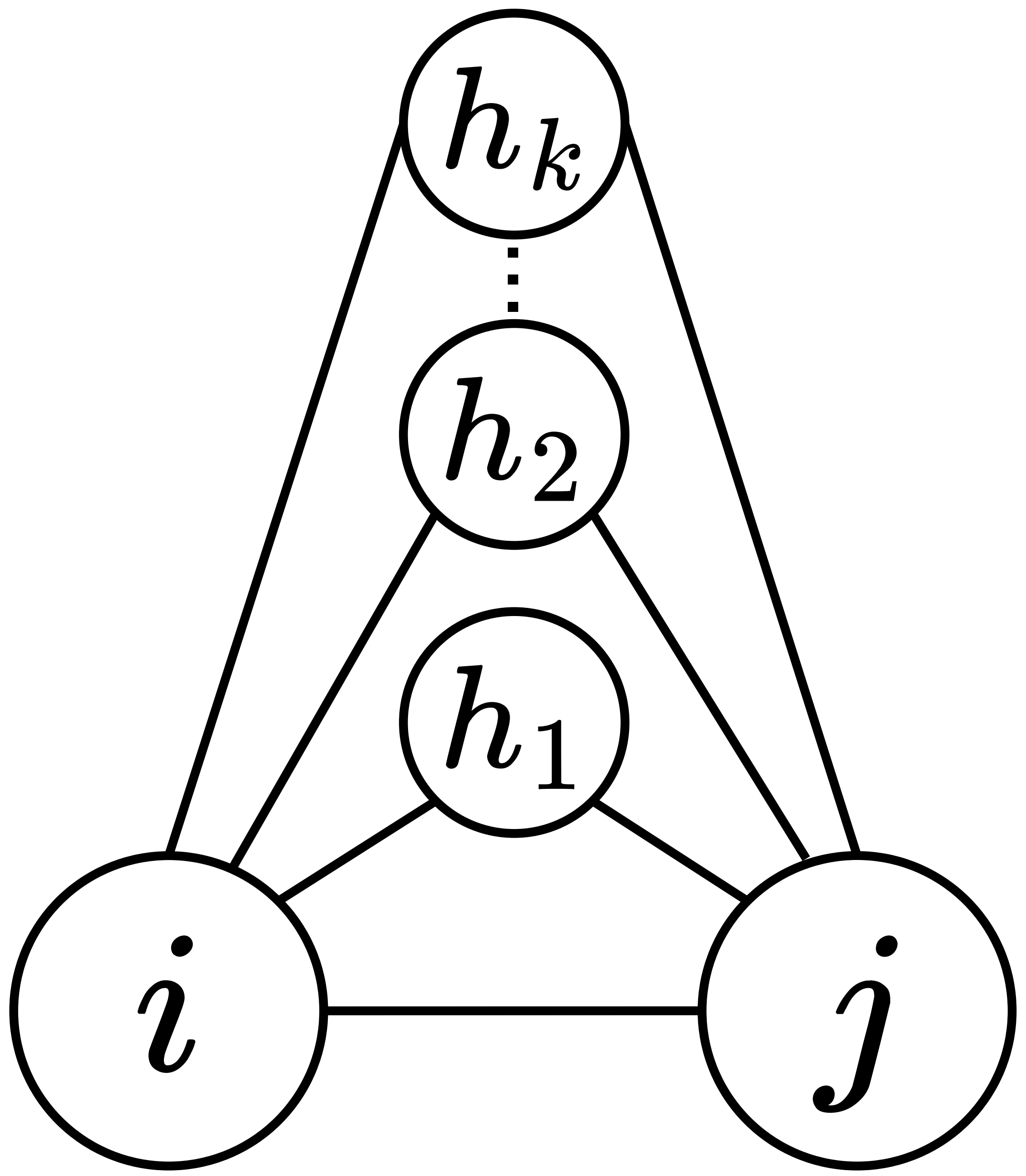
Estimate an ERGM

Code Snippet (see also help(“ergm-terms”))

```
friendship_model <- ergm(friendship_network~ edges+
                           gwesp(log(2), fixed = T) +
                           gwdegree(log(2), fixed = T) +
                           nodematch("grade")+
                           nodematch("race")+
                           nodematch("sex"))
```







Estimate an ERGM

Code Snippet

```
> summary(friendship_model)
Call:
ergm(formula = friendship_network ~ edges + gwesp(log(2), fixed = T) +
    gwdegree(log(2), fixed = T) + nodematch("grade") + nodematch("race") +
    nodematch("sex"))

Monte Carlo Maximum Likelihood Results:

                              Estimate Std. Error MCMC % z value Pr(>|z|)
edges                          -4.3530     0.1985      0 -21.926   <1e-04 ***
gwesp.fixed.0.693147180559945   0.6011     0.0810      0   7.421   <1e-04 ***
gwdeg.fixed.0.693147180559945  -0.3884     0.4253      0  -0.913   0.3610
nodematch.grade                 1.8985     0.1234      0  15.379   <1e-04 ***
nodematch.race                  0.0537     0.1188      0   0.452   0.6512
nodematch.sex                   0.2796     0.1105      0   2.529   0.0114 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 6862  on 4950  degrees of freedom
 Residual Deviance: 1912  on 4944  degrees of freedom

AIC: 1924  BIC: 1963  (Smaller is better. MC Std. Err. = 0.4417)
```

Estimate an ERGM

Code Snippet

```
> summary(friendship_model)
Call:
ergm(formula = friendship_network ~ edges + gwesp(log(2), fixed = T) +
    gwdegree(log(2), fixed = T) + nodematch("grade") + nodematch("race") +
    nodematch("sex"))

Monte Carlo Maximum Likelihood Results:

                              Estimate Std. Error MCMC % z value Pr(>|z|)
edges                          -4.3530     0.1985      0 -21.926   <1e-04 ***
gwesp.fixed.0.693147180559945   0.6011     0.0810      0   7.421   <1e-04 ***
gwdeg.fixed.0.693147180559945  -0.3884     0.4253      0  -0.913   0.3610
nodematch.grade                 1.8985     0.1234      0  15.379   <1e-04 ***
nodematch.race                  0.0537     0.1188      0   0.452   0.6512
nodematch.sex                   0.2796     0.1105      0   2.529   0.0114 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

     Null Deviance: 6862  on 4950  degrees of freedom
 Residual Deviance: 1912  on 4944  degrees of freedom

AIC: 1924  BIC: 1963  (Smaller is better. MC Std. Err. = 0.4417)
```

- Negative edge parameter ⇒ Sparse graph

Estimate an ERGM

Code Snippet

```
> summary(friendship_model)
Call:
ergm(formula = friendship_network ~ edges + gwesp(log(2), fixed = T) +
    gwdegree(log(2), fixed = T) + nodematch("grade") + nodematch("race") +
    nodematch("sex"))

Monte Carlo Maximum Likelihood Results:

                              Estimate Std. Error MCMC % z value Pr(>|z|)
edges                          -4.3530     0.1985      0 -21.926   <1e-04 ***
gwesp.fixed.0.693147180559945   0.6011     0.0810      0   7.421   <1e-04 ***
gwdeg.fixed.0.693147180559945  -0.3884     0.4253      0  -0.913   0.3610
nodematch.grade                 1.8985     0.1234      0  15.379   <1e-04 ***
nodematch.race                  0.0537     0.1188      0   0.452   0.6512
nodematch.sex                   0.2796     0.1105      0   2.529   0.0114 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 6862  on 4950  degrees of freedom
 Residual Deviance: 1912  on 4944  degrees of freedom

AIC: 1924  BIC: 1963  (Smaller is better. MC Std. Err. = 0.4417)
```

- Non-significant degree term ⇒ No clear centralization pattern

Estimate an ERGM

Code Snippet

```
> summary(friendship_model)
Call:
ergm(formula = friendship_network ~ edges + gwesp(log(2), fixed = T) +
    gwdegree(log(2), fixed = T) + nodematch("grade") + nodematch("race") +
    nodematch("sex"))

Monte Carlo Maximum Likelihood Results:

                                  Estimate Std. Error MCMC % z value Pr(>|z|)
edges                              -4.3530     0.1985      0 -21.926   <1e-04 ***
gwesp.fixed.0.693147180559945      0.6011     0.0810      0   7.421   <1e-04 ***
gwdeg.fixed.0.693147180559945     -0.3884     0.4253      0  -0.913   0.3610
nodematch.grade                    1.8985     0.1234      0  15.379   <1e-04 ***
nodematch.race                     0.0537     0.1188      0   0.452   0.6512
nodematch.sex                      0.2796     0.1105      0   2.529   0.0114 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 6862  on 4950  degrees of freedom
 Residual Deviance: 1912  on 4944  degrees of freedom

AIC: 1924  BIC: 1963  (Smaller is better. MC Std. Err. = 0.4417)
```

- More clustering than expected by randomness

Estimate an ERGM

Code Snippet

```
> summary(friendship_model)
Call:
ergm(formula = friendship_network ~ edges + gwesp(log(2), fixed = T) +
    gwdegree(log(2), fixed = T) + nodematch("grade") + nodematch("race") +
    nodematch("sex"))

Monte Carlo Maximum Likelihood Results:

                                Estimate Std. Error MCMC % z value Pr(>|z|)
edges                            -4.3530     0.1985      0 -21.926   <1e-04 ***
gwesp.fixed.0.693147180559945     0.6011     0.0810      0   7.421   <1e-04 ***
gwdeg.fixed.0.693147180559945    -0.3884     0.4253      0  -0.913   0.3610
nodematch.grade                   1.8985     0.1234      0  15.379   <1e-04 ***
nodematch.race                    0.0537     0.1188      0   0.452   0.6512
nodematch.sex                     0.2796     0.1105      0   2.529   0.0114 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 6862  on 4950  degrees of freedom
 Residual Deviance: 1912  on 4944  degrees of freedom

AIC: 1924  BIC: 1963  (Smaller is better. MC Std. Err. = 0.4417)
```

- Common grade and sex has a positive impact on edge formation

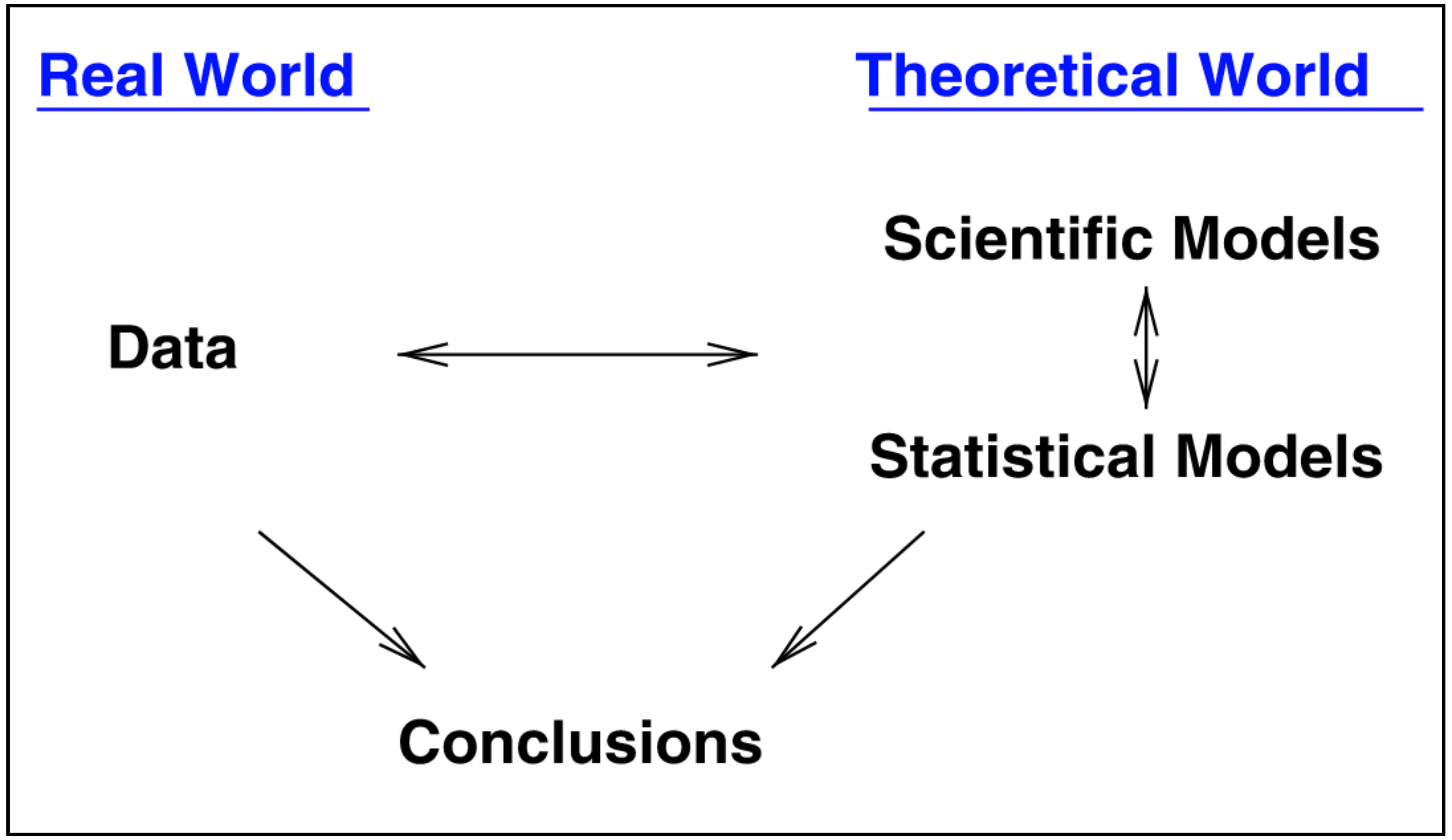
Estimate an ERGM

Code Snippet

```
> summary(friendship_model)
Call:
ergm(formula = friendship_network ~ edges + gwesp(log(2), fixed = T) +
    gwdegree(log(2), fixed = T) + nodematch("grade") + nodematch("race") +
    nodematch("sex"))

Monte Carlo Maximum Likelihood Results:

                              Estimate Std. Error MCMC % z value Pr(>|z|)
edges                          -4.3530     0.1985      0 -21.926   <1e-04 ***
gwesp.fixed.0.693147180559945   0.6011     0.0810      0   7.421   <1e-04 ***
gwdeg.fixed.0.693147180559945  -0.3884     0.4253      0  -0.913   0.3610
nodematch.grade                 1.8985     0.1234      0  15.379   <1e-04 ***
nodematch.race                  0.0537     0.1188      0   0.452   0.6512
nodematch.sex                   0.2796     0.1105      0   2.529   0.0114 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

     Null Deviance: 6862  on 4950  degrees of freedom
 Residual Deviance: 1912  on 4944  degrees of freedom

AIC: 1924  BIC: 1963  (Smaller is better. MC Std. Err. = 0.4417)
```

- Matching race no significantly positive impact

Model Assessment



Code Snippet

```
friendship_gof <- gof(friendship_model)
plot(friendship_gof)
```

Model Assessment

Code Snippet

```
friendship_gof <- gof(friendship_model)
plot(friendship_gof)
```

Current Research

- ERGMs cover the general class of discrete exponential families
  - Extending the probability distribution other samples spaces than binary relations (signed, count, or rank networks)
  - Discover temporal patterns in networks
- Analyzing Large Networks with scalable methods and models
  - Scalable models via local dependence (dependence constrained to neighborhoods)
  - Scalable methods via approximative solutions (composite likelihood or variational approximations)
  - Joint models for networks and attributes
- ERGMs for sampled networks

Recap

1. Why is modeling networks important?
2. Why are common regression models not sufficient?
3. What are random graph models?
4. How can we capture particular aspects of network data?
5. How can we interpret ERGMs?
6. How are ERGMs estimated?
7. What are degeneracy issues?
8. How can we address degeneracy issues?
9. How can we estimate the ERGM in R?
10. How can we assess the fit of an ERGM?

Questions?