Statistical Analysis of Networks

Introduction to Exponential Random Graph Models

Learning Goals

- Understand the logic of the exponential random graph model.
- Understand the historical development of exponential random graph models.
- Describe the properties of various types of exponential random graph models.
- Understand the notion of "network configurations" as operationalizations of theoretical concepts.

Introduction

- **So far**:
  - We have been doing "descriptive statistics" with network data.
    - Example: how are the degrees distributed in this network?
- **Now**:
  - We want to shift toward "inferential statistics" with network data.
    - Example: is the distribution of degrees different from a network where ties form at random?

Introduction

- **New questions**:
  - *How do networks form*?
  - *What are the micro patterns that generate global structure?*
  - *How likely is it that we would observe these configurations if ties formed at random?*

- Exponential Random Graph Models or ERGMs, provide a **model** (an account of what **governs the formation of a network**) for examining such questions.

Modeling Networks

- Why should we care? Why not be satisfied with descriptive statistics?
  - Complexity and randomness
  - Statistical inference and hypothesis testing
  - Global structure from local structure (micro-macro problem)

Logic of Random Graph Modeling

- We have an **observed network**, and want to know about the **stochastic** process by which it came about.

- Conceptual Analog:
  - Sampling from a normal distribution.
  - We don't get the **exact** same data (i.e. sample *statistics* differ from population *parameters*).
    - But, there is some **process** generating our sample statistics (i.e. central tendency and dispersion).

Logic of Random Graph Modeling

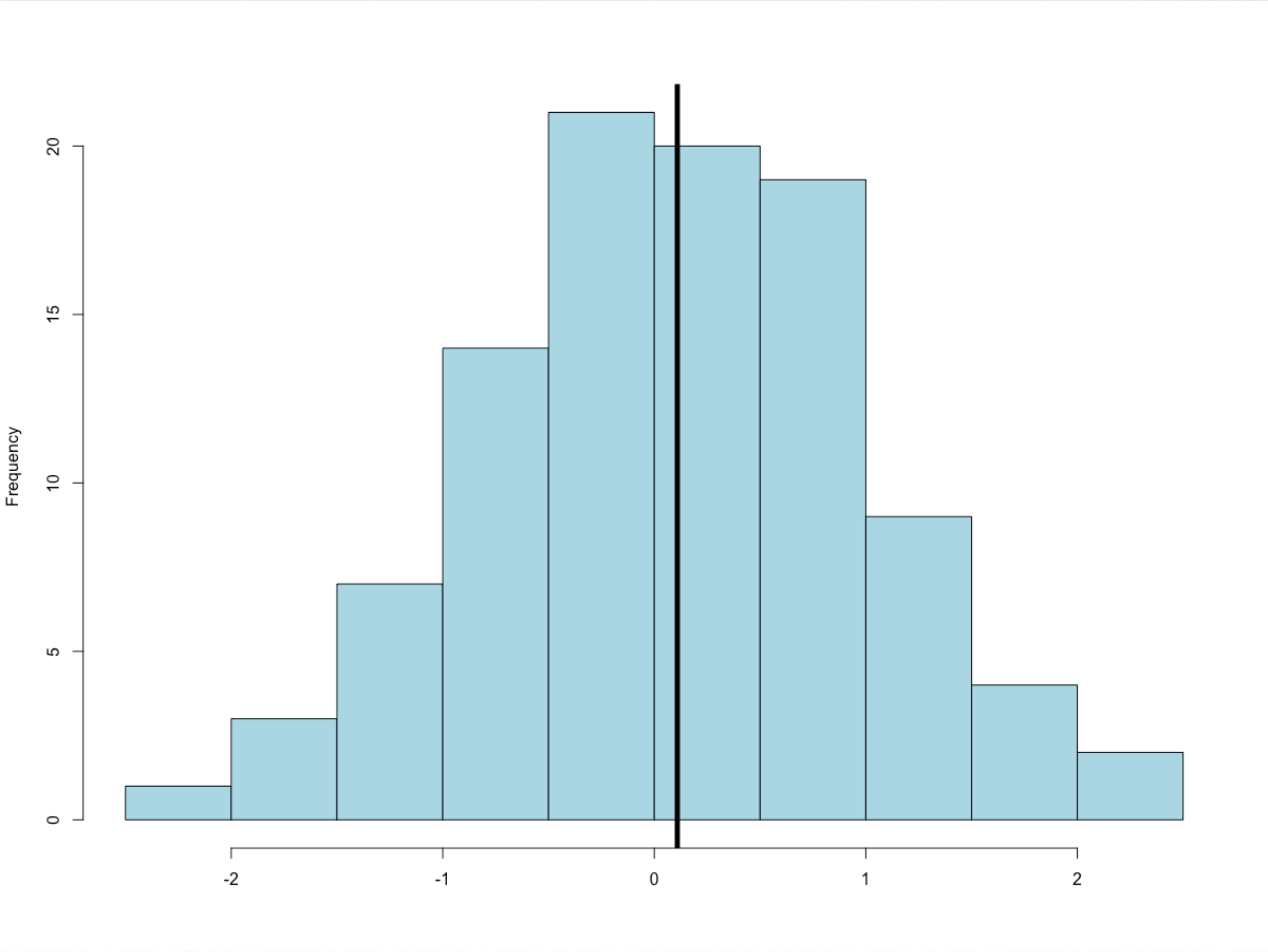
- Simulate some data from a standard normal distribution and calculate the mean.

Logic of Random Graph Modeling

- Simulate some data from a standard normal distribution and calculate the mean.

100 random draws from a distribution with a mean = 0

Sample mean = 0.108

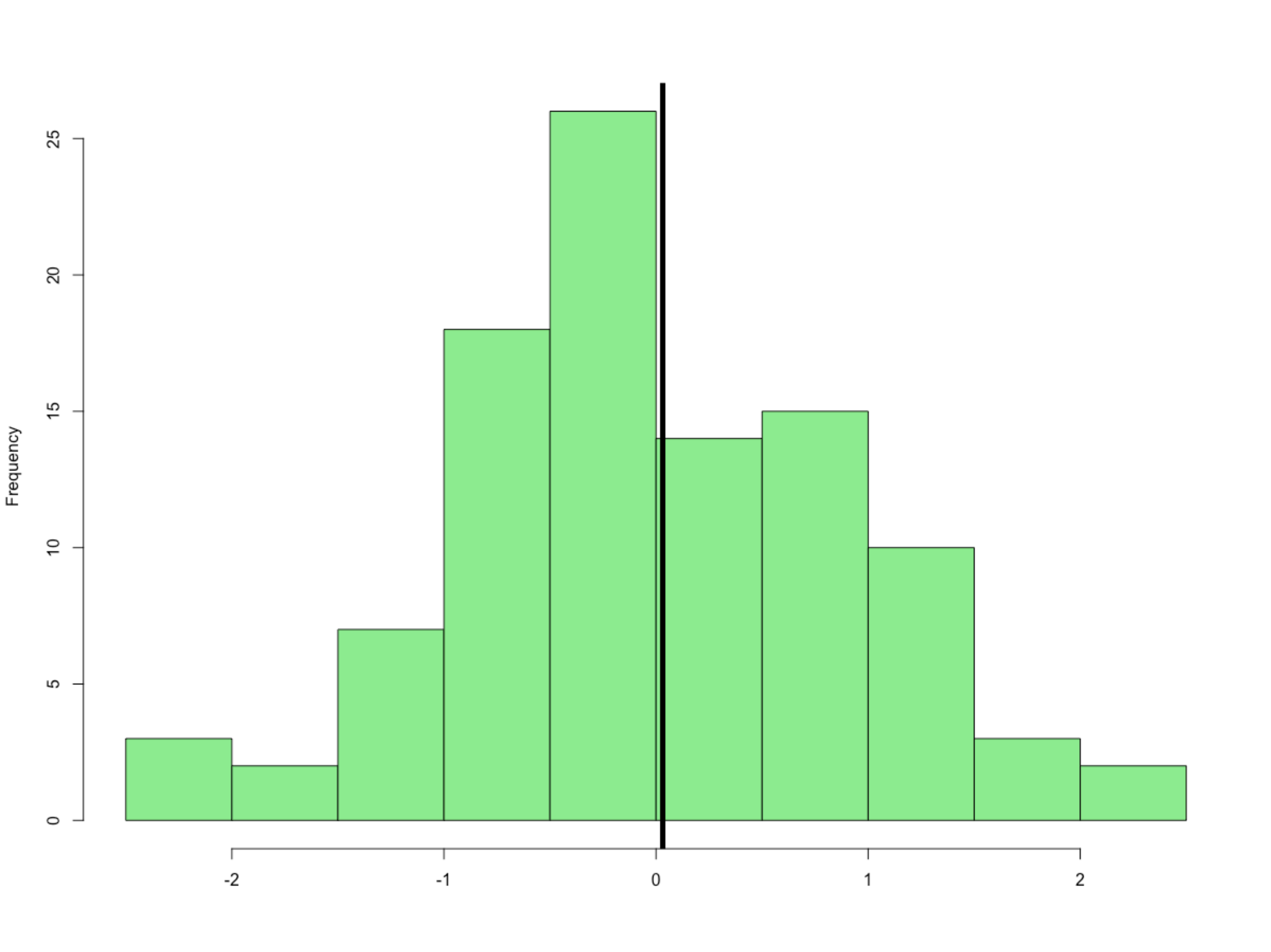


Logic of Random Graph Modeling

- Simulate some data from a standard normal distribution and calculate the mean.

100 random draws from a distribution with a mean = 0

Sample mean = 0.031

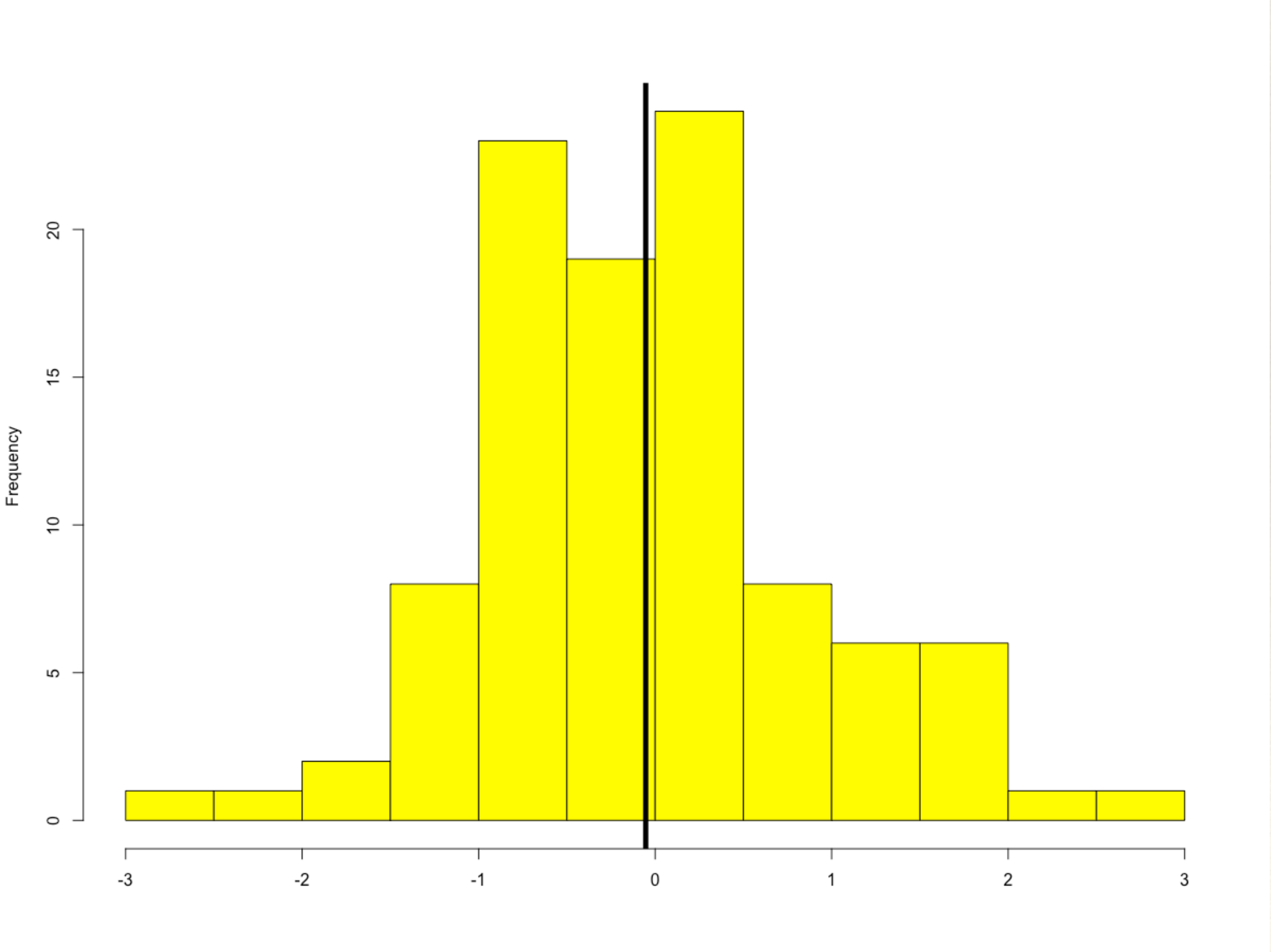


Logic of Random Graph Modeling

- Simulate some data from a standard normal distribution and calculate the mean.

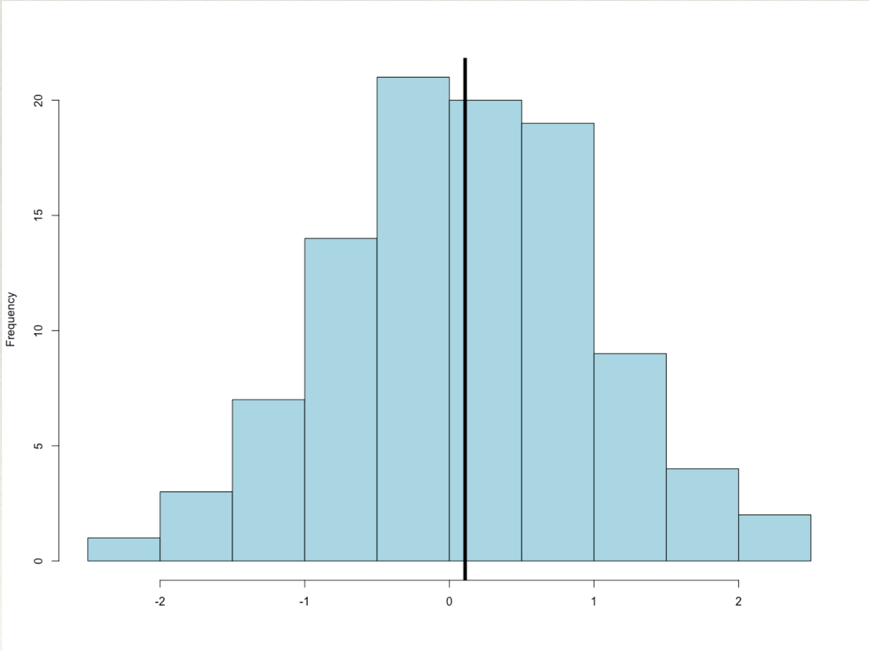
100 random draws from a distribution with a mean = 0

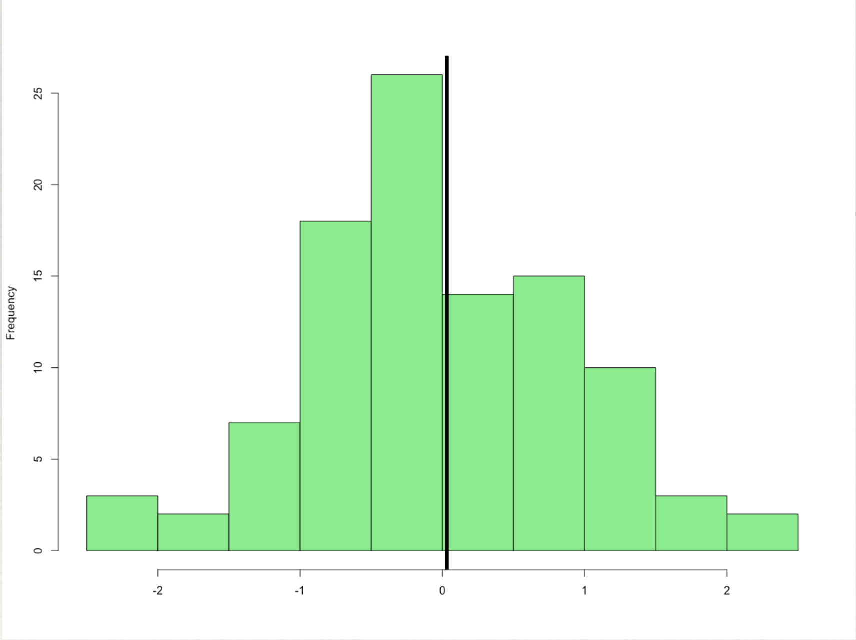
Sample mean = -0.053

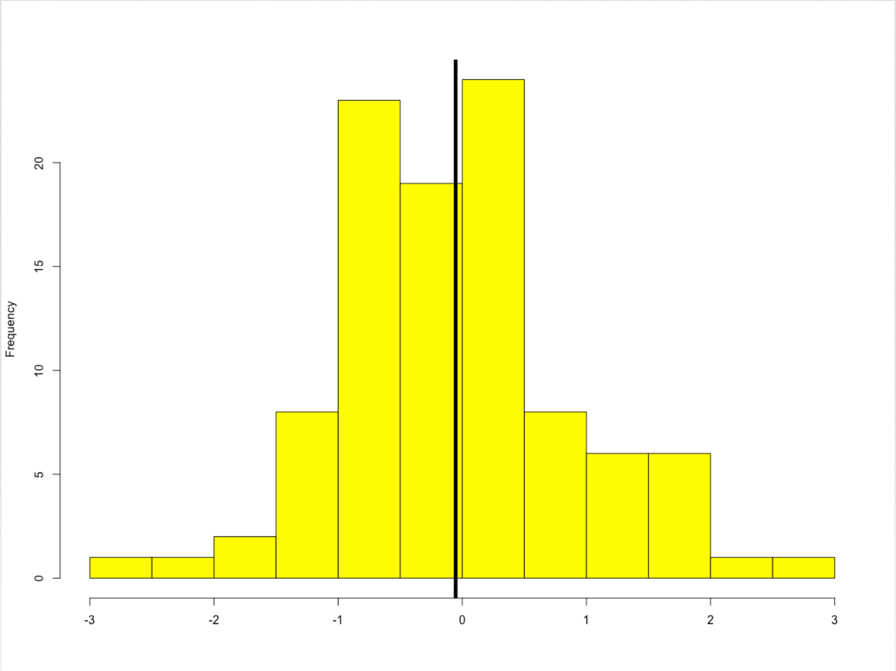


Logic of Random Graph Modeling

- There are data generated from the same distribution, but have different means.





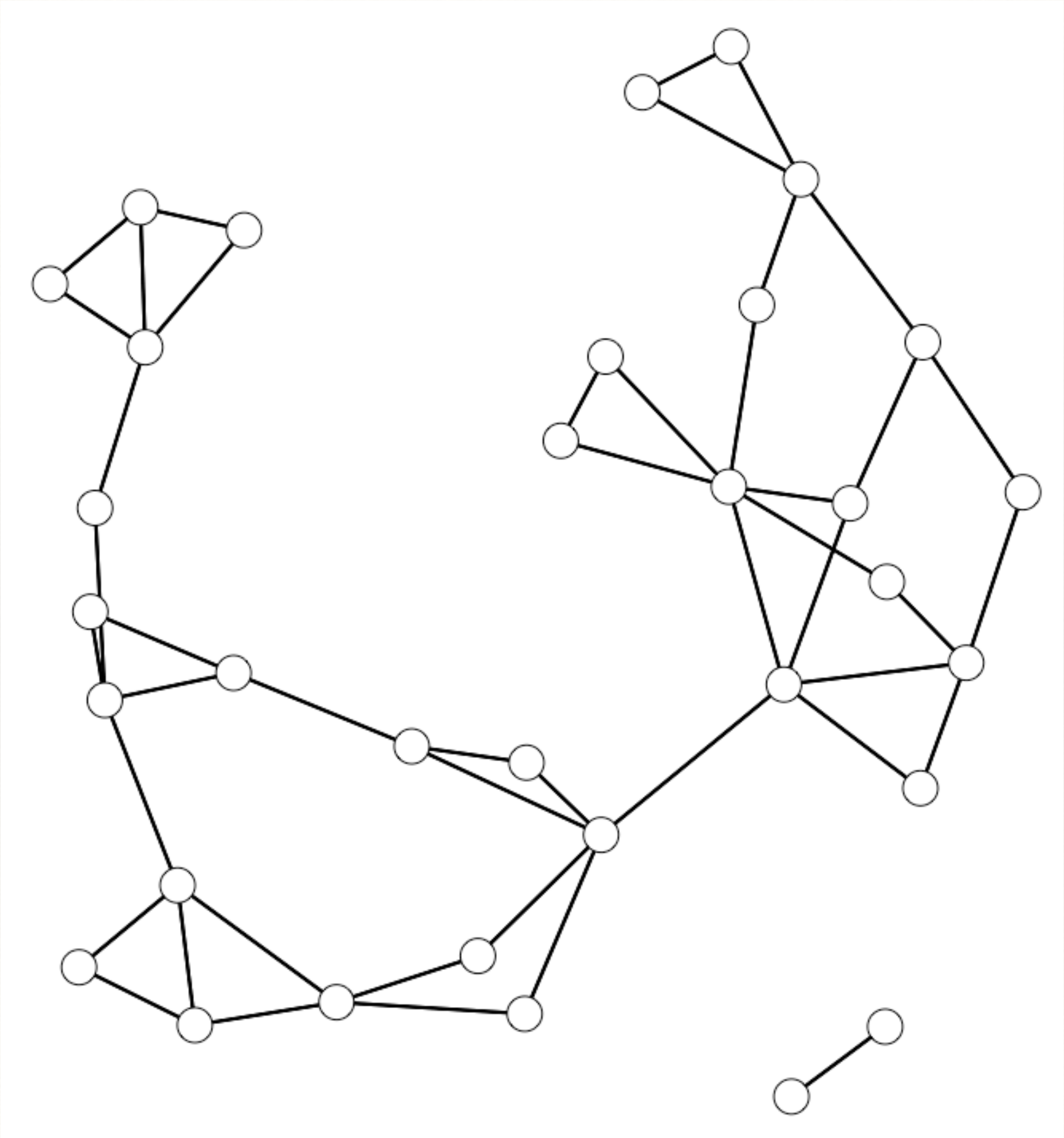


Logic of Random Graph Modeling

- In a similar way, we want to examine the *parameters* which generate the network we have observed.
  - Did it come about because people:
    - Reciprocate relationships?
    - Nominate popular others?
    - Close triads?

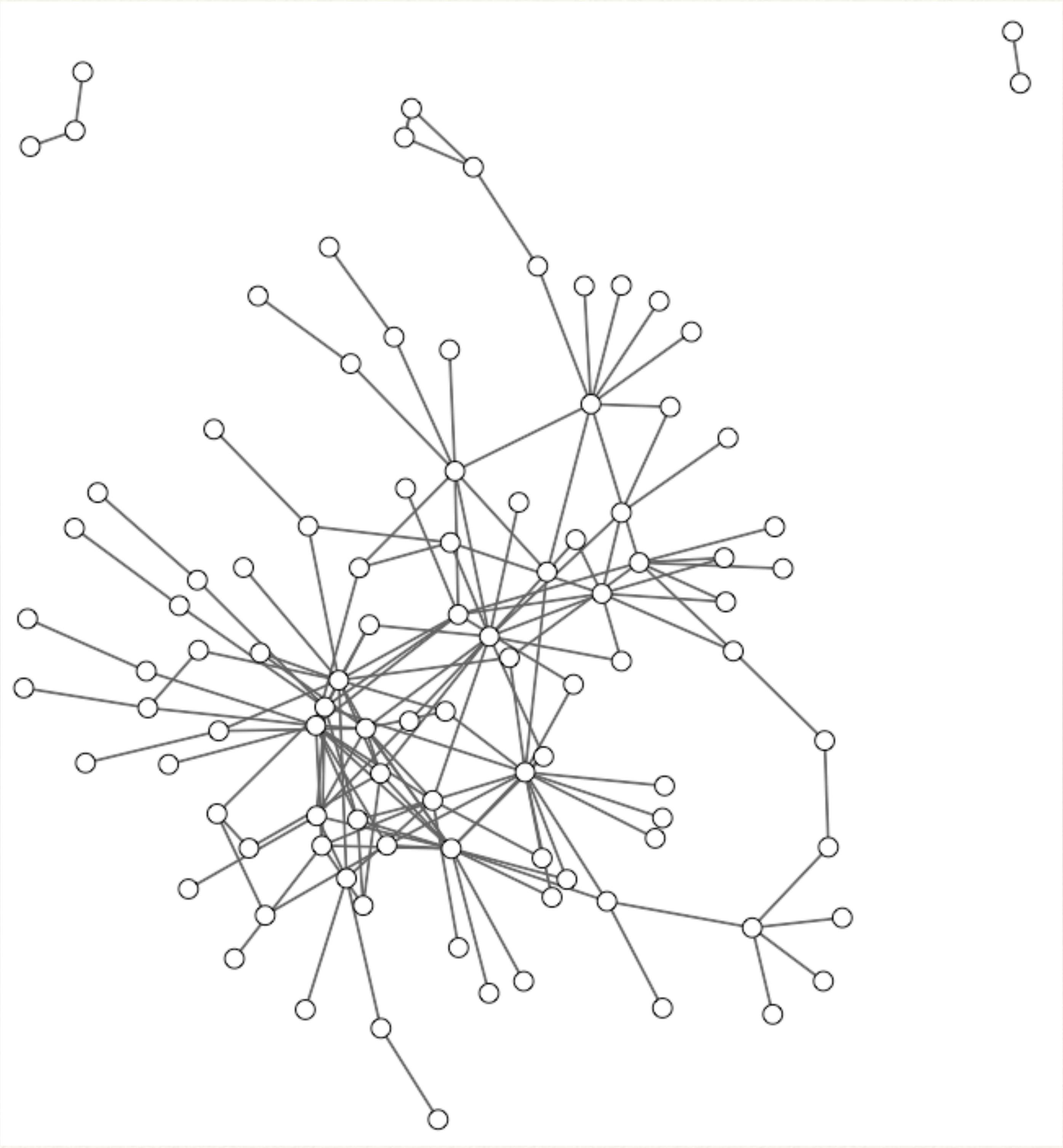
Example

What <u>pattern</u> do you see in these data?



Example

What pattern do you see in these data?



Emergent Structure

- Different processes can lead to similar outcomes.

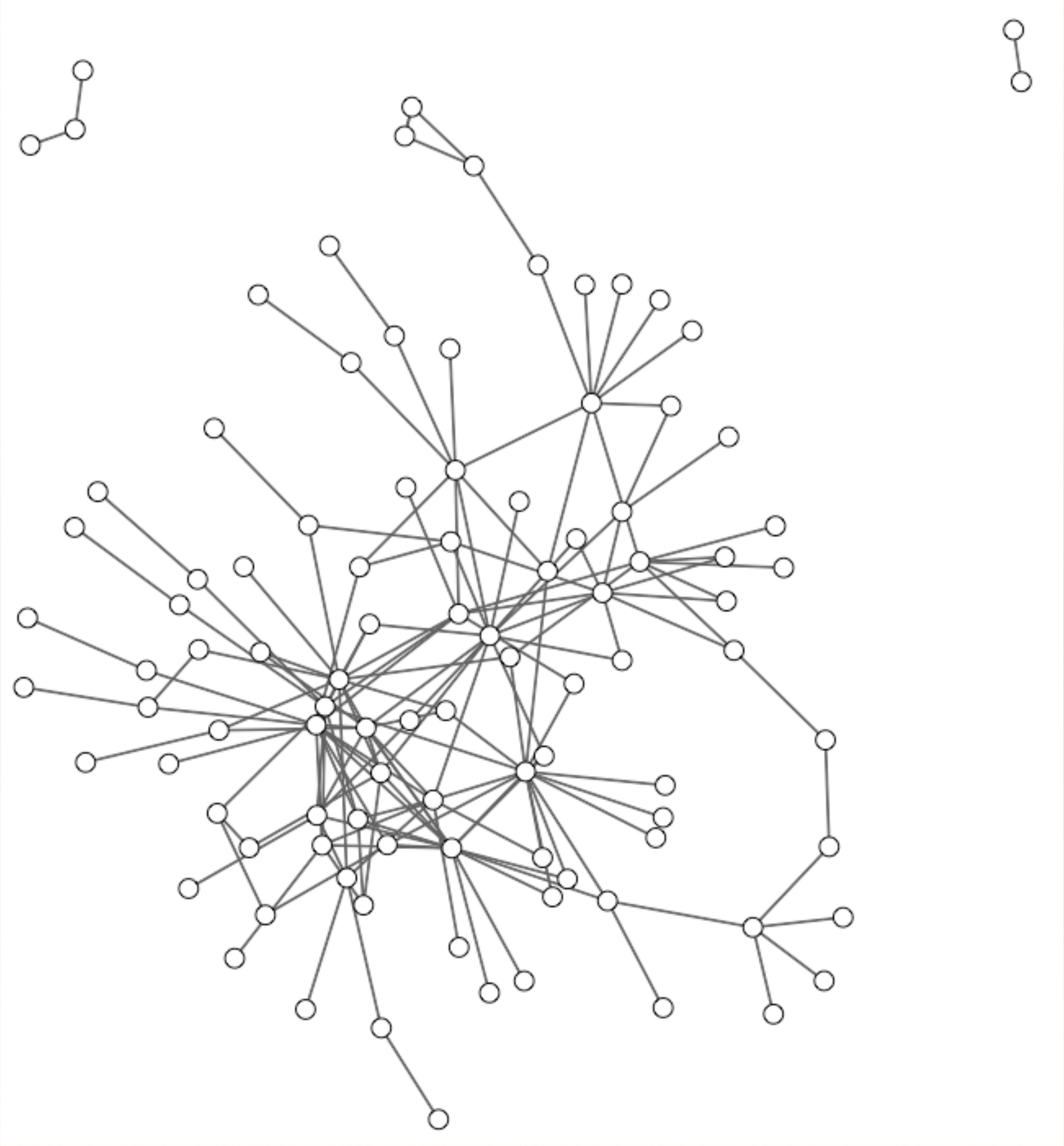
Example

Different processes can lead to similar outcomes:

Sociality-highly active persons create clusters.

Homophily-assortative mixing by attribute creates clusters.

Transitivity-triangles create clusters.



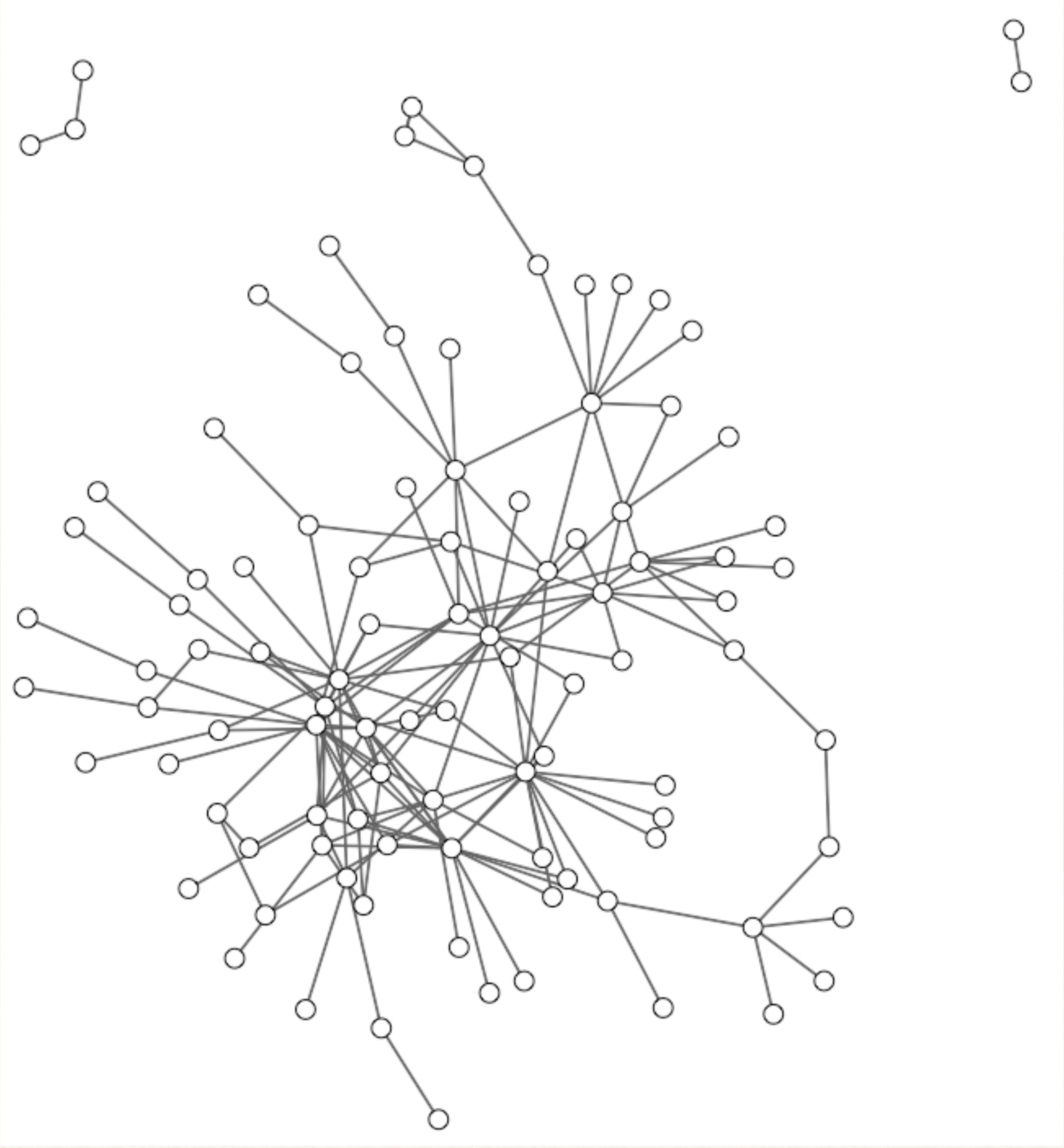
Example

Different processes can lead to similar outcomes:

Sociality-highly active persons create clusters.

Homophily-assortative mixing by attribute creates clusters.

Transitivity-triangles create clusters.



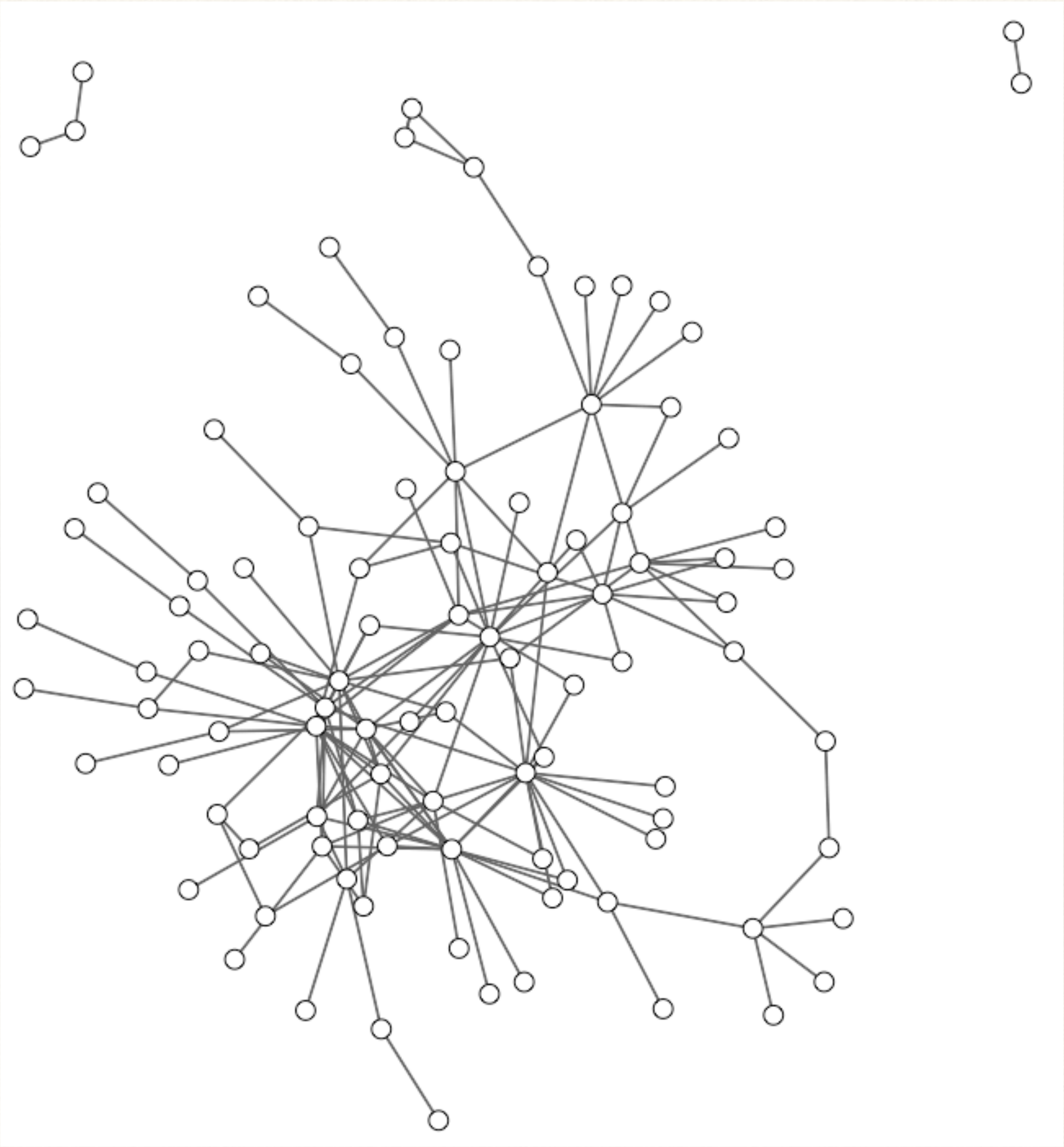
Example

Different processes can lead to similar outcomes:

Sociality-highly active persons create clusters.

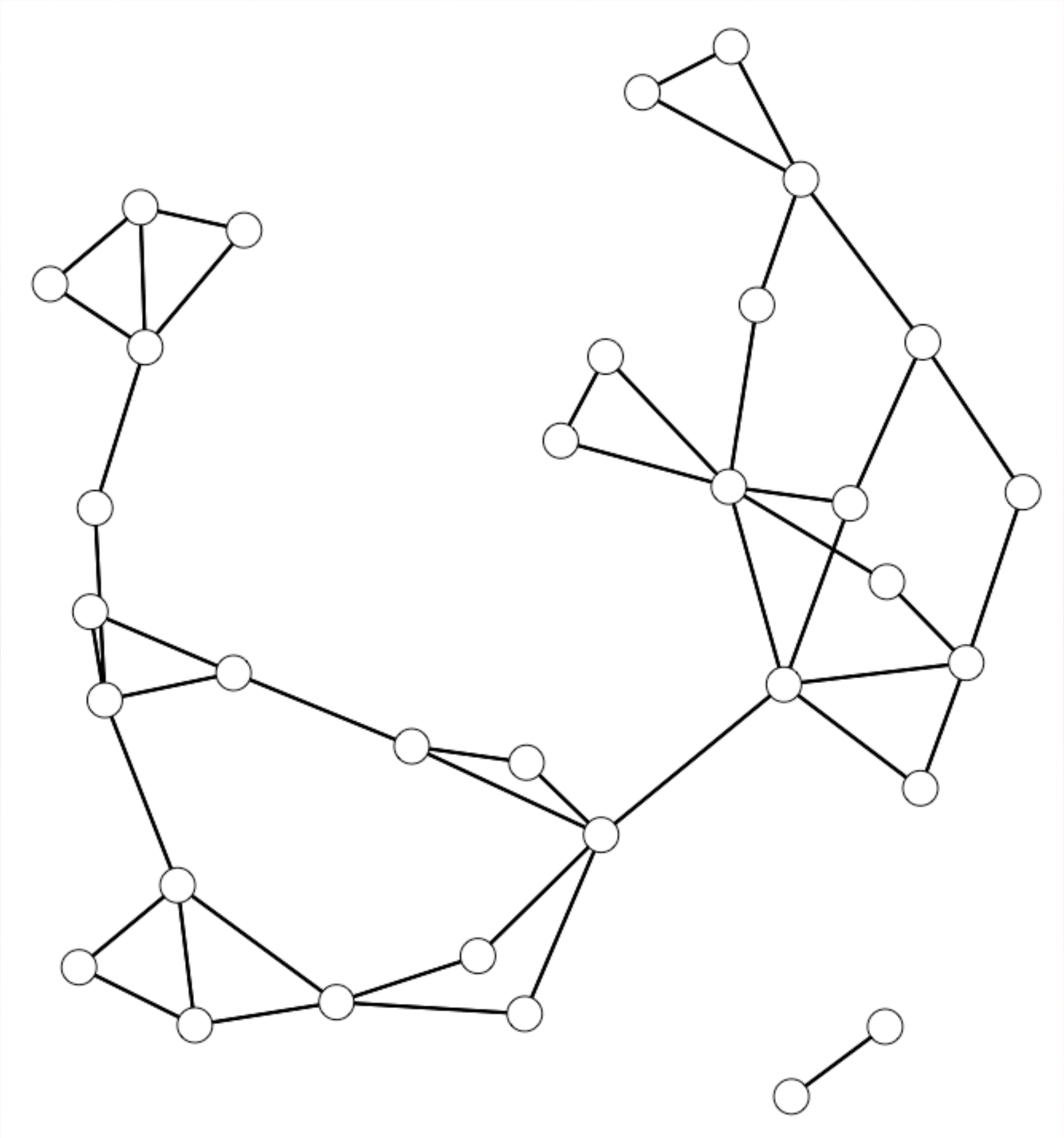
Homophily-assortative mixing by attribute creates clusters.

Transitivity-triangles create clusters.



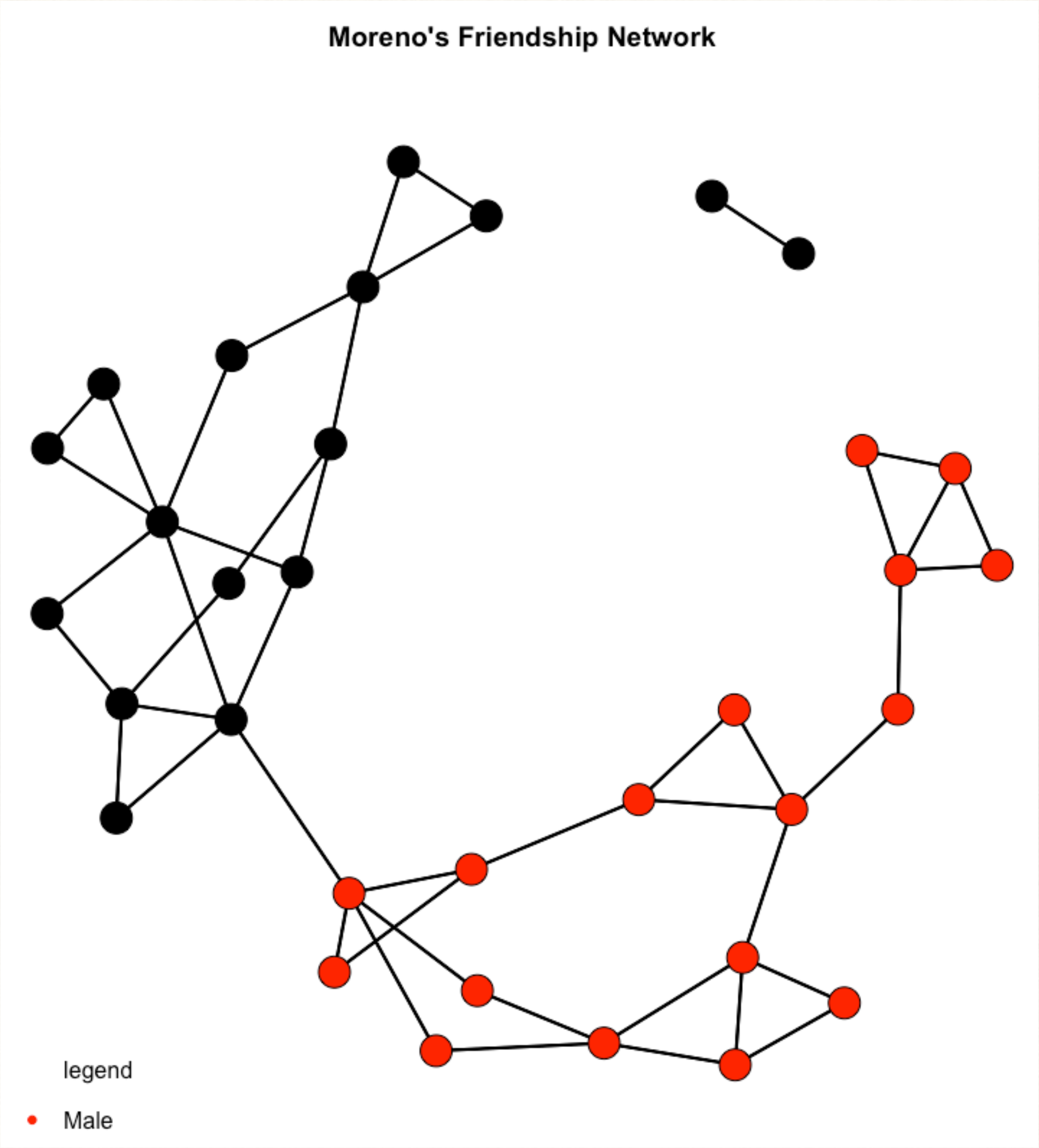
Example

What pattern do you see in these data?



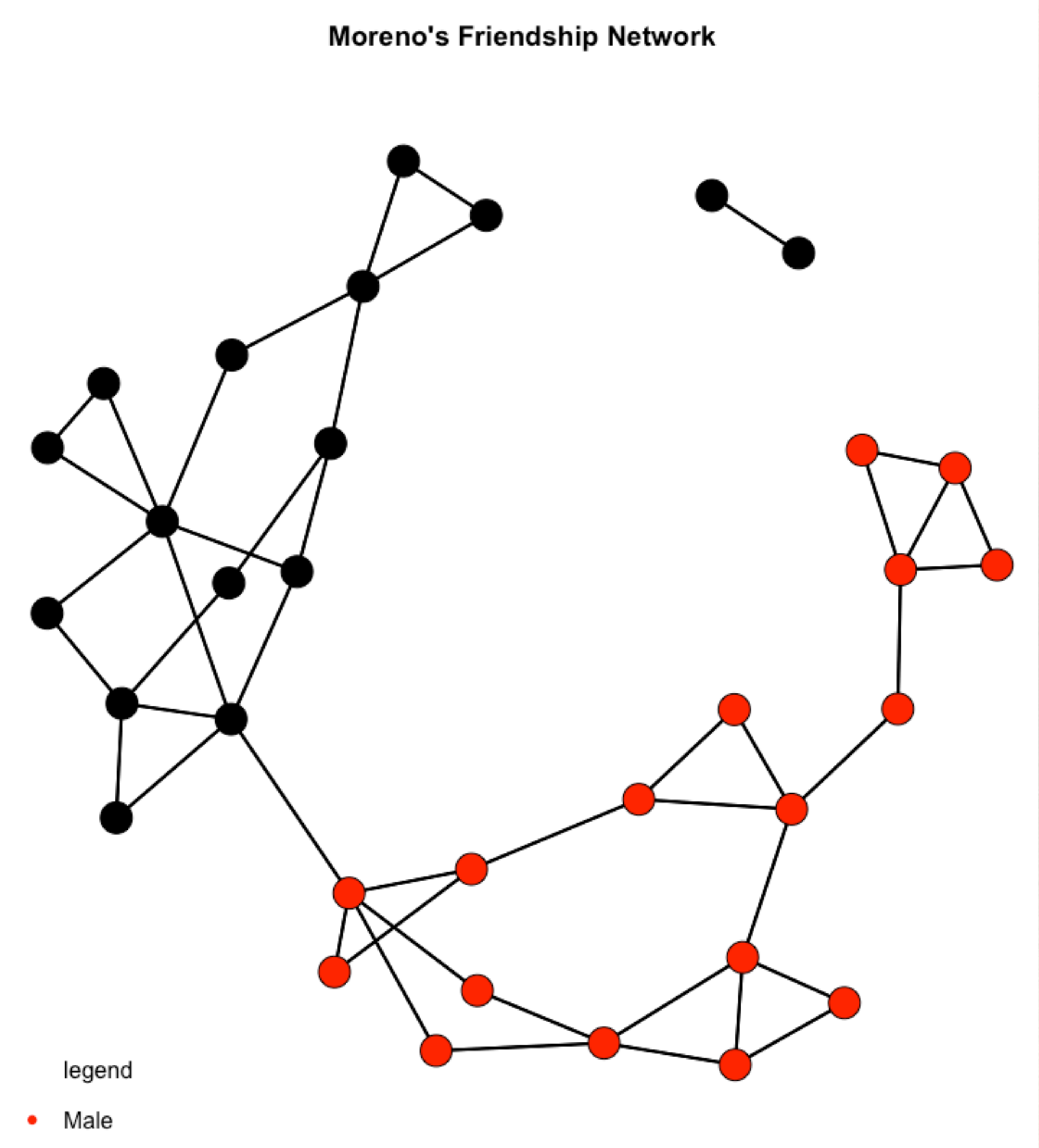
Example

Here, it is probably matching based on attributes.



Example

But, we can <u>test</u> that **hypothesis**!

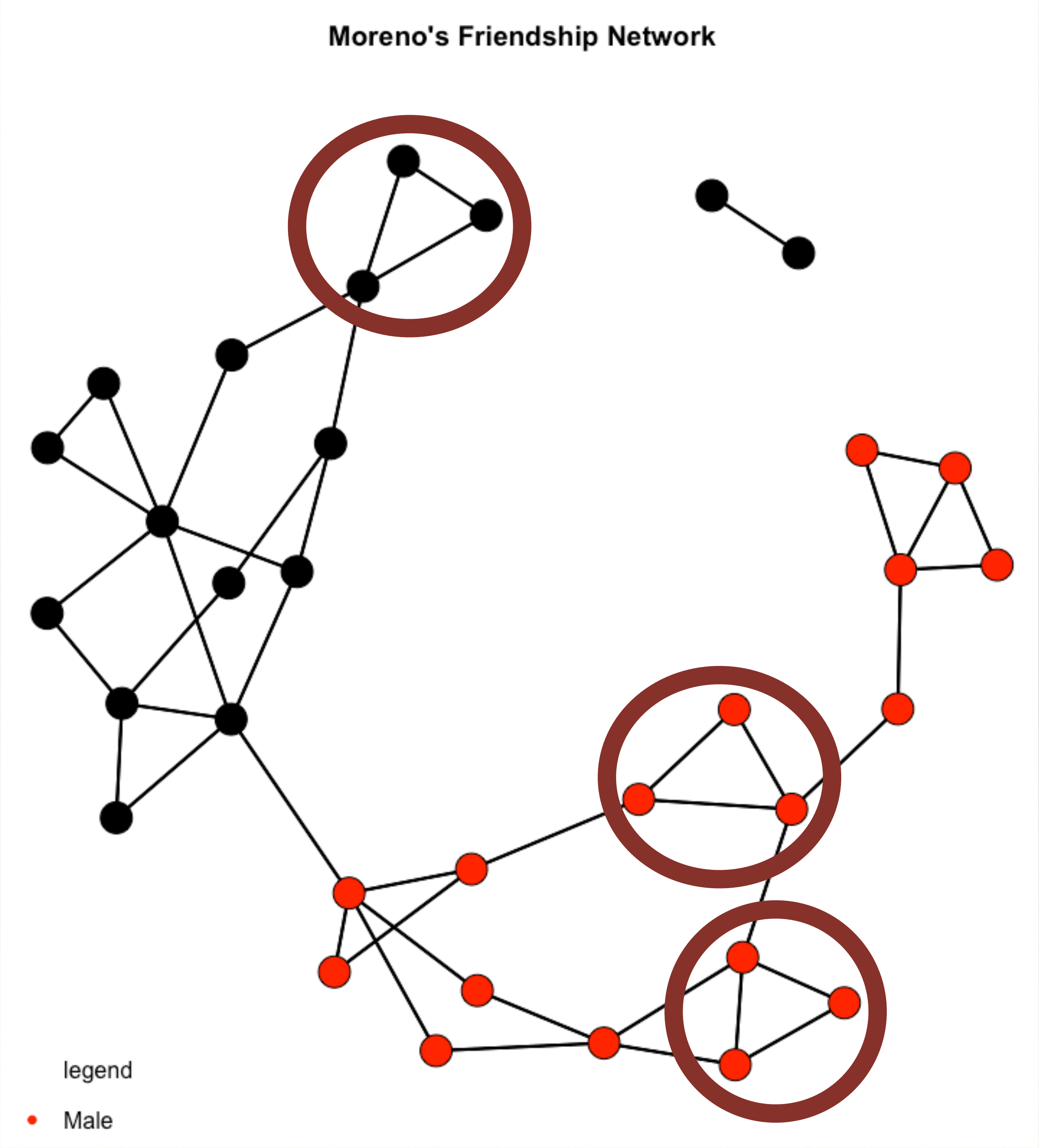


Example

- Think about a triad of individuals with the same attribute…

Example

- Think about a triad of individuals with the same attribute…



Example

- Think about a triad of individuals with the same attribute…
  - Did this configuration occur because:
    - People tend to choose friends who are like them? (“birds of a feather”)
    - People who have friends in common tend to become friends? (“friend of a friend”)

Emergent Structure

- Different processes can lead to similar outcomes.
  - We want to be able to fit these terms simultaneously and identify the independent effects of each process on the overall outcome.

Logic of Random Graph Modeling

- We can then evaluate how probable a particular network is among all possible networks that could exist on a set of actors.
  - The observed network is only one realization from a set of possible networks with similar characteristics (think back to the sampling example).
- Robins et al. (2007: 176)
  - "The range of possible networks, and their probability of occurrence under the model, is represented by a *probability distribution* on the set of all possible graphs"

Logic of Random Graph Modeling

- Once we have estimated the parameters of the probability distribution, we can sample graphs at random and compare their characteristics with those of the observed network.
  - If the model is good, then sampled graphs will resemble the observed network (visually and descriptively)
  - If this is the case, we can conjecture that the modeled structural effects could explain the emergence of the network.

Exponential Random Graph Formulation

- Express the probability of observing a tie between nodes i and j given some terms (i.e. network configurations).
  - A general framework for expressing different types of models.
    - Think of each model as "theory of network **dependence**".
      - We will look at four model types.

Dependence Assumptions

- <u>Edge independence (Bernoulii/Simple Random graphs)</u>
  - *How likely is a tie between* i *and* j?
  - **Erdos and Renyi (1959)**
    - The probability of a tie is the number of edges.

Exponential Random Graph Formulation

$$P(Y = y) = \left(\frac{1}{c}\right) exp\{\theta L(y)\}$$

Exponential Random Graph Formulation

Probability of a tie
being observed

$$P(Y = y) = \left(\frac{1}{c}\right) exp\{\theta L(y)\}$$

Exponential Random Graph Formulation

Probability of a tie
being observed

$$P(Y = y) = \left(\frac{1}{c}\right) exp\{\theta L(y)\}$$

Normalizing
constant

Exponential Random Graph Formulation

Probability of a tie being observed

$$P(Y = y) = \left(\frac{1}{c}\right) exp\{\theta L(y)\}$$

Normalizing constant

coefficient for the edge term

Exponential Random Graph Formulation

Probability of a tie being observed

Number of edges in y

$$P(Y = y) = \left(\frac{1}{c}\right) exp\{\theta L(y)\}$$

Normalizing constant

coefficient for the edge term

Exponential Random Graph Formulation

$$P(Y = y) = \left(\frac{1}{c}\right) exp\{\theta L(y)\}$$

Note the absence of
any dependence
with other edges

Dependence Assumptions

- Edge independence (Bernoulii/Simple Random graphs)
  - Assumes all ties are independent of one another.
  - Assumes nodes do not vary in their tie propensity.
    - Does a poor job capturing: clustering, the degree distribution(s).
  - But, a **baseline** model: a good reference for comparing more complex models.

Dependence Assumptions

- <u>Dyadic independence (p1 models)</u>
  - *How likely is a tie between* i *and* j*?*
  - **Holland and Leinhardt (1981)**
    - Depends on the <u>attractiveness</u> of the node, as nodes differ in their indegree.
    - Depends on whether the tie is <u>reciprocal</u>, Did *j* send a tie to *i*?

Exponential Random Graph Formulation

$$P(Y=y) \propto exp\left(\mu L(y) + \sum_{i}^{N} \alpha_i y_{i+} + \sum_{j}^{N} \beta_j y_{+j} + \rho M(y)\right)$$

Number of edges in the network

Exponential Random Graph Formulation

Number of
outgoing ties

$$P(Y = y) \propto exp\left(\mu L(y) + \sum_{i}^{N} \alpha_i y_{i+} + \sum_{j}^{N} \beta_j y_{+j} + \rho M(y) \right)$$

Number of
incoming ties

Number of edges in
the network

Exponential Random Graph Formulation

Number of outgoing ties

Number of mutual ties

$$P(Y = y) \propto exp\left(\mu L(y) + \sum_{i}^{N} \alpha_i y_{i+} + \sum_{j}^{N} \beta_j y_{+j} + \rho M(y)\right)$$

Number of incoming ties

Number of edges in the network

Exponential Random Graph Formulation

$$P(Y = y) \propto exp\left(\mu L(y) + \sum_{i}^{N} \alpha_i y_{i+} + \sum_{j}^{N} \beta_j y_{+j} + \rho M(y)\right)$$

Note the absence of terms for **other** dyads.

Dependence Assumptions

- Dyadic independence (p1 models)
  - Assumes that two dyads are conditionally **independent**.
    - A tie between *i* and *j* does not depend on a tie with *k*.
      - Does a poor job capturing transitivity in networks.

Dependence Assumptions

- Dyadic dependence (p* models/Markov graphs)
  - *How likely is a tie between* i *and* j*?*
  - **Frank and Strauss (1986)**
    - Tie probability between *i* and *j* depends on ties that *i* and *j* have with others.
      - Example: Tie between Chris and Lisa is dependent on Lisa's relationship with Ewan.
    - Edges that do not have a node in common are conditionally independent (Markov assumption).

Exponential Random Graph Formulation

$$P(Y = y) = \left(\frac{1}{c}\right) exp\{\theta L(y) + \sigma_k S_k(y) + ... + \tau T(y)\}$$

Number of edges in
the network

Exponential Random Graph Formulation

Number of k-star configurations in the network

$$P(Y=y)=\left(\frac{1}{c}\right)exp\{\theta L(y)+\sigma_k S_k(y)+...+\tau T(y)\}$$

Number of edges in the network

Exponential Random Graph Formulation

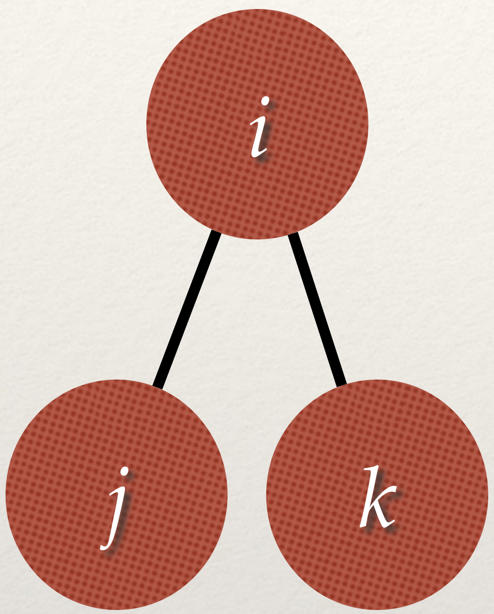
Number of k-star configurations in the network

$$P(Y = y) = \left(\frac{1}{c}\right) exp\{\theta L(y) + \sigma_k S_k(y) + ... + \tau T(y)\}$$

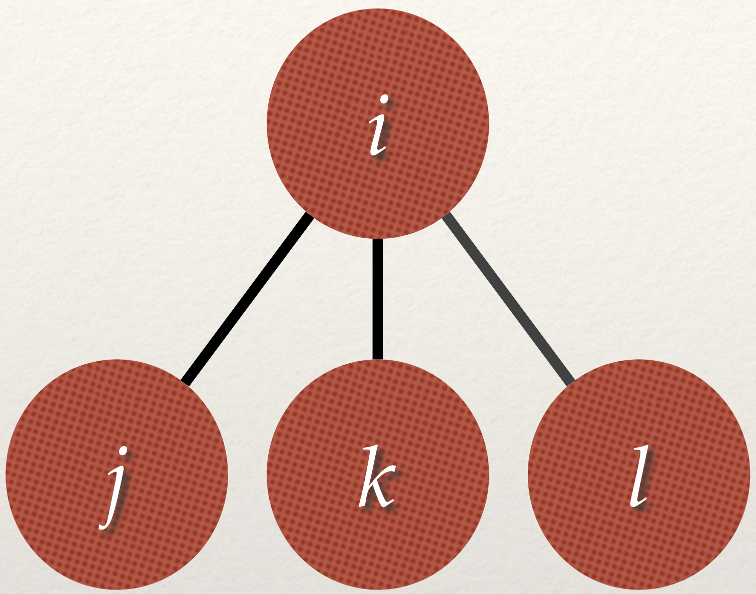
Number of edges in the network

Number of of triangles in the network

Configurations

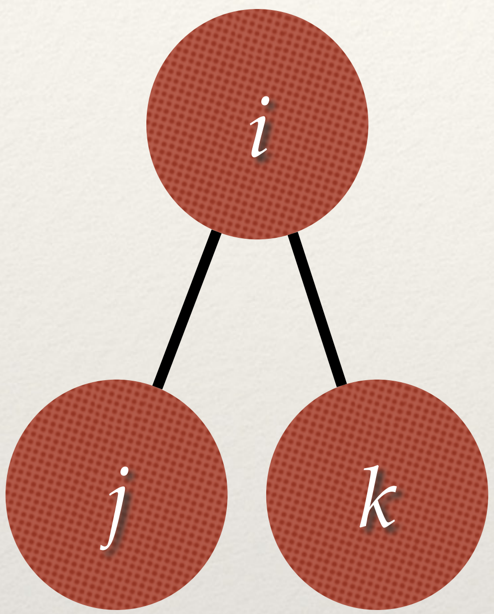


2-star

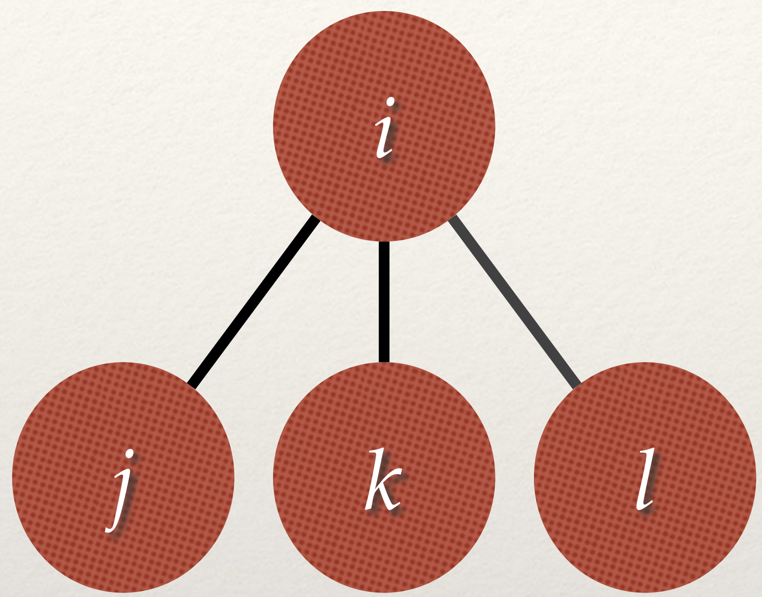


3-star

Configurations

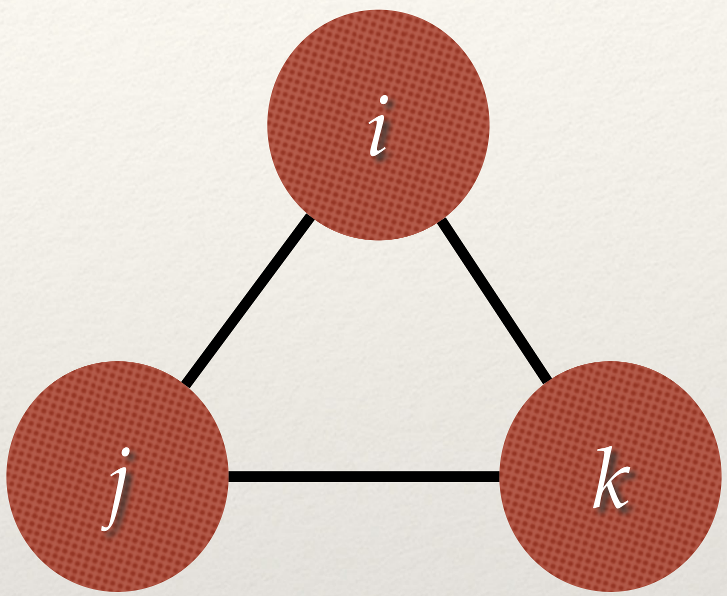


2-star



3-star

…



Triangle

Exponential Random Graph Formulation

$$P(Y = y) = \left(\frac{1}{c}\right) exp\{\theta L(y) + \sigma_k S_k(y) + ... + \tau T(y)\}$$

Counting these up,
allows us to
represent the
dependence
between the dyads.

Dependence Assumptions

- Dyadic dependence (p* models/Markov graphs)
    - The original formulation could not handle actor covariates.
        - Example: Do males send more ties?
    - The model was later extended to include actor covariates.

Exponential Random Graph Formulation

All dyads except *i* and *j*.

$$logit\left(P\left(Y_{ij} = 1 \mid n\ actors, Y_{ij}^{C}\right)\right) = \sum_{k=1}^{\kappa} \theta_k \delta_{z_k(y)}$$

Exponential Random Graph Formulation

All dyads except i and j.

$$logit\left(P\left(Y_{ij} = 1 \mid n\ actors, Y_{ij}^{C}\right)\right) = \sum_{k=1}^{\kappa} \theta_k \delta_{z_k(y)}$$

Coefficients for network statistics

Exponential Random Graph Formulation

All dyads except *i* and *j*.

Change statistic

$$logit\left(P\left(Y_{ij} = 1 \mid n\ actors, Y_{ij}^{C}\right)\right) = \sum_{k=1}^{\kappa} \theta_k \delta_{z_k(y)}$$

Coefficients for network statistics

Exponential Random Graph Formulation

All dyads except *i* and *j*.

Change statistic

$$logit\left(P\left(Y_{ij} = 1 \mid n\; actors, Y_{ij}^{C}\right)\right) = \sum_{k=1}^{\kappa} \theta_k \delta_{z_k(y)}$$

Looks like a **logistic regression**, right?!?

Coefficients for network statistics

Dependence Assumptions

- Dyadic dependence (p* models/Markov graphs)
  - It is like a logistic regression, except:
    - Change statistic is not just the value of the independent variable.
    - Conditional statement of left-hand side of equation (logistic regression assumes independence across units).

Dependence Assumptions

- Higher-Order Dependence Models
    - p* models struggle with *degeneracy*, meaning that the networks simulated from the model do not match well with the observed network.
        - Recent work has addressed this problem by defining more complex dependencies among dyads.
            - Examples:
                - Social circuit models.
                - Geometrically weighted terms.

ERGM Theory

- *So what do we do with this?*
  - Robins and Lusher (2013: 11)
    - "It is a theoretical and empirical task to delineate the various forms of dependence that are exhibited in actual social structures. We regard this as social network theory at a fundamental level…"
    - "The process of theory translation requires the alignment of theoretical concepts with network configurations."

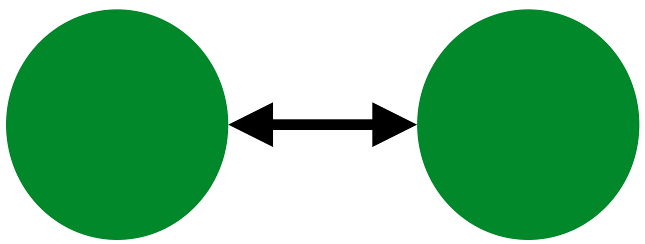
Network Configurations and Processes

- Conceptualization and Operationalization!
    - What is happening? (Conceptualization)
    - What would that “look like”? (Operationalization)

Configurations and Processes

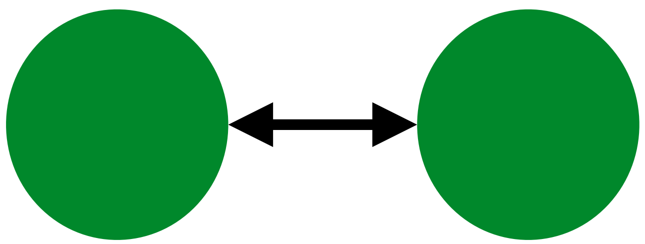
	Configuration	Process
		Friendship is reciprocated

Configurations and Processes



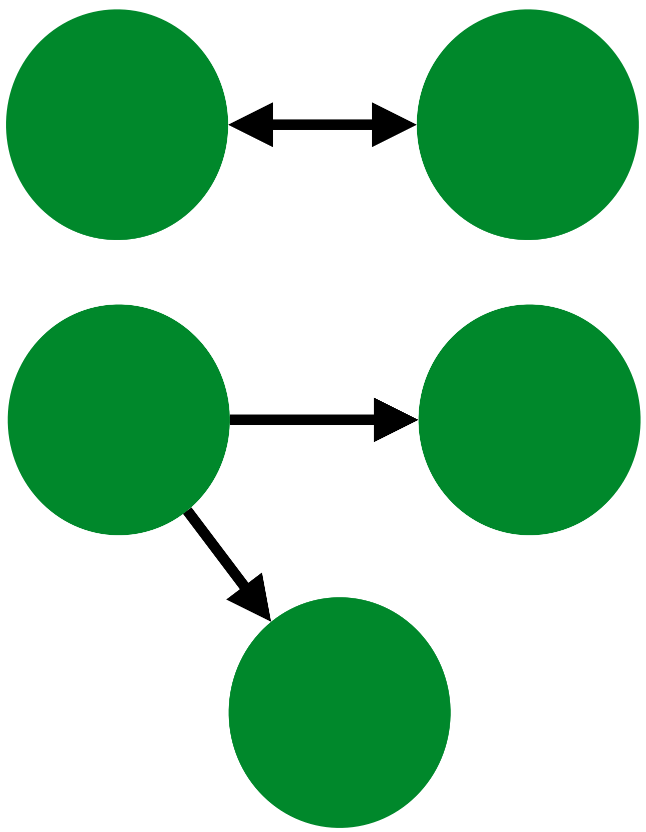
	Configuration	Process
	Mutuality	Friendship is reciprocated

Configurations and Processes



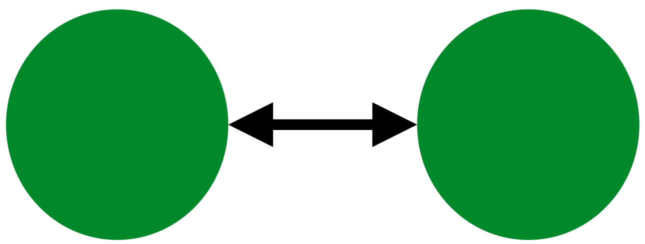
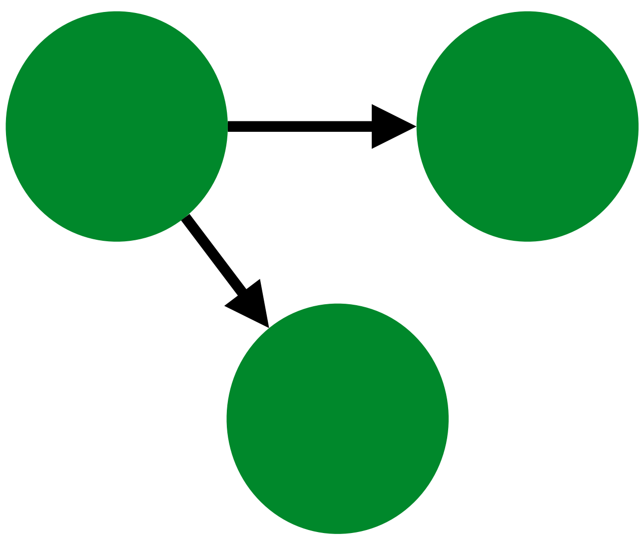
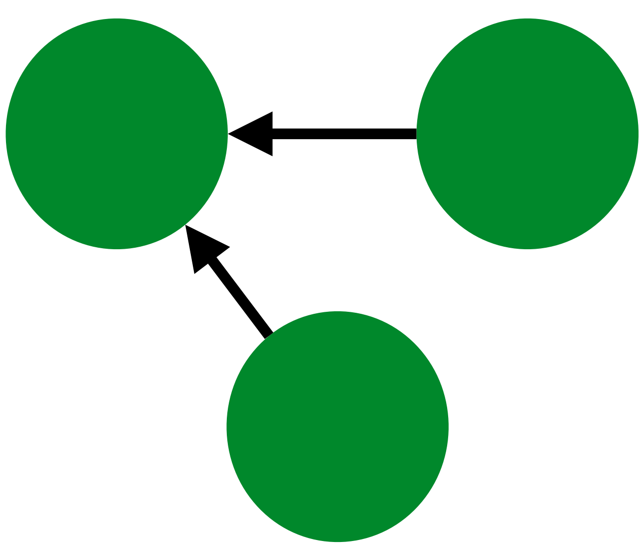
Configuration	Process
Mutuality	Friendship is reciprocated
	Some people are very social

Configurations and Processes

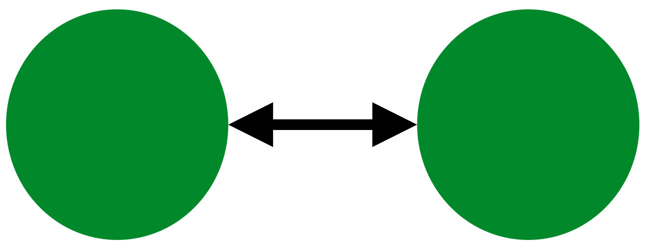
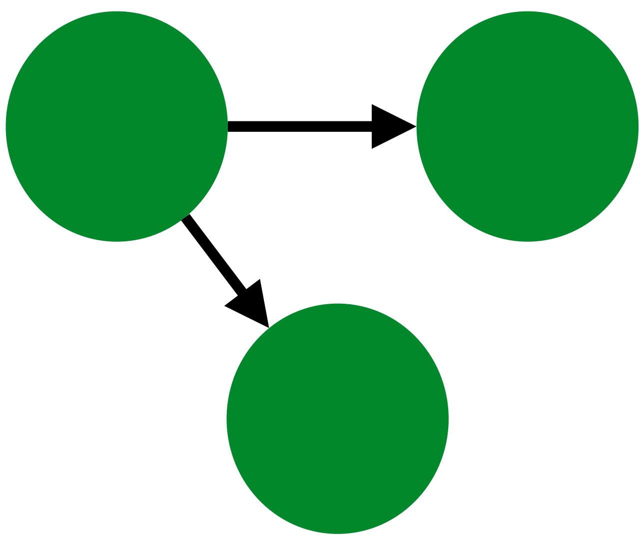
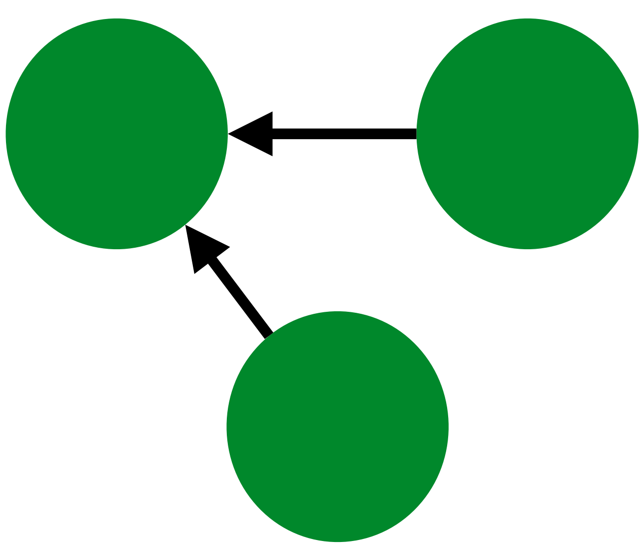


Configuration	Process
Mutuality	Friendship is reciprocated
Outdegree	Some people are very social

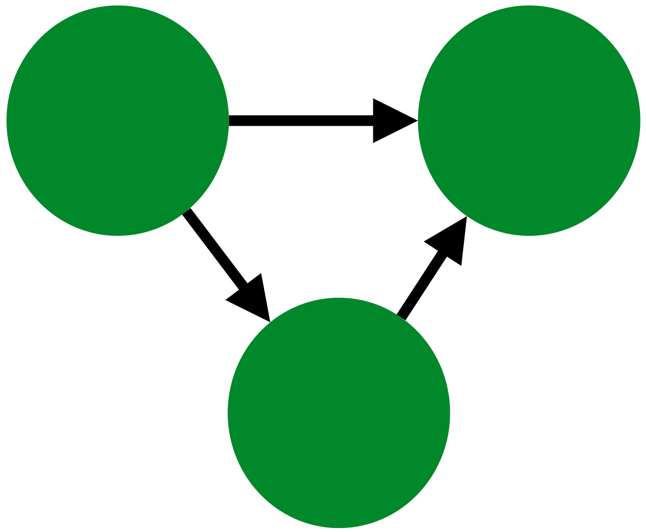
Configurations and Processes

	Configuration	Process
	Mutuality	Friendship is reciprocated
	Outdegree	Some people are very social
		Some people are very popular

Configurations and Processes

	Configuration	Process
	Mutuality	Friendship is reciprocated
	Outdegree	Some people are very social
	Indegree	Some people are very popular

Configurations and Processes

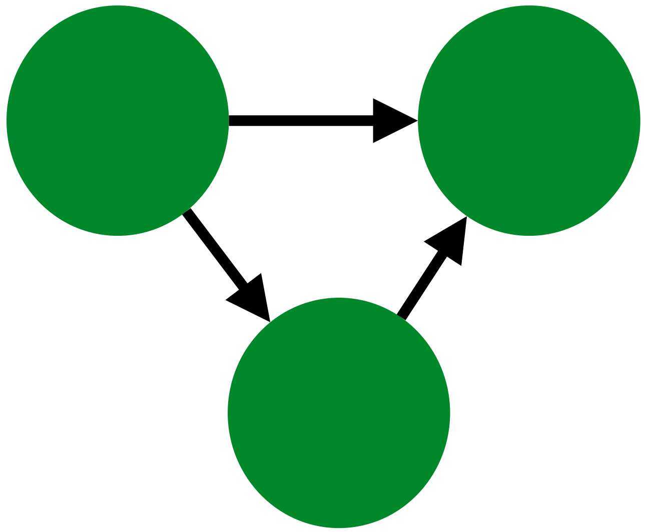


Configuration

Process

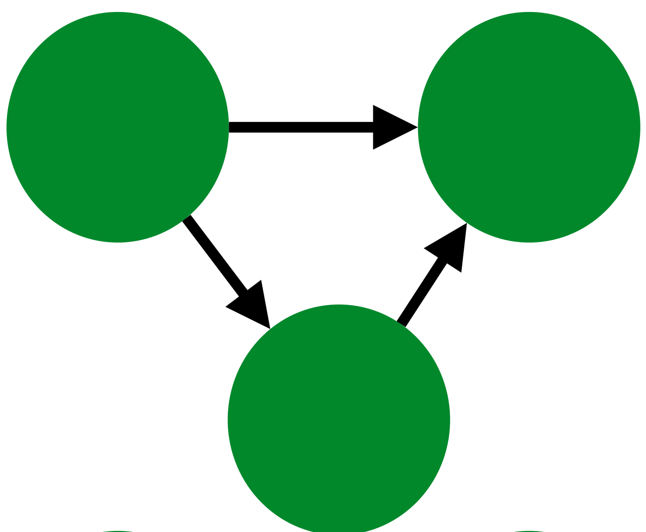
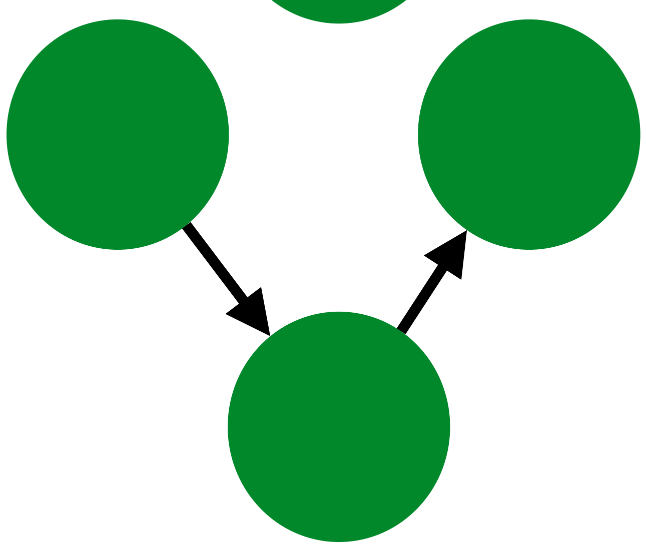
Friends of friends become friends

Configurations and Processes

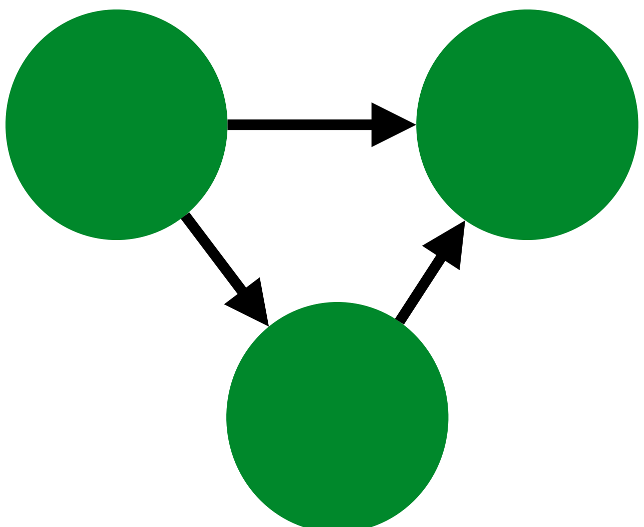
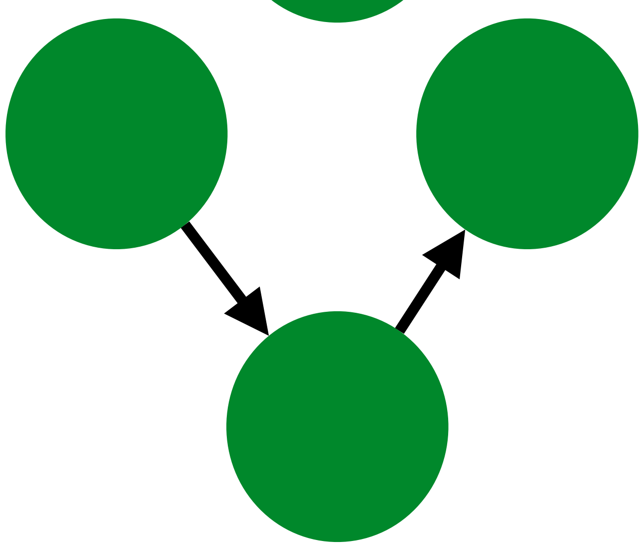


Configuration	Process
Transitivity	Friends of friends become friends

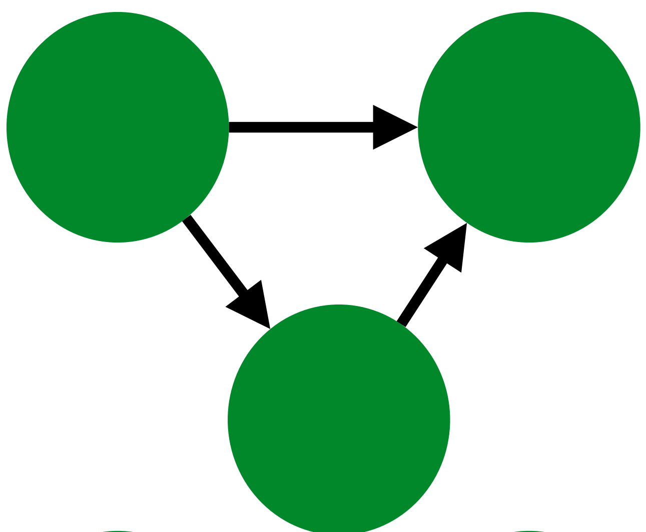
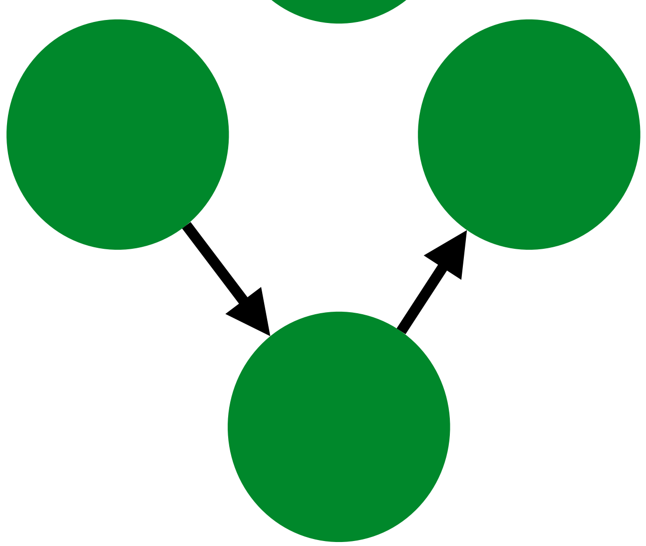
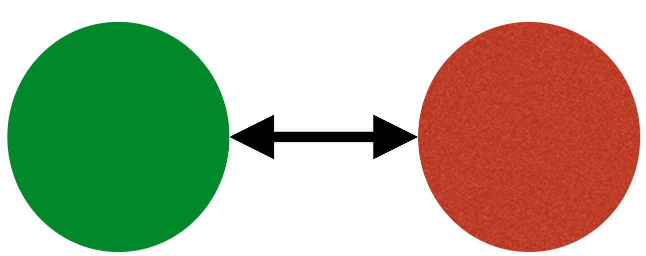
Configurations and Processes

	Configuration	Process
	Transitivity	Friends of friends become friends
		Brokerage structures are preferred

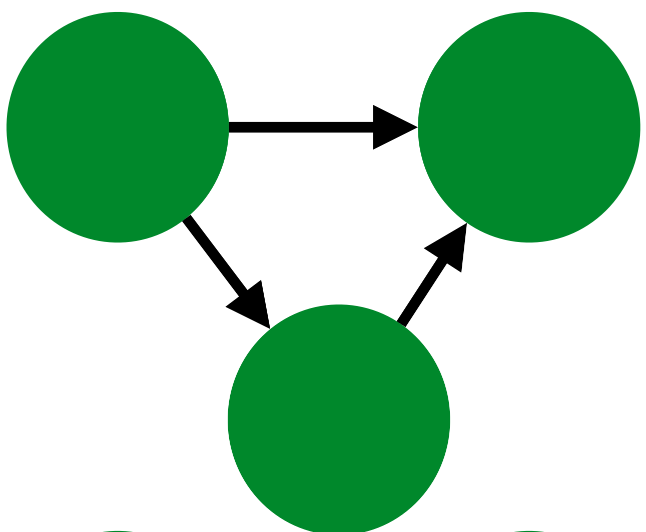
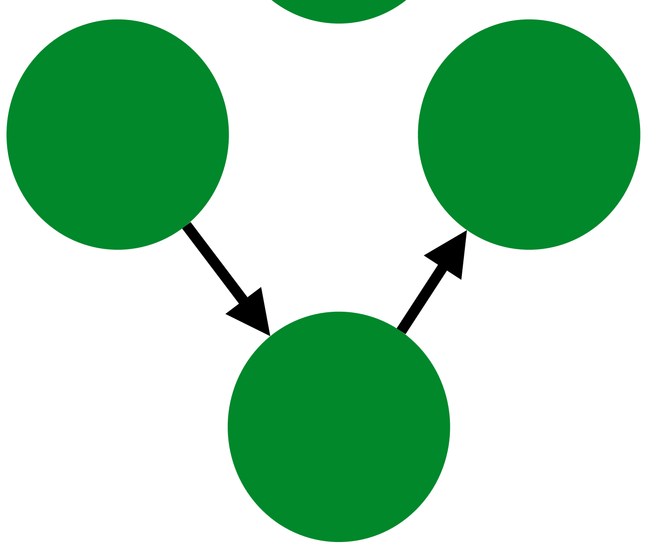
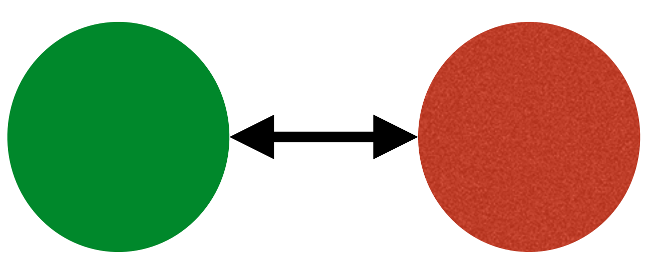
Configurations and Processes

	Configuration	Process
	Transitivity	Friends of friends become friends
	2-path	Brokerage structures are preferred

Configurations and Processes

	Configuration	Process
	Transitivity	Friends of friends become friends
	2-path	Brokerage structures are preferred
		Opposites attract

Configurations and Processes

	Configuration	Process
	Transitivity	Friends of friends become friends
	2-path	Brokerage structures are preferred
	Heterophily	Opposites attract

From Lusher, Koskinen, and Robins "Exponential Random Graph Models of Social Networks: Theory, Methods, and Applications

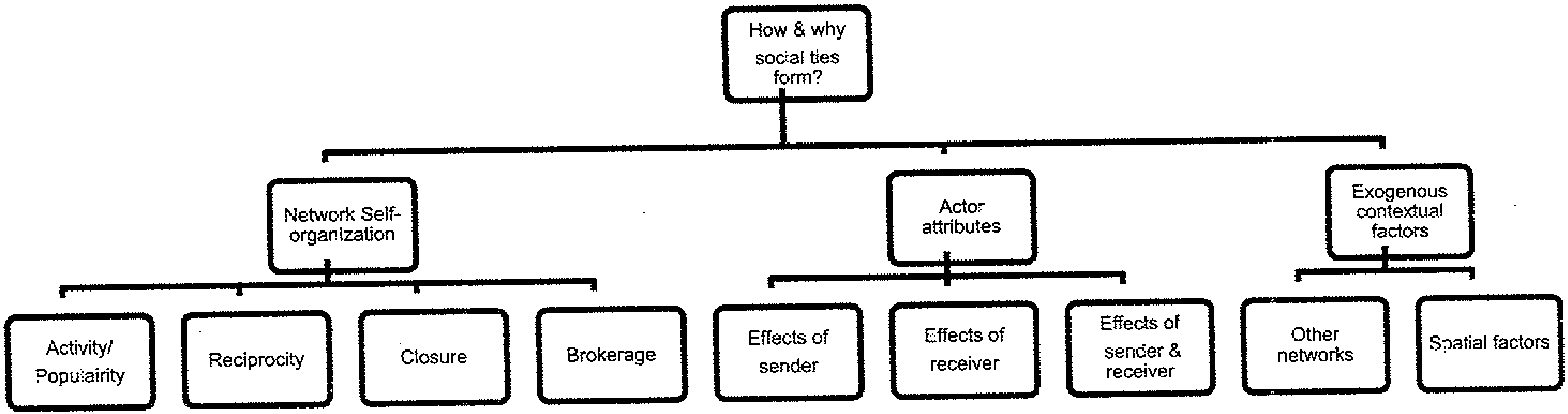


Figure 3.3. Conceptual framework for processes of social tie formation discussed in this book.

Network Configurations and Processes

- As we work through developing a model, we want to identify the **network configurations** that capture the theoretical process we are interested in testing.

Learning Goals

- Understand the logic of the exponential random graph model.
- Understand the historical development of exponential random graph models.
- Describe the properties of various types of exponential random graph models.
- Understand the notion of “network configurations” as operationalizations of theoretical concepts.

Questions?