Network Modeling

Viviana Amati Jürgen Lerner Bobo Nick

Dept. Computer & Information Science
University of Konstanz

Winter 2011/2012
(last updated: November 30, 2011)

Outline.

Outline.

Topic of this lecture.

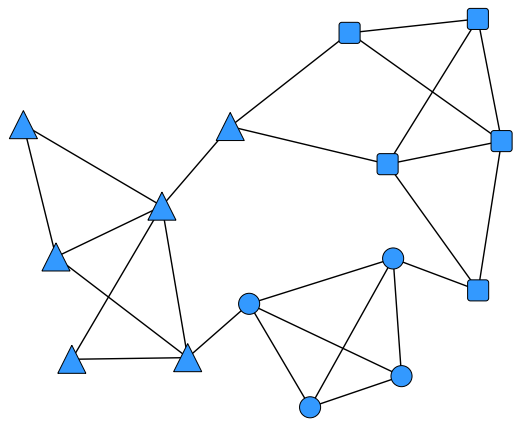
Statistical models for social network data.

Topic of this lecture.

Statistical models for social network data.

Social networks consist of **actors** and **relations** among them.

- **actors:** persons, organizations, companies, countries, ...
- **relations:** friendship, asking for advice, communication, collaboration, trade, war, ...



Topic of this lecture.

Statistical models for social network data.

Data availability improved largely over the last decade.

- traditional data collection, e. g., by questionaires "*please name your best friends*"
- more and more **automatically logged data** from electronic communication and collaboration: telephone calls, email, online social networks, online markets, recommender systems, wikis, ...

$\Rightarrow$ opportunity and challenge for data-driven social science

Topic of this lecture.

Statistical models for social network data.

Statistics can formulate precise statements about **uncertainty**.

Different sources of uncertainty:

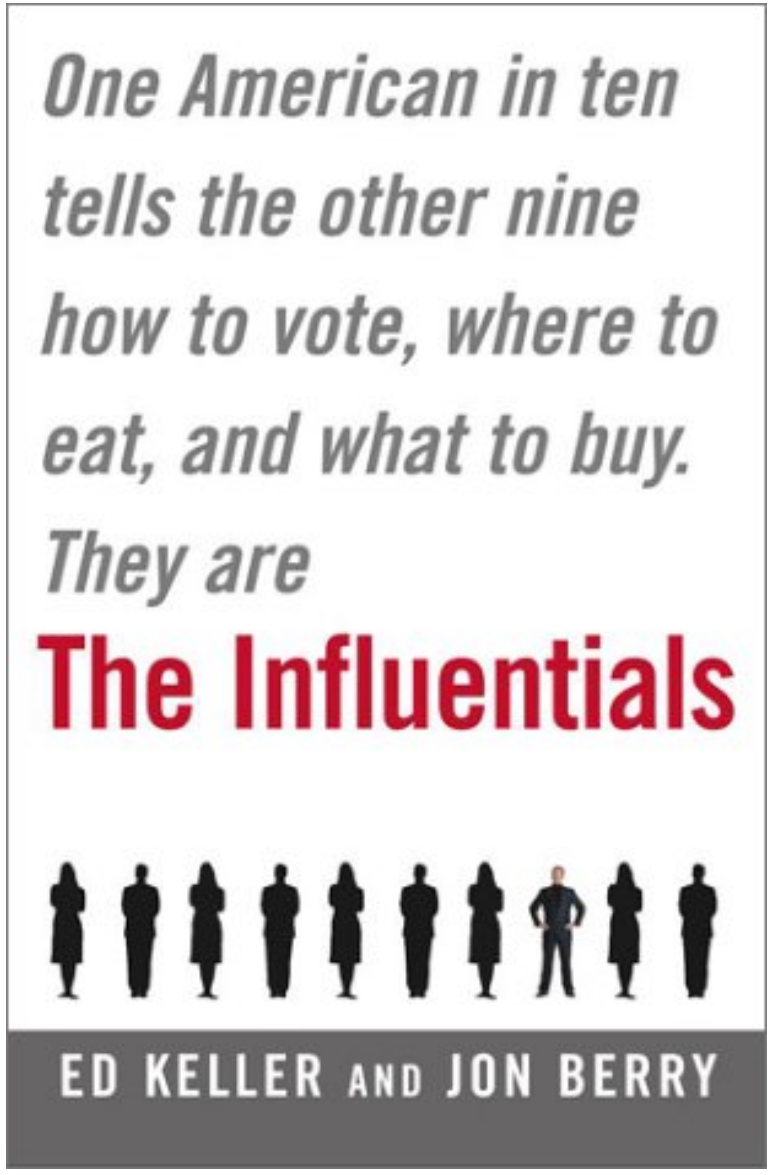
- (potential errors in the data);
- *what would happen, if we measured the data again?*
  - at a different point in time,
  - on a different set of actors,
  - with different environmental factors, ...

  expected outcome $\pm$ **variability**

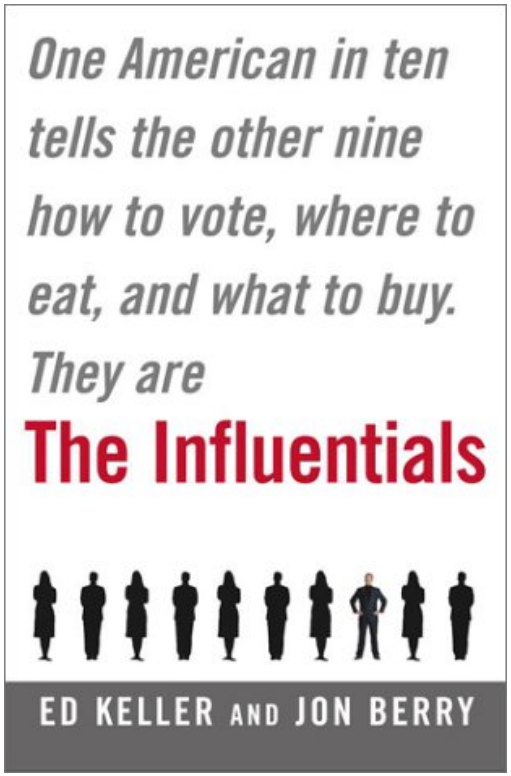
  $\Rightarrow$ to explain and predict social relations and behavior

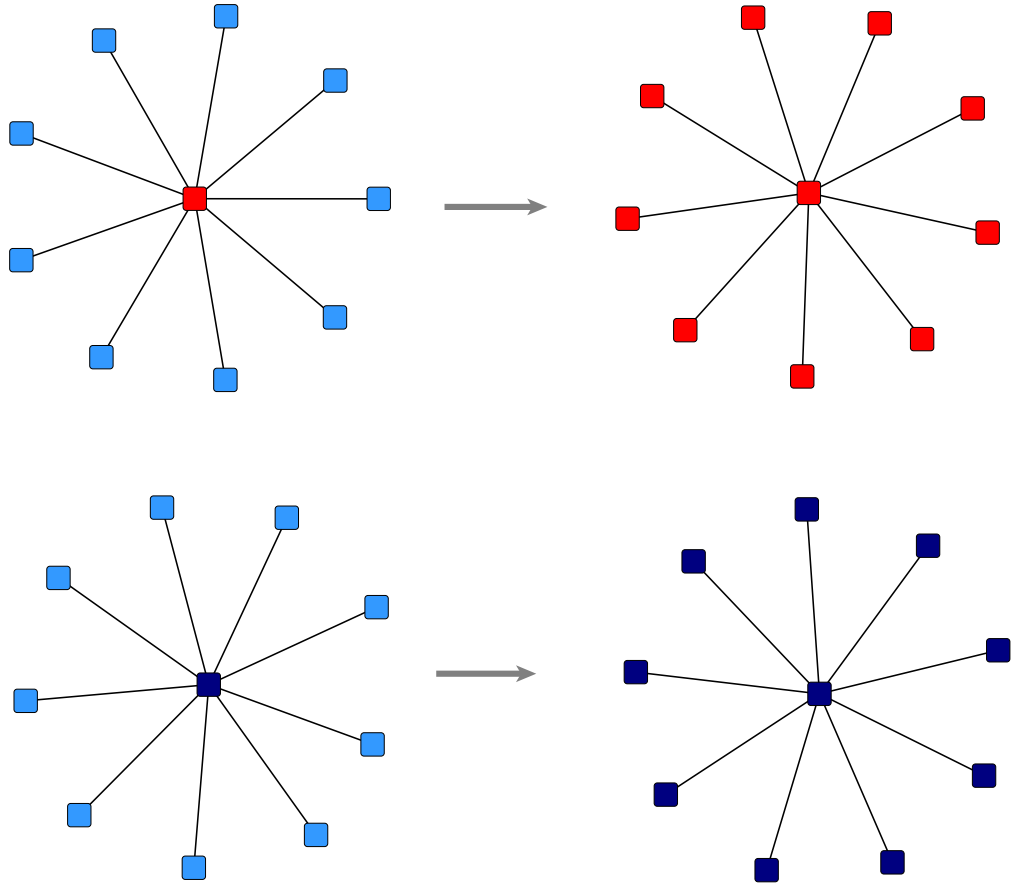
Illustrative application: assessment of **social influence**.

Social influence and network-based marketing.



Social influence and network-based marketing.





Proposition is a bit too optimistic. . .
but approximate pattern might be empirically observable.

Social influence and network-based marketing.

Empirical validation.

Hill, Provost, and Volinsky (2006): Network-Based Marketing. *Statistical Science* 21(2):256–276.

Data: Derived from a direct-mail marketing campaign of a telecommunications firm to promote a new product/service.

- "Traditional" variables: loyalty to firm, interest in high-tech products, early adopter, ...
- Network variable: communicated with product adopter.
- Outcome: subsequently purchase or not.

Hypothesis: Customers connected to an adopter have a higher probability to purchase the product.

Social influence and network-based marketing.

Empirical validation.

Hill, Provost, and Volinsky (2006): Network-Based Marketing. *Statistical Science* 21(2):256–276.

Data: Derived from a direct-mail marketing campaign of a telecommunications firm to promote a new product/service.

- "Traditional" variables: loyalty to firm, interest in high-tech products, early adopter, ...
- Network variable: communicated with product adopter.
- Outcome: subsequently purchase or not.

Hypothesis: Customers connected to an adopter have a higher probability to purchase the product.

Social influence and network-based marketing.

Empirical validation.

Hill, Provost, and Volinsky (2006): Network-Based Marketing. *Statistical Science* 21(2):256–276.

Data: Derived from a direct-mail marketing campaign of a telecommunications firm to promote a new product/service.

- "Traditional" variables: loyalty to firm, interest in high-tech products, early adopter, ...
- Network variable: communicated with product adopter.
- Outcome: subsequently purchase or not.

Hypothesis: Customers connected to an adopter have a higher probability to purchase the product.

Social influence and network-based marketing.

Empirical validation.

Hill, Provost, and Volinsky (2006): Network-Based Marketing. *Statistical Science* 21(2):256–276.

Data: Derived from a direct-mail marketing campaign of a telecommunications firm to promote a new product/service.

- "Traditional" variables: loyalty to firm, interest in high-tech products, early adopter, ...
- Network variable: communicated with product adopter.
- Outcome: subsequently purchase or not.

Hypothesis: Customers connected to an adopter have a higher probability to purchase the product.

Social influence and network-based marketing.

Empirical validation.

Hill, Provost, and Volinsky (2006): Network-Based Marketing. *Statistical Science* 21(2):256–276.

Data: Derived from a direct-mail marketing campaign of a telecommunications firm to promote a new product/service.

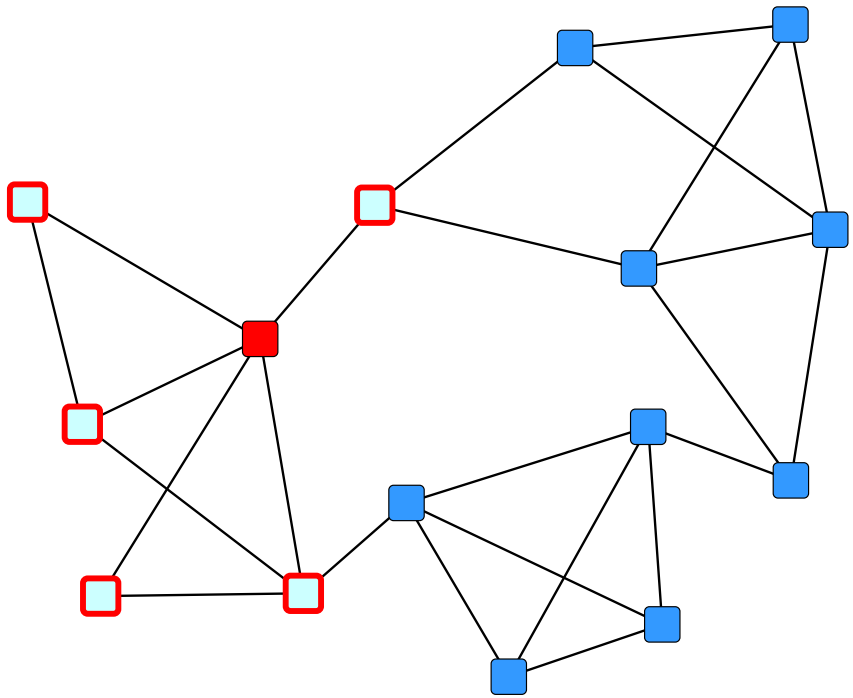
- "Traditional" variables: loyalty to firm, interest in high-tech products, early adopter, ...
- Network variable: communicated with product adopter.
- Outcome: subsequently purchase or not.

Hypothesis: Customers connected to an adopter have a higher probability to purchase the product.

Social influence and network-based marketing.

graphically

One customer has already bought the product; some are connected to this early adopter.



Do customers connected to an early adopter have a higher probability to purchase the product?

Social influence and network-based marketing.

Empirical validation (continued).

Research question: do customers connected to an early adopter have a higher probability to purchase the product?

Method: all potential customers are classified into *marketing segments* determined by traditional variables.

For each segment separately, compare

$$\frac{P(\text{purchase} \mid \text{connected to adopter})}{P(\text{purchase})}$$

Social influence and network-based marketing.

Empirical validation (continued).

Method: all potential customers are classified into *marketing segments* determined by traditional variables.

For each segment separately, compare

$$\frac{P(\text{purchase} \mid \text{connected to adopter})}{P(\text{purchase})}$$

Results: customers connected to product adopter have a purchase probability that is 3–5 times higher.

Network analysis can identify prospective customers ignored by traditional marketing strategies.

Social influence and network-based marketing.

Empirical validation (continued).

Method: all potential customers are classified into *marketing segments* determined by traditional variables.

For each segment separately, compare

$$\frac{P(\text{purchase} \mid \text{connected to adopter})}{P(\text{purchase})}$$

Results: customers connected to product adopter have a purchase probability that is 3–5 times higher.

Network analysis can identify prospective customers ignored by traditional marketing strategies.

Social influence and network-based marketing.

Empirical validation (continued).

Method: all potential customers are classified into *marketing segments* determined by traditional variables.

For each segment separately, compare

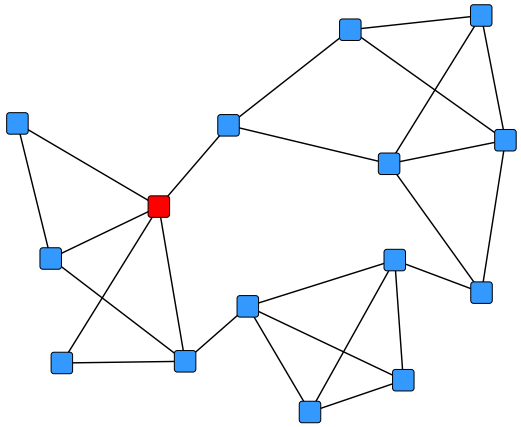
$$\frac{P(\text{purchase} \mid \text{connected to adopter})}{P(\text{purchase})}$$

Results: customers connected to product adopter have a purchase probability that is 3–5 times higher.

Network analysis can identify prospective customers ignored by traditional marketing strategies.

Social influence and network-based marketing.

What is the mechanism of this social influence?



One customer has bought the product.

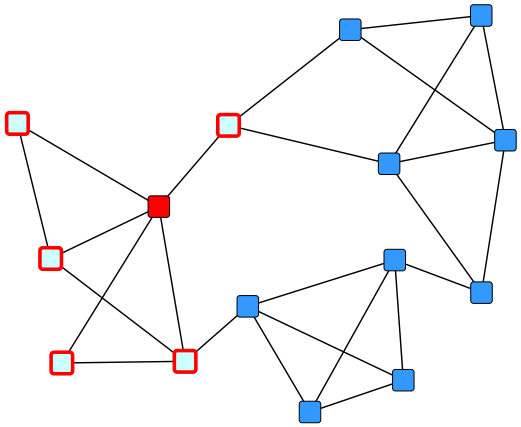
His/her friends

- become aware of the product;
- receive recommendation;
- and/or desire to have it.

Detailed mechanism of social influence has not been validated in Hill, Provost, and Volinsky (2006).

Social influence and network-based marketing.

What is the mechanism of this social influence?



One customer has bought the product.

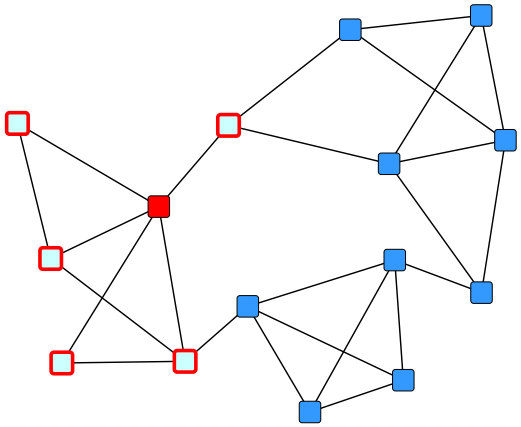
His/her friends

- become aware of the product;
- receive recommendation;
- and/or desire to have it.

Detailed mechanism of social influence has not been validated in Hill, Provost, and Volinsky (2006).

Social influence and network-based marketing.

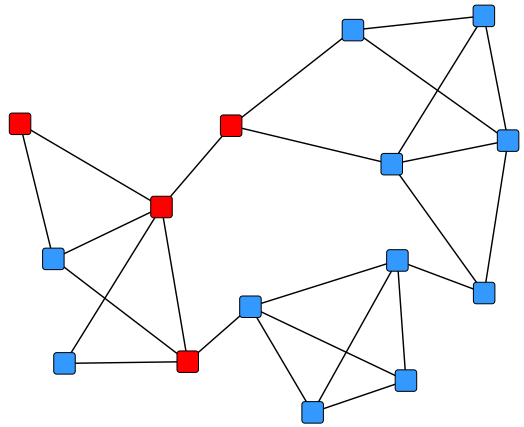
What is the mechanism of this social influence?



One customer has bought the product.

His/her friends

- become aware of the product;
- receive recommendation;
- and/or desire to have it.

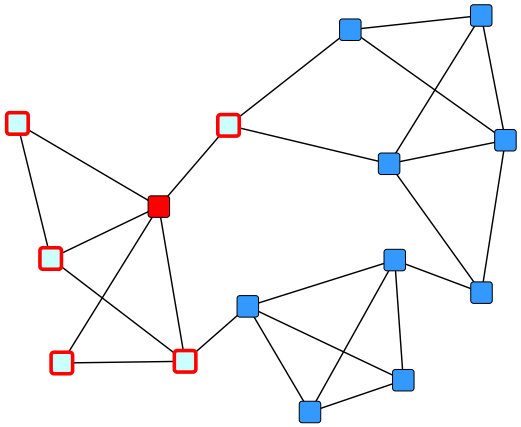


Some of them buy the product.

Detailed mechanism of social influence has not been validated in Hill, Provost, and Volinsky (2006).

Social influence and network-based marketing.

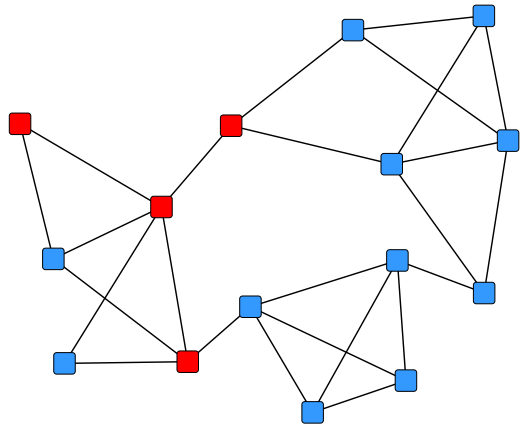
What is the mechanism of this social influence?



One customer has bought the product.

His/her friends

- become aware of the product;
- receive recommendation;
- and/or desire to have it.



Some of them buy the product.

Detailed mechanism of social influence has not been validated in Hill, Provost, and Volinsky (2006).

Other stories about social influence . . .

Spread of obesity.

Christakis and Fowler (2007): The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine* 357:370–379.

Data (Framingham Heart Study): health data about 12,000 people from 1971–2003, including body mass index and various social relations.

Key result:

> *A person's chances of becoming obese increased by 57% [. . .] if he or she had a friend who became obese [. . .].*

What are the mechanisms of obesity contagion?

Spread of obesity.

Christakis and Fowler (2007): The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine* 357:370–379.

Data (Framingham Heart Study): health data about 12,000 people from 1971–2003, including body mass index and various social relations.

Key result:

> *A person's chances of becoming obese increased by 57% [. . .] if he or she had a friend who became obese [. . .].*

What are the mechanisms of obesity contagion?

Spread of obesity.

Christakis and Fowler (2007): The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine* 357:370–379.

Data (Framingham Heart Study): health data about 12,000 people from 1971–2003, including body mass index and various social relations.

Key result:

> *A person's chances of becoming obese increased by 57% [. . .] if he or she had a friend who became obese [. . .].*

What are the mechanisms of obesity contagion?

Spread of happiness and smoking behavior.

A person's chances of becoming obese increased by 57% [. . .] if he or she had a friend who became obese [. . .].

Other results on the same or similar datasets

- happy friends make an individual happier;
- individuals have a higher probability to start smoking if they are friends of smokers; . . .

Everything seems to spread through networks.

Spread of happiness and smoking behavior.

A person's chances of becoming obese increased by 57% [. . .] if he or she had a friend who became obese [. . .].

Other results on the same or similar datasets

- happy friends make an individual happier;
- individuals have a higher probability to start smoking if they are friends of smokers; . . .

Everything seems to spread through networks.

Spread of happiness and smoking behavior.

A person's chances of becoming obese increased by 57% [. . .] if he or she had a friend who became obese [. . .].

Other results on the same or similar datasets

- happy friends make an individual happier;
- individuals have a higher probability to start smoking if they are friends of smokers; . . .

Everything seems to spread through networks.

Spread of happiness and smoking behavior.

A person's chances of becoming obese increased by 57% [. . .] if he or she had a friend who became obese [. . .].

Other results on the same or similar datasets

- happy friends make an individual happier;
- individuals have a higher probability to start smoking if they are friends of smokers; . . .

Everything seems to spread through networks.

Criticism of popular network analysis methods.

Cohen-Cole and Fletcher (2008): Detecting Implausible Social Network Effects in Acne, Height, and Headaches. *British Medical Journal* 337:a2533.

Data: Add Health Study.

Results: using popular SNA methods it can be validated that

- people whose friends have skin diseases tend to develop skin diseases;
- individuals with tall friends become taller;
- the likelihood of headaches increases with the presence of a friend with headaches.

Conclusion?

Criticism of popular network analysis methods.

Cohen-Cole and Fletcher (2008): Detecting Implausible Social Network Effects in Acne, Height, and Headaches. *British Medical Journal* 337:a2533.

Data: Add Health Study.

Results: using popular SNA methods it can be validated that

- people whose friends have skin diseases tend to develop skin diseases;
- individuals with tall friends become taller;
- the likelihood of headaches increases with the presence of a friend with headaches.

Conclusion?

Criticism of popular network analysis methods.

Cohen-Cole and Fletcher (2008): Detecting Implausible Social Network Effects in Acne, Height, and Headaches. *British Medical Journal* 337:a2533.

Data: Add Health Study.

Results: using popular SNA methods it can be validated that

- people whose friends have skin diseases tend to develop skin diseases;
- individuals with tall friends become taller;
- the likelihood of headaches increases with the presence of a friend with headaches.

Conclusion?

Criticism of popular network analysis methods.

Cohen-Cole and Fletcher (2008): Detecting Implausible Social Network Effects in Acne, Height, and Headaches. *British Medical Journal* 337:a2533.

Data: Add Health Study.

Results: using popular SNA methods it can be validated that

- people whose friends have skin diseases tend to develop skin diseases;
- individuals with tall friends become taller;
- the likelihood of headaches increases with the presence of a friend with headaches.

Conclusion?

Criticism of popular network analysis methods.

Cohen-Cole and Fletcher (2008): Detecting Implausible Social Network Effects in Acne, Height, and Headaches. *British Medical Journal* 337:a2533.

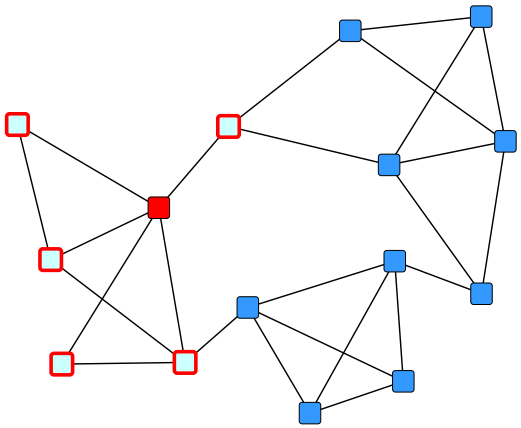
Data: Add Health Study.

Results: using popular SNA methods it can be validated that

- people whose friends have skin diseases tend to develop skin diseases;
- individuals with tall friends become taller;
- the likelihood of headaches increases with the presence of a friend with headaches.

Conclusion?

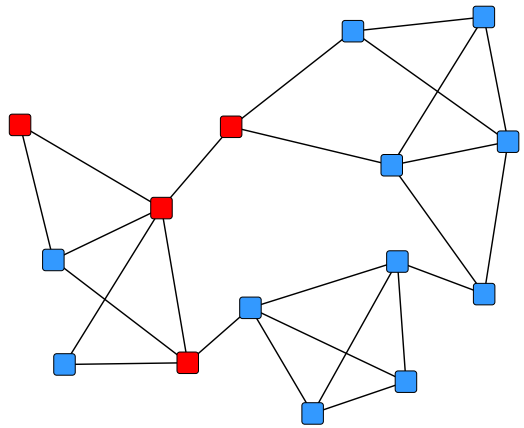
Revisiting social influence in purchase decisions.



One actor has bought the product.

His/her friends

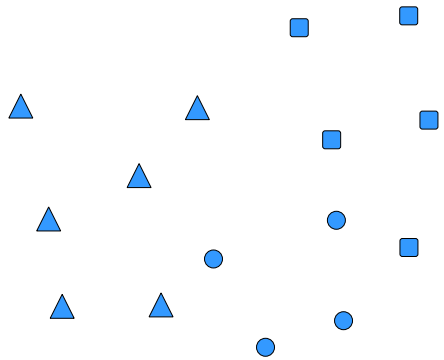
- become aware of the product;
- receive recommendation;
- and/or desire to have it.



Some of them buy the product.

Detailed mechanism (*middle*) has not been validated.

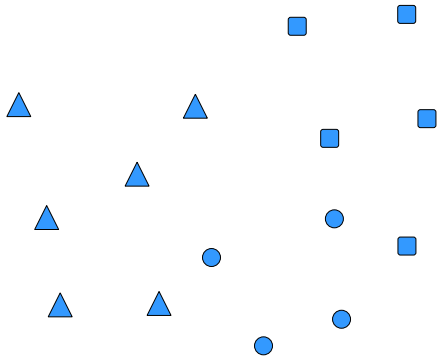
Alternative explanation of observed social influence.

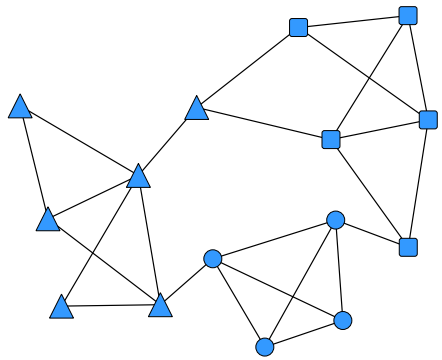


(1) Actors have different characteristics (e. g., age).

Alternative explanation of observed social influence.

(2) Similar actors have a higher probability to become friends.

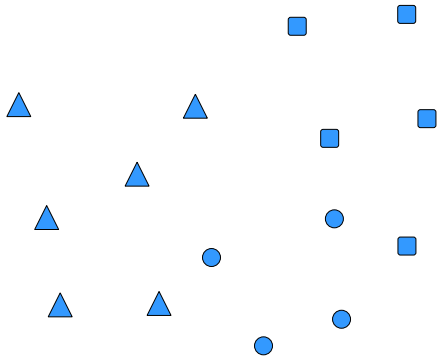


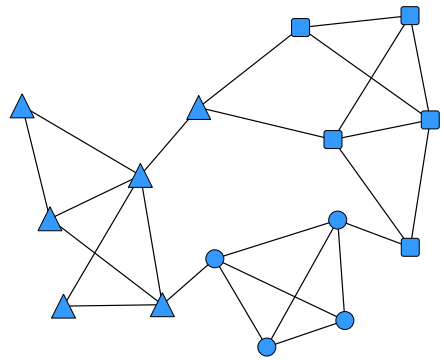


(1) Actors have different characteristics (e. g., age).

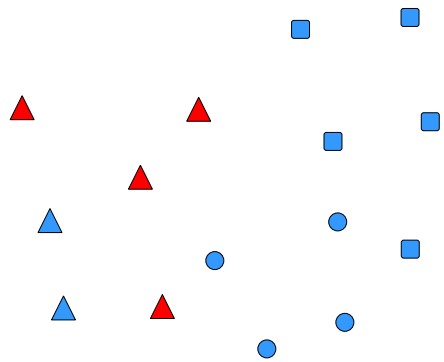
Alternative explanation of observed social influence.

(2) Similar actors have a higher probability to become friends.





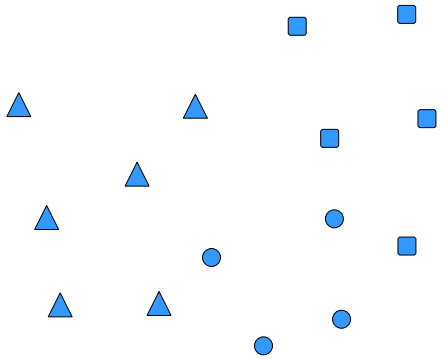
(1) Actors have different characteristics (e. g., age).

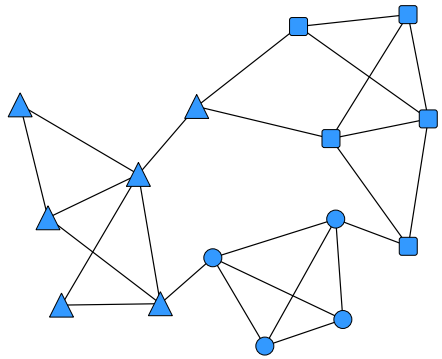


(3) Actors' characteristics influence purchase probabilities.

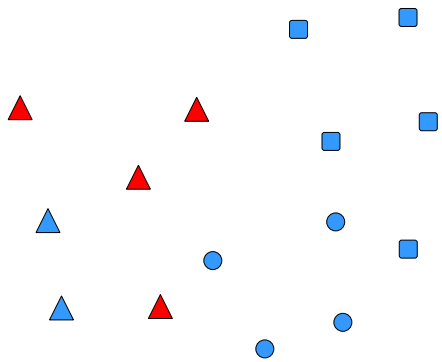
Alternative explanation of observed social influence.

(2) Similar actors have a higher probability to become friends.

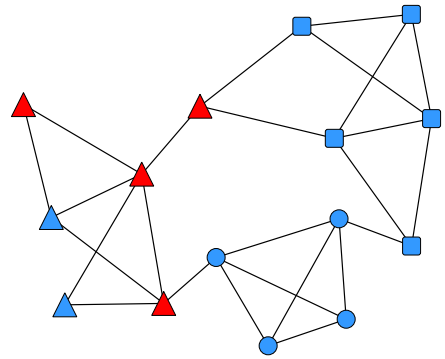




(1) Actors have different characteristics (e. g., age).



(4) Together it looks like social influence.



(3) Actors' characteristics influence purchase probabilities.

Alternative explanations for network effects.

Smokers' friends are often smokers.

- individuals are influenced by their friends;
- OR individuals chose those that are like them as friends.

Chances of becoming obese increase with obese friends.

- individuals are influenced by their friends (e. g., eating behavior or acceptance of obesity);
- OR individuals chose those that are like them as friends;
- OR there are more fast-food restaurants in some regions; these cause obesity; and people living in the same region are more likely to become friends.

Alternative explanations for network effects.

Smokers' friends are often smokers.

individuals are influenced by their friends;

OR individuals chose those that are like them as friends.

Chances of becoming obese increase with obese friends.

individuals are influenced by their friends (e. g., eating behavior or acceptance of obesity);

OR individuals chose those that are like them as friends;

OR there are more fast-food restaurants in some regions; these cause obesity; and people living in the same region are more likely to become friends.

Alternative explanations for network effects.

Smokers' friends are often smokers.

individuals are influenced by their friends;

OR individuals chose those that are like them as friends.

Chances of becoming obese increase with obese friends.

individuals are influenced by their friends (e. g., eating behavior or acceptance of obesity);

OR individuals chose those that are like them as friends;

OR there are more fast-food restaurants in some regions; these cause obesity; and people living in the same region are more likely to become friends.

Alternative explanations for network effects.

Smokers' friends are often smokers.

individuals are influenced by their friends;

OR individuals chose those that are like them as friends.

Chances of becoming obese increase with obese friends.

individuals are influenced by their friends (e. g., eating behavior or acceptance of obesity);

OR individuals chose those that are like them as friends;

OR there are more fast-food restaurants in some regions; these cause obesity; and people living in the same region are more likely to become friends.

Alternative explanations for network effects.

Smokers' friends are often smokers.

individuals are influenced by their friends;

OR individuals chose those that are like them as friends.

Chances of becoming obese increase with obese friends.

individuals are influenced by their friends (e. g., eating behavior or acceptance of obesity);

OR individuals chose those that are like them as friends;

OR there are more fast-food restaurants in some regions; these cause obesity; and people living in the same region are more likely to become friends.

Alternative explanations for network effects.

Smokers' friends are often smokers.

individuals are influenced by their friends;

OR individuals chose those that are like them as friends.

Chances of becoming obese increase with obese friends.

individuals are influenced by their friends (e. g., eating behavior or acceptance of obesity);

OR individuals chose those that are like them as friends;

OR there are more fast-food restaurants in some regions; these cause obesity; and people living in the same region are more likely to become friends.

Alternative explanations for network effects.

Smokers' friends are often smokers.

individuals are influenced by their friends;

OR individuals chose those that are like them as friends.

Chances of becoming obese increase with obese friends.

individuals are influenced by their friends (e. g., eating behavior or acceptance of obesity);

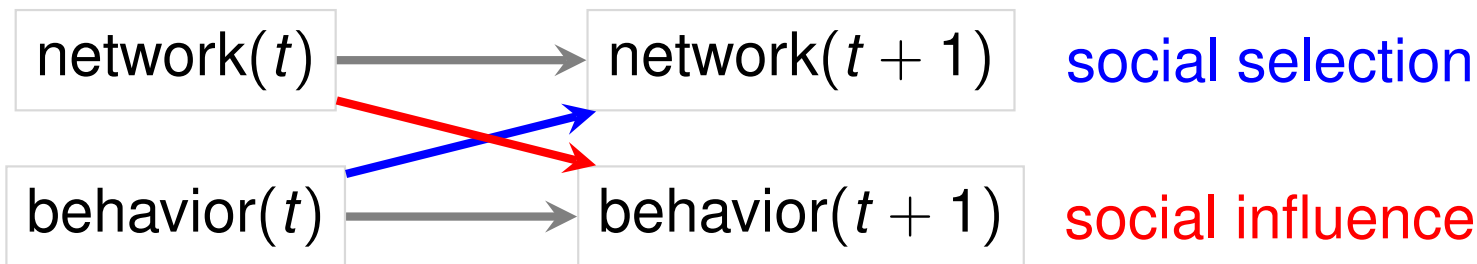
OR individuals chose those that are like them as friends;

OR there are more fast-food restaurants in some regions; these cause obesity; and people living in the same region are more likely to become friends.

Social influence cannot be treated in isolation.

Actors who adopt a behavior might just have a higher probabiliy to be friends of early adopters. (not the other way round)

Need to model the relations as well—not just the behavior.



Ignoring some of these dependencies may lead to spurious conclusions.

Statistical dependencies in network data.

Social influence.

- Network ties influence actors' behavior.
- E. g., friends of smokers start smoking.

Social selection.

- Actor characteristics influence network ties.
- E. g., smokers choose smokers as friends (*homophily*).

Network auto-correlation.

- Ties influence other ties.
- E. g., friends of friends become friends (*transitivity*).

Correlation of individual attributes.

- E. g., eating behavior causes obesity.

Statistical dependencies in network data.

Social influence.

- Network ties influence actors' behavior.
- E. g., friends of smokers start smoking.

Social selection.

- Actor characteristics influence network ties.
- E. g., smokers choose smokers as friends (*homophily*).

Network auto-correlation.

- Ties influence other ties.
- E. g., friends of friends become friends (*transitivity*).

Correlation of individual attributes.

- E. g., eating behavior causes obesity.

Statistical dependencies in network data.

Social influence.

- Network ties influence actors' behavior.
- E. g., friends of smokers start smoking.

Social selection.

- Actor characteristics influence network ties.
- E. g., smokers choose smokers as friends (*homophily*).

Network auto-correlation.

- Ties influence other ties.
- E. g., friends of friends become friends (*transitivity*).

Correlation of individual attributes.

- E. g., eating behavior causes obesity.

Statistical dependencies in network data.

Social influence.

- Network ties influence actors' behavior.
- E. g., friends of smokers start smoking.

Social selection.

- Actor characteristics influence network ties.
- E. g., smokers choose smokers as friends (*homophily*).

Network auto-correlation.

- Ties influence other ties.
- E. g., friends of friends become friends (*transitivity*).

Correlation of individual attributes.

- E. g., eating behavior causes obesity.

Topic of this lecture.

Statistical models for social network data.

Specify realistic **probability distributions** for social networks (ties and behavior), where

- tie probabilities depend on other ties and behavior;
- behavior depends on social ties and behavior of others.

Statistical network models serve several purposes.

Random generation of networks that look like real data

- algorithm engineering; empirical estimation of average runtime or performance;
- simulation of network processes (e. g., information spreding, spread of disease).

Predicting social relations and/or behavior

- learn from given data and predict the data yet to come.

Explaining social relations and/or behavior

- search for rules that govern the evolution of social networks.

Structure of this lecture.

Varying amount of **time information** in the data requires different network models.

Networks observed at a single point in time

- model the probability of single networks $P(G)$.

Continuously observed network changes or events

- model the next network event, given the network of previous events $P(e_t|G_{<t})$.

Networks observed at two or more points in time

- model the conditional probability of later networks, given the previous ones $P(G_t|G_{t-1})$.

Outline.

Background: finite probability space.

Definition

A *finite probability space* is a pair (Ω, P), where

- Ω is a finite set (*possible outcomes*)
- $P\colon \Omega \to [0,1]$ a function satisfying $\sum_{\omega\in\Omega} P(\omega) = 1$.

Notation

- $P(\omega)$ is called the *probability* of $\omega \in \Omega$.
- The probability of a subset $\Omega' \subseteq \Omega$ is defined by $P(\Omega') = \sum_{\omega\in\Omega'} P(\omega)$.

Background: finite probability space.

Definition

A *finite probability space* is a pair (Ω, P), where

- Ω is a finite set (*possible outcomes*)
- $P\colon \Omega \to [0,1]$ a function satisfying $\sum_{\omega\in\Omega} P(\omega) = 1$.

Notation

- $P(\omega)$ is called the *probability* of $\omega \in \Omega$.
- The probability of a subset $\Omega' \subseteq \Omega$ is defined by $P(\Omega') = \sum_{\omega\in\Omega'} P(\omega)$.

Example (dice)

$\Omega = \{1,2,3,4,5,6\}$	(possible outcomes when throwing a die)
$P(\omega) = 1/6$ for all $\omega \in \Omega$	(uniform probability)
$\Omega' = \{1,3,5\}$	(throwing an odd number)

Background: finite probability space.

Definition

A *finite probability space* is a pair (Ω, P), where

- Ω is a finite set (*possible outcomes*)
- $P\colon \Omega \to [0,1]$ a function satisfying $\sum_{\omega\in\Omega} P(\omega) = 1$.

Notation

- $P(\omega)$ is called the *probability* of $\omega \in \Omega$.
- The probability of a subset $\Omega' \subseteq \Omega$ is defined by $P(\Omega') = \sum_{\omega\in\Omega'} P(\omega)$.

Example (lottery)

$\Omega = \{X \subset \{1,\dots,49\}\,;\ |X| = 6\}$ (set of 6 different numbers)

$P(\omega) = \binom{49}{6}^{-1} = \frac{6!43!}{49!}$ for all $\omega \in \Omega$ (uniform probability)

Background: graphs.

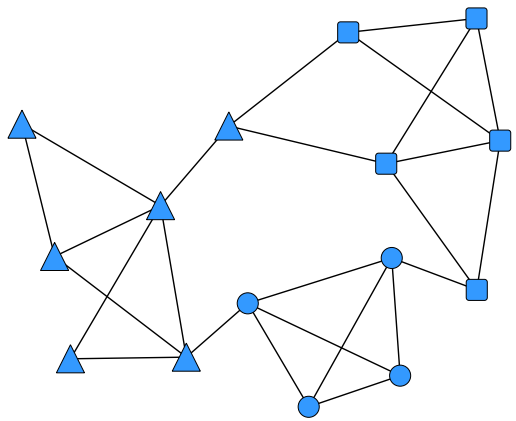
Definition

A *graph* is a pair $G = (V, E)$, where V is a finite set of *vertices* and E the set of *edges*.

- *undirected* graph: $E \subseteq \binom{V}{2} = \{\{u, v\}\,;\ u, v \in V\}$
- *directed* graph: $E \subseteq V \times V = \{(u, v)\,;\ u, v \in V\}$
- *loop*: edge from a vertex to itself

Interpretation:

- vertices correspond to actors
- edges form the relation among them



Random graph models.

Definition

A *random graph model* is a probability space $(\mathcal{G}, P)$, where $\mathcal{G}$ is a (finite) set of graphs.

Example (uniform random graph model)

Let $\mathcal{G}$ be the set of all undirected, loopless graphs with vertex set $V = \{1, \ldots, n\}$ and let

$$P \colon \mathcal{G} \to \mathbb{R};\ P(G) = \frac{1}{2^{\frac{n(n-1)}{2}}} \quad .$$

Then $(\mathcal{G}, P)$ is a random graph model.

Random graph models.

Definition

A *random graph model* is a probability space $(\mathcal{G}, P)$, where $\mathcal{G}$ is a (finite) set of graphs.

Example (uniform random graph model)

Let $\mathcal{G}$ be the set of all undirected, loopless graphs with vertex set $V = \{1, \ldots, n\}$ and let

$$P\colon \mathcal{G} \to \mathbb{R};\ P(G) = \frac{1}{2^{\frac{n(n-1)}{2}}} \quad .$$

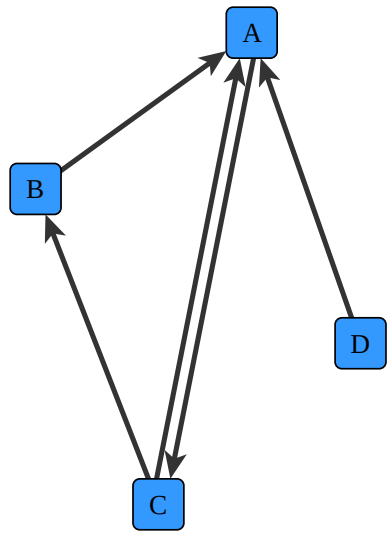
Then $(\mathcal{G}, P)$ is a random graph model.

Background: adjacency matrices.

Definition

A graph $G = (V, E)$ is associated with its *adjacency matrix* X

- vertices index the rows and columns of X;
- $X_{uv} = 1 \Leftrightarrow (u, v) \in E$; $\quad X_{uv} = 0 \Leftrightarrow (u, v) \notin E$.



	A	B	C	D
A	0	0	1	0
B	1	0	0	0
C	1	1	0	0
D	1	0	0	0

A random graph model can be defined via matrices.

Random graph models: notation.

- We consider only random graph models $(\mathcal{G}, P)$ in which all graphs in $\mathcal{G}$ have the same set of vertices; usually $V = \{1, \ldots, n\}$.
- The set of *dyads* D consists of all elements that can be edges in a graph in $\mathcal{G}$.
  - For undirected, loopless graphs $D = \{\{i, j\}\,;\ i, j \in V,\ i \neq j\}$.
  - For directed, loopless graphs $D = \{(i, j)\,;\ i, j \in V,\ i \neq j\}$.

  Correspond to the entries of the adjacency matrix.
- A dyad $e \in D$ is associated with a subset

$$\mathcal{G}_e = \{G \in \mathcal{G}\,;\ e \in E_G\} \ .$$

  When we say *"probability of an edge e"*, we mean $P(\mathcal{G}_e)$, that is, the probability of $\mathcal{G}_e$.

Random graph models: notation.

- We consider only random graph models $(\mathcal{G}, P)$ in which all graphs in $\mathcal{G}$ have the same set of vertices; usually $V = \{1, \ldots, n\}$.
- The set of *dyads* D consists of all elements that can be edges in a graph in $\mathcal{G}$.
  - For undirected, loopless graphs $D = \{\{i, j\}\,;\ i, j \in V,\ i \neq j\}$.
  - For directed, loopless graphs $D = \{(i, j)\,;\ i, j \in V,\ i \neq j\}$.

  Correspond to the entries of the adjacency matrix.
- A dyad $e \in D$ is associated with a subset

$$\mathcal{G}_e = \{G \in \mathcal{G}\,;\ e \in E_G\}\ .$$

When we say *"probability of an edge e"*, we mean $P(\mathcal{G}_e)$, that is, the probability of $\mathcal{G}_e$.

Random graph models: notation.

- We consider only random graph models $(\mathcal{G}, P)$ in which all graphs in $\mathcal{G}$ have the same set of vertices; usually $V = \{1, \ldots, n\}$.
- The set of *dyads* D consists of all elements that can be edges in a graph in $\mathcal{G}$.
  - For undirected, loopless graphs $D = \{\{i, j\}\,;\ i, j \in V,\ i \neq j\}$.
  - For directed, loopless graphs $D = \{(i, j)\,;\ i, j \in V,\ i \neq j\}$.

  Correspond to the entries of the adjacency matrix.
- A dyad $e \in D$ is associated with a subset

$$\mathcal{G}_e = \{G \in \mathcal{G}\,;\ e \in E_G\}\ .$$

When we say *"probability of an edge e"*, we mean $P(\mathcal{G}_e)$, that is, the probability of $\mathcal{G}_e$.

Random graph models: notation.

- We consider only random graph models $(\mathcal{G}, P)$ in which all graphs in $\mathcal{G}$ have the same set of vertices; usually $V = \{1, \ldots, n\}$.
- The set of *dyads* D consists of all elements that can be edges in a graph in $\mathcal{G}$.
  - For undirected, loopless graphs $D = \{\{i, j\}\,;\ i, j \in V,\ i \neq j\}$.
  - For directed, loopless graphs $D = \{(i, j)\,;\ i, j \in V,\ i \neq j\}$.

  Correspond to the entries of the adjacency matrix.
- A dyad $e \in D$ is associated with a subset

$$\mathcal{G}_e = \{G \in \mathcal{G}\,;\ e \in E_G\}\ .$$

When we say *"probability of an edge e"*, we mean $P(\mathcal{G}_e)$, that is, the probability of $\mathcal{G}_e$.

Random graph models: notation.

- We consider only random graph models $(\mathcal{G}, P)$ in which all graphs in $\mathcal{G}$ have the same set of vertices; usually $V = \{1, \ldots, n\}$.
- The set of *dyads* D consists of all elements that can be edges in a graph in $\mathcal{G}$.
  - For undirected, loopless graphs $D = \{\{i, j\}\,;\ i, j \in V,\ i \neq j\}$.
  - For directed, loopless graphs $D = \{(i, j)\,;\ i, j \in V,\ i \neq j\}$.

  Correspond to the entries of the adjacency matrix.
- A dyad $e \in D$ is associated with a subset

$$\mathcal{G}_e = \{G \in \mathcal{G}\,;\ e \in E_G\} \ .$$

When we say *"probability of an edge e"*, we mean $P(\mathcal{G}_e)$, that is, the probability of $\mathcal{G}_e$.

Random graph models: edge probability.

A dyad $e \in D$ is associated with a subset of graphs

$$\mathcal{G}_e = \{G \in \mathcal{G}\,;\ e \in E_G\}\ .$$

When we say *"probability of an edge e"*, we mean $P(\mathcal{G}_e)$.

Thus, assigning a probability to each graph also determines the probability of individual edges.

Does this also hold the other way round?

Random graph models: edge probability.

A dyad $e \in D$ is associated with a subset of graphs

$$\mathcal{G}_e = \{G \in \mathcal{G}\,;\ e \in E_G\}\ .$$

When we say *"probability of an edge e"*, we mean $P(\mathcal{G}_e)$.

Thus, assigning a probability to each graph also determines the probability of individual edges.

Does this also hold the other way round?

Background: independence and conditional prob.

Definition

- Two subsets $A, B \subseteq \Omega$ are *independent* if

$$P(A \cap B) = P(A) \cdot P(B) \ .$$

- If $P(B) > 0$, then the *conditional probability of A, given B* is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \ .$$

- A and A' are *conditionally independent, given B* if

$$P(A \cap A'|B) = P(A|B) \cdot P(A'|B) \ .$$

Background: independence and conditional prob.

Definition

- Two subsets $A, B \subseteq \Omega$ are *independent* if

$$P(A \cap B) = P(A) \cdot P(B) \ .$$

- If $P(B) > 0$, then the *conditional probability of A, given B* is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \ .$$

- A and A' are *conditionally independent, given B* if

$$P(A \cap A'|B) = P(A|B) \cdot P(A'|B) \ .$$

Example (probability space: dice)

$A_{\text{odd}} = \{1, 3, 5\}$ and $A_{\leq 4} = \{1, 2, 3, 4\}$ are independent.

Background: independence and conditional prob.

Definition

- Two subsets $A, B \subseteq \Omega$ are *independent* if

$$P(A \cap B) = P(A) \cdot P(B) \ .$$

- If $P(B) > 0$, then the *conditional probability of A, given B* is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \ .$$

- A and A' are *conditionally independent, given B* if

$$P(A \cap A'|B) = P(A|B) \cdot P(A'|B) \ .$$

Example (probability space: dice)

$A_{\text{odd}} = \{1, 3, 5\}$ and $A_{\leq 3} = \{1, 2, 3\}$ are **not** independent.

Background: independence and conditional prob.

Definition

- Two subsets $A, B \subseteq \Omega$ are *independent* if

$$P(A \cap B) = P(A) \cdot P(B) \ .$$

- If $P(B) > 0$, then the *conditional probability of A, given B* is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \ .$$

- A and A' are *conditionally independent, given B* if

$$P(A \cap A'|B) = P(A|B) \cdot P(A'|B) \ .$$

Background: independence and conditional prob.

Definition

- Two subsets $A, B \subseteq \Omega$ are *independent* if

$$P(A \cap B) = P(A) \cdot P(B) \ .$$

- If $P(B) > 0$, then the *conditional probability of A, given B* is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \ .$$

- A and A' are *conditionally independent, given B* if

$$P(A \cap A'|B) = P(A|B) \cdot P(A'|B) \ .$$

Example (probability space: dice)

$P(A_{\text{odd}}|A_{\leq 4}) = 1/2$, but $P(A_{\text{odd}}|A_{\leq 3}) = 2/3$

Background: independence and conditional prob.

Definition

- Two subsets $A, B \subseteq \Omega$ are *independent* if

$$P(A \cap B) = P(A) \cdot P(B) \ .$$

- If $P(B) > 0$, then the *conditional probability of A, given B* is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \ .$$

- A and A' are *conditionally independent, given B* if

$$P(A \cap A'|B) = P(A|B) \cdot P(A'|B) \ .$$

Random graph models: notation (continued).

A dyad $e \in D$ is associated with a subset of graphs

$$\mathcal{G}_e = \{G \in \mathcal{G}\,;\ e \in E_G\}\ .$$

- if $\mathcal{G}_{e_1}$ and $\mathcal{G}_{e_2}$ are independent, we say that "the dyads e_1 and e_2 are independent"

Definition

Let $D' \subset D$. A dyad $e \in D \setminus D'$ is said to be *independent* of D' if for all partitions $D' = D^+ \cup D^-$, the subset $\mathcal{G}_e$ is independent of

$$\mathcal{G}_{D^+ \cup D^-} = \{G \in \mathcal{G}\,;\ D^+ \subseteq E_G \text{ and } D^- \cap E_G = \emptyset\}$$

(*all dyads in D^+ are edges in G and no dyad in D^- is an edge*).

Random graph models: notation (continued).

A dyad $e \in D$ is associated with a subset of graphs

$$\mathcal{G}_e = \{G \in \mathcal{G}\,;\ e \in E_G\}\ .$$

- if $\mathcal{G}_{e_1}$ and $\mathcal{G}_{e_2}$ are independent, we say that "the dyads e_1 and e_2 are independent"

Definition

Let $D' \subset D$. A dyad $e \in D \setminus D'$ is said to be *independent* of D' if for all partitions $D' = D^+ \cup D^-$, the subset $\mathcal{G}_e$ is independent of

$$\mathcal{G}_{D^+ \cup D^-} = \{G \in \mathcal{G}\,;\ D^+ \subseteq E_G \text{ and } D^- \cap E_G = \emptyset\}$$

(*all dyads in* D^+ *are edges in* G *and no dyad in* D^- *is an edge*).

Random graph models: notation (continued).

A dyad $e \in D$ is associated with a subset of graphs

$$\mathcal{G}_e = \{G \in \mathcal{G}\,;\ e \in E_G\}\ .$$

- if $\mathcal{G}_{e_1}$ and $\mathcal{G}_{e_2}$ are independent, we say that "the dyads e_1 and e_2 are independent"

Definition

Let $D' \subset D$. A dyad $e \in D \setminus D'$ is said to be *independent* of D' if for all partitions $D' = D^+ \cup D^-$, the subset $\mathcal{G}_e$ is independent of

$$\mathcal{G}_{D^+ \cup D^-} = \{G \in \mathcal{G}\,;\ D^+ \subseteq E_G \text{ and } D^- \cap E_G = \emptyset\}$$

(*all dyads in* D^+ *are edges in* G *and no dyad in* D^- *is an edge*).

Structural balance theory (illustrating dependence).

Structural balance theory (Heider 1946) applies to triplets of 3 actors with **positive** or **negative** ties:

balanced **not balanced**

SBT claims that actors prefer balanced networks.

In an appropriate random graph model, it holds that

- all edges are pairwise independent;
- every edge depends on the two others.

Structural balance theory (illustrating dependence).

Structural balance theory (Heider 1946) applies to triplets of 3 actors with **positive** or **negative** ties:

balanced **not balanced**

SBT claims that actors prefer balanced networks.

In an appropriate random graph model, it holds that

- all edges are pairwise independent;
- every edge depends on the two others.

Before ending this section, we treat

- edge probability,
- independence of dyads,
- and expected number of edges

of the uniform random graph model.

Uniform graph model: edge probability.

Lemma

The edge probability of a dyad $e \in D$ in the uniform random graph model is equal to $1/2$.

Proof.

The two sets

$$\begin{aligned} \mathcal{G}_e &= \{G \in \mathcal{G};\ e \in E_G\}, \\ \overline{\mathcal{G}_e} &= \{G \in \mathcal{G};\ e \notin E_G\} \end{aligned}$$

- have the same cardinality $\Rightarrow P(\mathcal{G}_e) = P(\overline{\mathcal{G}_e})$,
- are disjoint $\Rightarrow P(\mathcal{G}_e) + P(\overline{\mathcal{G}_e}) = P(\mathcal{G}_e \cup \overline{\mathcal{G}_e})$,
- and their union equals $\mathcal{G} \Rightarrow P(\mathcal{G}_e \cup \overline{\mathcal{G}_e}) = 1$.

$\Rightarrow\ P(\mathcal{G}_e) = 1/2$.

Uniform graph model: edge probability.

Lemma

The edge probability of a dyad $e \in D$ in the uniform random graph model is equal to $1/2$.

Proof.

The two sets

$$\begin{aligned} \mathcal{G}_e &= \{G \in \mathcal{G};\ e \in E_G\}, \\ \overline{\mathcal{G}_e} &= \{G \in \mathcal{G};\ e \notin E_G\} \end{aligned}$$

- have the same cardinality $\Rightarrow P(\mathcal{G}_e) = P(\overline{\mathcal{G}_e})$,
- are disjoint $\Rightarrow P(\mathcal{G}_e) + P(\overline{\mathcal{G}_e}) = P(\mathcal{G}_e \cup \overline{\mathcal{G}_e})$,
- and their union equals $\mathcal{G} \Rightarrow P(\mathcal{G}_e \cup \overline{\mathcal{G}_e}) = 1$.

$\Rightarrow\ P(\mathcal{G}_e) = 1/2$.

Uniform graph model: edge probability.

Lemma

The edge probability of a dyad $e \in D$ in the uniform random graph model is equal to $1/2$.

Proof.

The two sets

$$\begin{aligned} \mathcal{G}_e &= \{G \in \mathcal{G}\,;\ e \in E_G\}, \\ \overline{\mathcal{G}_e} &= \{G \in \mathcal{G}\,;\ e \notin E_G\} \end{aligned}$$

- have the same cardinality $\Rightarrow P(\mathcal{G}_e) = P(\overline{\mathcal{G}_e})$,
- are disjoint $\Rightarrow P(\mathcal{G}_e) + P(\overline{\mathcal{G}_e}) = P(\mathcal{G}_e \cup \overline{\mathcal{G}_e})$,
- and their union equals $\mathcal{G} \Rightarrow P(\mathcal{G}_e \cup \overline{\mathcal{G}_e}) = 1$.

$\Rightarrow P(\mathcal{G}_e) = 1/2$.

Uniform graph model: edge probability.

Lemma

The edge probability of a dyad $e \in D$ in the uniform random graph model is equal to $1/2$.

Proof.

The two sets

$$\begin{aligned} \mathcal{G}_e &= \{G \in \mathcal{G}\,;\ e \in E_G\}, \\ \overline{\mathcal{G}_e} &= \{G \in \mathcal{G}\,;\ e \notin E_G\} \end{aligned}$$

- have the same cardinality $\Rightarrow P(\mathcal{G}_e) = P(\overline{\mathcal{G}_e})$,
- are disjoint $\Rightarrow P(\mathcal{G}_e) + P(\overline{\mathcal{G}_e}) = P(\mathcal{G}_e \cup \overline{\mathcal{G}_e})$,
- and their union equals $\mathcal{G} \Rightarrow P(\mathcal{G}_e \cup \overline{\mathcal{G}_e}) = 1$.

$\Rightarrow\ P(\mathcal{G}_e) = 1/2$.

Uniform graph model: edge probability.

Lemma

The edge probability of a dyad $e \in D$ in the uniform random graph model is equal to $1/2$.

Proof.

The two sets

$$\begin{aligned} \mathcal{G}_e &= \{G \in \mathcal{G};\ e \in E_G\}, \\ \overline{\mathcal{G}_e} &= \{G \in \mathcal{G};\ e \notin E_G\} \end{aligned}$$

- have the same cardinality $\Rightarrow P(\mathcal{G}_e) = P(\overline{\mathcal{G}_e})$,
- are disjoint $\Rightarrow P(\mathcal{G}_e) + P(\overline{\mathcal{G}_e}) = P(\mathcal{G}_e \cup \overline{\mathcal{G}_e})$,
- and their union equals $\mathcal{G} \Rightarrow P(\mathcal{G}_e \cup \overline{\mathcal{G}_e}) = 1$.

$\Rightarrow P(\mathcal{G}_e) = 1/2$.

Uniform graph model: edge probability.

Lemma

The edge probability of a dyad $e \in D$ in the uniform random graph model is equal to $1/2$.

Proof.

The two sets

$$\begin{aligned} \mathcal{G}_e &= \{G \in \mathcal{G}\,;\ e \in E_G\}, \\ \overline{\mathcal{G}_e} &= \{G \in \mathcal{G}\,;\ e \notin E_G\} \end{aligned}$$

- have the same cardinality $\Rightarrow P(\mathcal{G}_e) = P(\overline{\mathcal{G}_e})$,
- are disjoint $\Rightarrow P(\mathcal{G}_e) + P(\overline{\mathcal{G}_e}) = P(\mathcal{G}_e \cup \overline{\mathcal{G}_e})$,
- and their union equals $\mathcal{G} \Rightarrow P(\mathcal{G}_e \cup \overline{\mathcal{G}_e}) = 1$.

$\Rightarrow\ P(\mathcal{G}_e) = 1/2$.

Uniform graph model: edge probability.

Lemma

The edge probability of a dyad $e \in D$ in the uniform random graph model is equal to $1/2$.

Proof.

The two sets

$$\begin{aligned} \mathcal{G}_e &= \{G \in \mathcal{G}\,;\ e \in E_G\}, \\ \overline{\mathcal{G}_e} &= \{G \in \mathcal{G}\,;\ e \notin E_G\} \end{aligned}$$

- have the same cardinality $\Rightarrow P(\mathcal{G}_e) = P(\overline{\mathcal{G}_e})$,
- are disjoint $\Rightarrow P(\mathcal{G}_e) + P(\overline{\mathcal{G}_e}) = P(\mathcal{G}_e \cup \overline{\mathcal{G}_e})$,
- and their union equals $\mathcal{G}$ $\Rightarrow P(\mathcal{G}_e \cup \overline{\mathcal{G}_e}) = 1$.

$\Rightarrow P(\mathcal{G}_e) = 1/2$.

Uniform graph model: edge probability.

Lemma

The edge probability of a dyad $e \in D$ in the uniform random graph model is equal to $1/2$.

Proof.

The two sets

$$\begin{aligned} \mathcal{G}_e &= \{G \in \mathcal{G};\ e \in E_G\}, \\ \overline{\mathcal{G}_e} &= \{G \in \mathcal{G};\ e \notin E_G\} \end{aligned}$$

- have the same cardinality $\Rightarrow P(\mathcal{G}_e) = P(\overline{\mathcal{G}_e})$,
- are disjoint $\Rightarrow P(\mathcal{G}_e) + P(\overline{\mathcal{G}_e}) = P(\mathcal{G}_e \cup \overline{\mathcal{G}_e})$,
- and their union equals $\mathcal{G} \Rightarrow P(\mathcal{G}_e \cup \overline{\mathcal{G}_e}) = 1$.

$\Rightarrow P(\mathcal{G}_e) = 1/2$.

Uniform graph model: edge probability.

Lemma

The edge probability of a dyad $e \in D$ in the uniform random graph model is equal to $1/2$.

Proof.

The two sets

$$\begin{aligned} \mathcal{G}_e &= \{G \in \mathcal{G}\,;\ e \in E_G\}, \\ \overline{\mathcal{G}_e} &= \{G \in \mathcal{G}\,;\ e \notin E_G\} \end{aligned}$$

- have the same cardinality $\Rightarrow P(\mathcal{G}_e) = P(\overline{\mathcal{G}_e})$,
- are disjoint $\Rightarrow P(\mathcal{G}_e) + P(\overline{\mathcal{G}_e}) = P(\mathcal{G}_e \cup \overline{\mathcal{G}_e})$,
- and their union equals $\mathcal{G} \Rightarrow P(\mathcal{G}_e \cup \overline{\mathcal{G}_e}) = 1$.

$\Rightarrow\ P(\mathcal{G}_e) = 1/2$.

Uniform graph model: edge probability.

Lemma

*The edge probability of a dyad $e \in D$ in the uniform random graph model is $1/2$, **independent of all sets of dyads.***

Proof.

Let $D^+, D^- \subseteq D \setminus \{e\}$ be two disjoint subsets of dyads, not containing e. Consider

$$\mathcal{G}' = \{G \in \mathcal{G}\,;\; D^+ \subseteq E_G, \text{ and } D^- \cap E_G = \emptyset\}$$

(*all dyads in D^+ are edges in G and no dyad in D^- is an edge*).

Then, with $\mathcal{G}'_e = \{G \in \mathcal{G}'\,;\; e \in E_G\}$ it follows $P(\mathcal{G}'_e|\mathcal{G}') = 1/2$ (as on the previous slide). □

Uniform graph model: edge probability.

Lemma

The edge probability of a dyad $e \in D$ in the uniform random graph model is $1/2$, ***independent of all sets of dyads.***

Proof.

Let $D^+, D^- \subseteq D \setminus \{e\}$ be two disjoint subsets of dyads, not containing e. Consider

$$\mathcal{G}' = \{G \in \mathcal{G}\,;\; D^+ \subseteq E_G, \text{ and } D^- \cap E_G = \emptyset\}$$

(*all dyads in* D^+ *are edges in* G *and no dyad in* D^- *is an edge*).

Then, with $\mathcal{G}'_e = \{G \in \mathcal{G}'\,;\; e \in E_G\}$ it follows $P(\mathcal{G}'_e|\mathcal{G}') = 1/2$ (as on the previous slide). □

Uniform graph model: edge probability.

Lemma

The edge probability of a dyad $e \in D$ in the uniform random graph model is $1/2$*,* ***independent of all sets of dyads.***

Proof.

Let $D^+, D^- \subseteq D \setminus \{e\}$ be two disjoint subsets of dyads, not containing e. Consider

$$\mathcal{G}' = \{G \in \mathcal{G}\,;\ D^+ \subseteq E_G, \text{ and } D^- \cap E_G = \emptyset\}$$

(*all dyads in* D^+ *are edges in* G *and no dyad in* D^- *is an edge*).

Then, with $\mathcal{G}'_e = \{G \in \mathcal{G}'\,;\ e \in E_G\}$ it follows $P(\mathcal{G}'_e|\mathcal{G}') = 1/2$ (as on the previous slide). □

Uniform graph model: edge probability.

Lemma

The edge probability of a dyad $e \in D$ in the uniform random graph model is $1/2$*,* ***independent of all sets of dyads.***

Proof.

Let $D^+, D^- \subseteq D \setminus \{e\}$ be two disjoint subsets of dyads, not containing e. Consider

$$\mathcal{G}' = \{G \in \mathcal{G}\,;\ D^+ \subseteq E_G, \text{ and } D^- \cap E_G = \emptyset\}$$

(*all dyads in* D^+ *are edges in* G *and no dyad in* D^- *is an edge*).

Then, with $\mathcal{G}'_e = \{G \in \mathcal{G}'\,;\ e \in E_G\}$ it follows $P(\mathcal{G}'_e|\mathcal{G}') = 1/2$ (as on the previous slide). □

Background: random variable and expectation.

Let (Ω, P) be a probability space.

Definition

A *random variable* is a function $X : \Omega \to \mathbb{R}$.

Let $S = X(\Omega)$ be the set of values of X.

The *expectation* of the random variable X is defined by

$$\mathbb{E}(X) = \sum_{x \in S} x \cdot P(X = x) \ .$$

Example

The prize assigned to lottery numbers is a random variable.
Its expectation is the average gain that could be expected after "many" lottery draws (to be compared with the cost of a ticket).

Background: random variable and expectation.

Let (Ω, P) be a probability space.

Definition

A *random variable* is a function $X : \Omega \to \mathbb{R}$.

Let $S = X(\Omega)$ be the set of values of X.

The *expectation* of the random variable X is defined by

$$\mathbb{E}(X) = \sum_{x \in S} x \cdot P(X = x) \ .$$

Example

The prize assigned to lottery numbers is a random variable.
Its expectation is the average gain that could be expected after "many" lottery draws (to be compared with the cost of a ticket).

Background: linearity of expectation.

Equivalently, the *expectation* of a random variable $X : \Omega \to \mathbb{R}$ is

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} P(\omega) \cdot X(\omega) \ .$$

Lemma

If $X, Y : \Omega \to \mathbb{R}$ are two random variables and α a real number, then it is

$$\begin{aligned} \mathbb{E}(X + Y) &= \mathbb{E}(X) + \mathbb{E}(Y) \\ \mathbb{E}(\alpha \cdot X) &= \alpha \cdot \mathbb{E}(X) \ . \end{aligned}$$

□

Background: linearity of expectation.

Equivalently, the *expectation* of a random variable $X : \Omega \to \mathbb{R}$ is

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} P(\omega) \cdot X(\omega) \ .$$

Lemma

If $X, Y : \Omega \to \mathbb{R}$ are two random variables and α a real number, then it is

$$\begin{aligned} \mathbb{E}(X + Y) &= \mathbb{E}(X) + \mathbb{E}(Y) \\ \mathbb{E}(\alpha \cdot X) &= \alpha \cdot \mathbb{E}(X) \ . \end{aligned}$$

Background: density.

The *density* of a graph is the ratio

$$\frac{\text{number of edges}}{\text{number of dyads}}.$$

The density is between zero and one.

For undirected, loopless graphs with n vertices the denominator is equal to $n(n-1)/2$.
For directed, loopless graphs with n vertices the denominator is equal to $n(n-1)$.

Background: density.

The *density* of a graph is the ratio

$$\frac{\text{number of edges}}{\text{number of dyads}} .$$

The density is between zero and one.

For undirected, loopless graphs with n vertices the denominator is equal to $n(n-1)/2$.
For directed, loopless graphs with n vertices the denominator is equal to $n(n-1)$.

Uniform graph model: expected density.

Lemma

The expected density of graphs in $\mathcal{G}(n)$ equals $1/2$.

Proof.

The number of edges of a graph G can be written as

$$m(G) = \sum_{e \in D} \chi_e(G)$$

Uniform graph model: expected density.

Lemma

The expected density of graphs in $\mathcal{G}(n)$ equals $1/2$.

Proof.

The number of edges of a graph G can be written as

$$m(G) = \sum_{e \in D} \chi_e(G)$$

where $\chi_e \colon \mathcal{G} \to \{0, 1\}$ is defined by

$$\chi_e(G) = \begin{cases} 1 & \text{if } e \in E_G \\ 0 & \text{else.} \end{cases}$$

Uniform graph model: expected density.

Lemma

The expected density of graphs in $\mathcal{G}(n)$ equals $1/2$.

Proof.

The number of edges of a graph G can be written as

$$m(G) = \sum_{e \in D} \chi_e(G)$$

Uniform graph model: expected density.

Lemma

The expected density of graphs in $\mathcal{G}(n)$ equals $1/2$.

Proof.

The number of edges of a graph G can be written as

$$m(G) = \sum_{e \in D} \chi_e(G)$$

From the linearity of the expectation it follows that

$$\begin{aligned} \mathbb{E}[m] &= \sum_{e \in D} \mathbb{E}[\chi_e] = \sum_{e \in D} P(e) \cdot 1 + (1 - P(e)) \cdot 0 \\ &= \sum_{e \in D} \frac{1}{2} \cdot 1 = \frac{1}{2} \frac{n(n-1)}{2} \end{aligned}$$

Uniform graph model: summary.

Characterizing properties:

- edges are mutually independent;
- all edges are equally likely;
- no preference for edges over non-edges or vice versa.

It has been found that empirical networks typically violate all of these properties:

- edges are not independent;
- have varying probabilities;
- networks are typically sparse (i. e., most dyads are non-edges).

Uniform graph model: summary.

Characterizing properties:

- edges are mutually independent;
- all edges are equally likely;
- no preference for edges over non-edges or vice versa.

It has been found that empirical networks typically violate all of these properties:

- edges are not independent;
- have varying probabilities;
- networks are typically sparse (i. e., most dyads are non-edges).

Outline.

$\mathcal{G}(n, p)$ definition.

$\mathcal{G}(n, p)$ is a model for undirected, loopless graphs.

Two parameters
$n \in \mathbb{N}_{\geq 1}$ (number of vertices)
$p \in [0, 1]$ (edge probability)

Definition of probability $P \colon \mathcal{G} \to [0, 1]$

Probability of graphs defined by specifying

- edge probability of each dyad is equal to p,
- each dyad is independent of all sets of dyads.

$\mathcal{G}(n, p)$ definition.

$\mathcal{G}(n, p)$ is a model for undirected, loopless graphs.

Two parameters
$n \in \mathbb{N}_{\geq 1}$ (number of vertices)
$p \in [0, 1]$ (edge probability)

Definition of probability $P \colon \mathcal{G} \to [0, 1]$

Probability of graphs defined by specifying

- edge probability of each dyad is equal to p,
- each dyad is independent of all sets of dyads.

$\mathcal{G}(n, p)$ definition.

$\mathcal{G}(n, p)$ is a model for undirected, loopless graphs.

Two parameters
$n \in \mathbb{N}_{\geq 1}$ (number of vertices)
$p \in [0, 1]$ (edge probability)

Definition of probability $P \colon \mathcal{G} \to [0, 1]$

Probability of graphs defined by specifying

- edge probability of each dyad is equal to p,
- each dyad is independent of all sets of dyads.

$\mathcal{G}(n,p)$ probability of a graph.

Lemma

The probability of a graph G with m edges in $\mathcal{G}(n,p)$ is

$$P(G) = p^m(1-p)^{\frac{n(n-1)}{2}-m}$$

Proof.

For dyad $e \in D$ it is $P(e \in E) = p$, $P(e \notin E) = 1-p$; multiply these probabilities over all dyads . . . done. (multiplication is valid since edge probabilities are independent) □

Remark

The uniform random graph model is identical with $\mathcal{G}(n, \frac{1}{2})$.

$\mathcal{G}(n,p)$ probability of a graph.

Lemma

The probability of a graph G with m edges in $\mathcal{G}(n,p)$ is

$$P(G) = p^m(1-p)^{\frac{n(n-1)}{2}-m}$$

Proof.

For dyad $e \in D$ it is $P(e \in E) = p$, $P(e \notin E) = 1-p$; multiply these probabilities over all dyads . . . done. (multiplication is valid since edge probabilities are independent) □

Remark

The uniform random graph model is identical with $\mathcal{G}(n, \frac{1}{2})$.

$\mathcal{G}(n,p)$ probability of a graph.

Lemma

The probability of a graph G with m edges in $\mathcal{G}(n,p)$ is

$$P(G) = p^m(1-p)^{\frac{n(n-1)}{2}-m}$$

Proof.

For dyad $e \in D$ it is $P(e \in E) = p$, $P(e \notin E) = 1-p$;
multiply these probabilities over all dyads . . . done.
(multiplication is valid since edge probabilities are independent) □

Remark

The uniform random graph model is identical with $\mathcal{G}(n, \frac{1}{2})$.

$\mathcal{G}(n,p)$ probability of a graph.

Lemma

The probability of a graph G with m edges in $\mathcal{G}(n,p)$ is

$$P(G) = p^m(1-p)^{\frac{n(n-1)}{2}-m}$$

Proof.

For dyad $e \in D$ it is $P(e \in E) = p$, $P(e \notin E) = 1-p$; multiply these probabilities over all dyads . . . done.
(multiplication is valid since edge probabilities are independent) □

Remark

The uniform random graph model is identical with $\mathcal{G}(n, \frac{1}{2})$.

$\mathcal{G}(n,p)$ probability of a graph.

Lemma

The probability of a graph G with m edges in $\mathcal{G}(n,p)$ is

$$P(G) = p^m(1-p)^{\frac{n(n-1)}{2}-m}$$

Proof.

For dyad $e \in D$ it is $P(e \in E) = p$, $P(e \notin E) = 1-p$; multiply these probabilities over all dyads . . . done. (multiplication is valid since edge probabilities are independent) □

Remark

The uniform random graph model is identical with $\mathcal{G}(n, \frac{1}{2})$.

$\mathcal{G}(n,p)$ probability of a graph.

Lemma

The probability of a graph G with m edges in $\mathcal{G}(n,p)$ is

$$P(G) = p^m(1-p)^{\frac{n(n-1)}{2}-m}$$

Proof.

For dyad $e \in D$ it is $P(e \in E) = p$, $P(e \notin E) = 1-p$; multiply these probabilities over all dyads . . . done. (multiplication is valid since edge probabilities are independent) □

Remark

The uniform random graph model is identical with $\mathcal{G}(n, \frac{1}{2})$.

$\mathcal{G}(n,p)$ probability of a graph.

$$P(G) = p^m (1-p)^{\frac{n(n-1)}{2} - m}$$

Proof.

(extended version)

$$\begin{aligned}
P(G) &= P(\{G\}) \\
&= P\left(\bigcap_{d \in E_G} \mathcal{G}_d \cap \bigcap_{d \in D \setminus E_G} \overline{\mathcal{G}_d} \right) \\
&= \prod_{d \in E_G} P(\mathcal{G}_d) \cdot \prod_{d \in D \setminus E_G} P(\overline{\mathcal{G}_d}) \\
&= \prod_{d \in E_G} p \cdot \prod_{d \in D \setminus E_G} 1 - p \\
&= p^m (1-p)^{\frac{n(n-1)}{2} - m} \ .
\end{aligned}$$

Remark.

If every edge is independent of all sets of dyads, then the probability of each graph is determined by the edge probabilities of all dyads.

In general (without independence), the edge probabilities do not uniquely determine the graph probability.

Remark.

If every edge is independent of all sets of dyads, then the probability of each graph is determined by the edge probabilities of all dyads.

In general (without independence), the edge probabilities do not uniquely determine the graph probability.

$\mathcal{G}(n,p)$ expected density.

Lemma

The expected density of graphs in $\mathcal{G}(n,p)$ equals p.

Proof.

The number of edges of a graph G can be written as

$$m(G) = \sum_{e \in D} \chi_e(G)$$

$\mathcal{G}(n,p)$ expected density.

Lemma

The expected density of graphs in $\mathcal{G}(n,p)$ equals p.

Proof.

The number of edges of a graph G can be written as

$$m(G) = \sum_{e \in D} \chi_e(G)$$

where $\chi_e \colon \mathcal{G} \to \{0,1\}$ is defined by

$$\chi_e(G) = \begin{cases} 1 & \text{if } e \in E_G \\ 0 & \text{else.} \end{cases}$$

$\mathcal{G}(n,p)$ expected density.

Lemma

The expected density of graphs in $\mathcal{G}(n,p)$ equals p.

Proof.

The number of edges of a graph G can be written as

$$m(G) = \sum_{e \in D} \chi_e(G)$$

$\mathcal{G}(n,p)$ expected density.

Lemma

The expected density of graphs in $\mathcal{G}(n,p)$ equals p.

Proof.

The number of edges of a graph G can be written as

$$m(G) = \sum_{e \in D} \chi_e(G)$$

From the linearity of the expectation it follows that

$$\begin{aligned} \mathbb{E}[m] &= \sum_{e \in D} \mathbb{E}[\chi_e] = \sum_{e \in D} P(e) \cdot 1 + (1 - P(e)) \cdot 0 \\ &= p\frac{n(n-1)}{2} \end{aligned}$$

$\mathcal{G}(n, p)$ efficient generation.

Task: design of a probabilistic algorithm returning a graph G with probability as in $\mathcal{G}(n, p)$.

Naive algorithm: iterate over all dyads $e \in D$

- draw a uniformly distributed random number $r \in [0, 1]$;
- if $r \leq p$ add e to the edge set.

$\mathcal{G}(n, p)$ efficient generation.

Task: design of a probabilistic algorithm returning a graph G with probability as in $\mathcal{G}(n, p)$.

Naive algorithm: iterate over all dyads $e \in D$

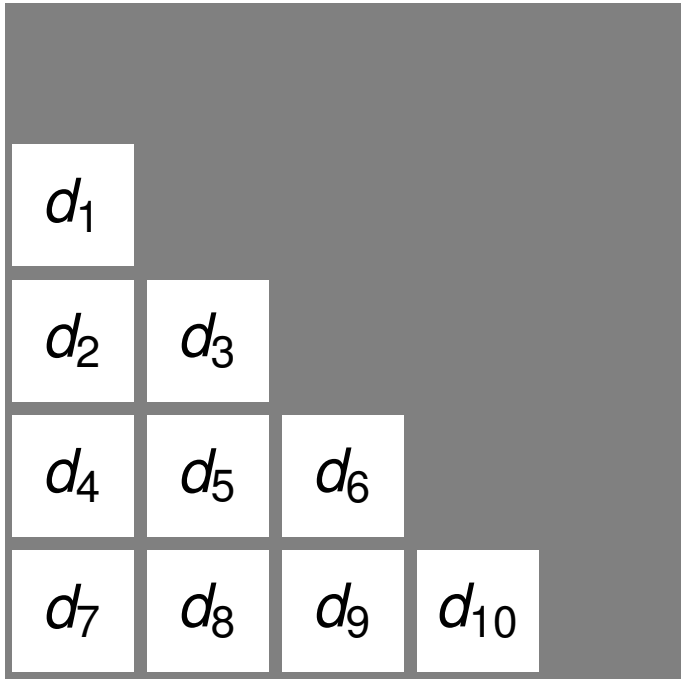
- draw a uniformly distributed random number $r \in [0, 1]$;
- if $r \leq p$ add e to the edge set.

Generating graphs from $\mathcal{G}(n, p)$ (naive algorithm).

Naive algorithm: iterate over all dyads $e \in D$

- draw a uniformly distributed random number $r \in [0, 1]$;
- if $r \leq p$ add e to the edge set.

enumerate dyads



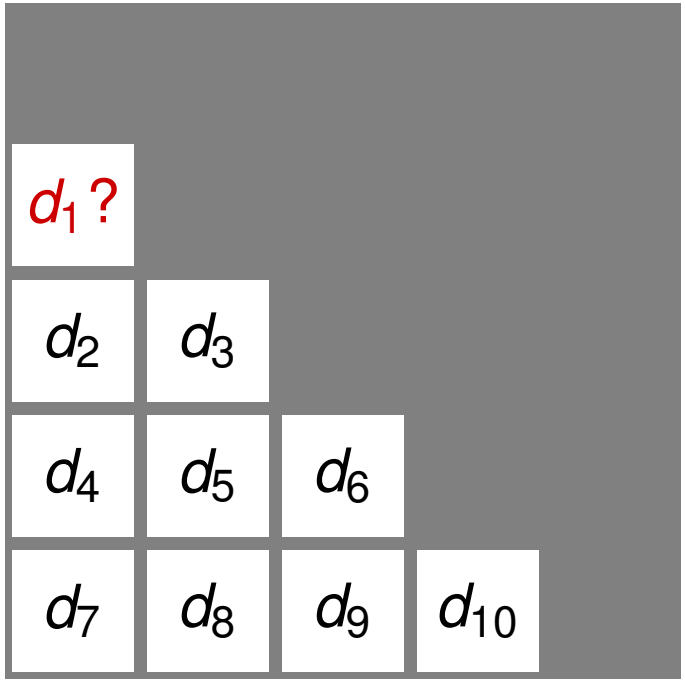
inefficient, when p is small (too many NOs)

Generating graphs from $\mathcal{G}(n, p)$ (naive algorithm).

Naive algorithm: iterate over all dyads $e \in D$

- draw a uniformly distributed random number $r \in [0, 1]$;
- if $r \leq p$ add e to the edge set.

Is d_1 an edge?
(draw a random number...)



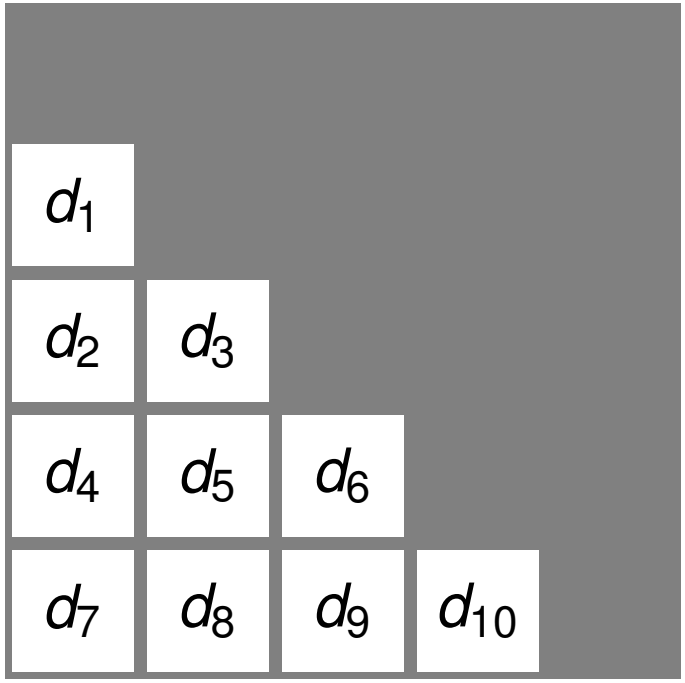
inefficient, when p is small (too many NOs)

Generating graphs from $\mathcal{G}(n, p)$ (naive algorithm).

Naive algorithm: iterate over all dyads $e \in D$

- draw a uniformly distributed random number $r \in [0, 1]$;
- if $r \leq p$ add e to the edge set.

Is d_1 an edge?
$\rightarrow$ NO (for instance)



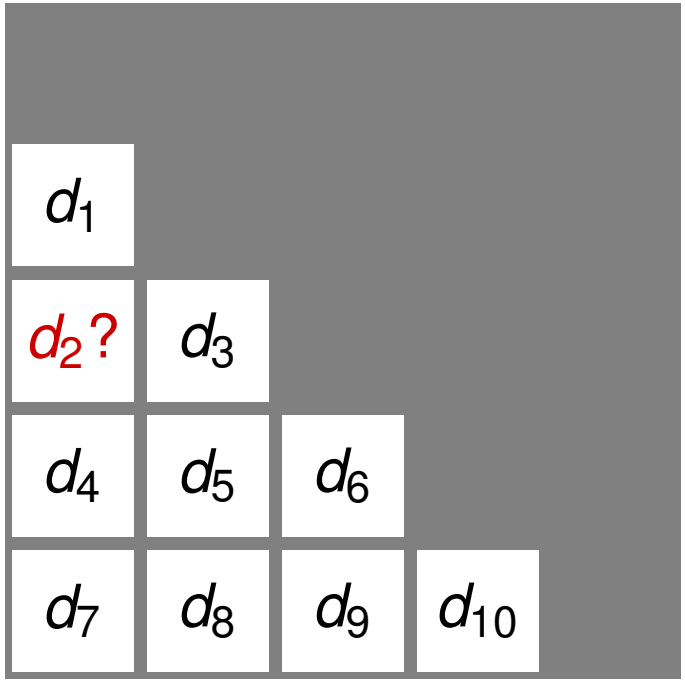
inefficient, when p is small (too many NOs)

Generating graphs from $\mathcal{G}(n, p)$ (naive algorithm).

Naive algorithm: iterate over all dyads $e \in D$

- draw a uniformly distributed random number $r \in [0, 1]$;
- if $r \leq p$ add e to the edge set.

Is d_2 an edge?
(draw a random number. . .)



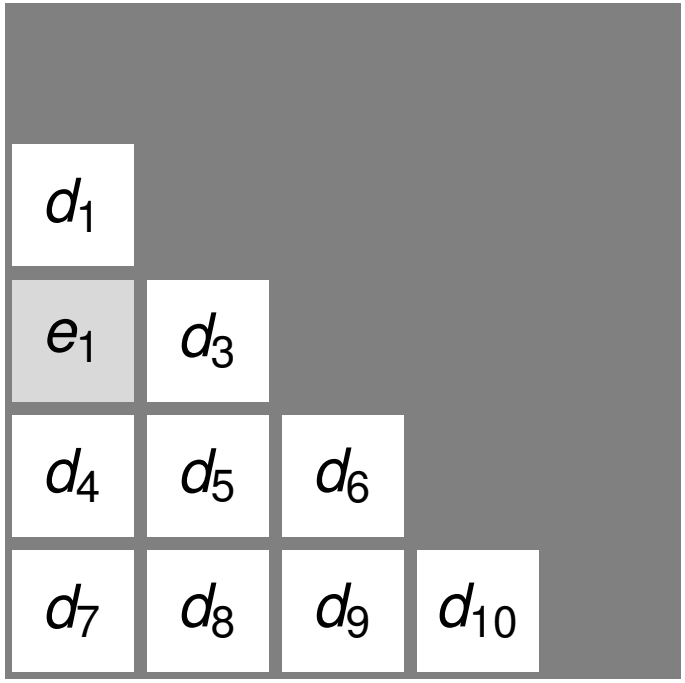
inefficient, when p is small (too many NOs)

Generating graphs from $\mathcal{G}(n, p)$ (naive algorithm).

Naive algorithm: iterate over all dyads $e \in D$

- draw a uniformly distributed random number $r \in [0, 1]$;
- if $r \leq p$ add e to the edge set.

Is d_2 an edge?
$\rightarrow$ YES (for instance)
$\Rightarrow$ turn d_2 into the first edge



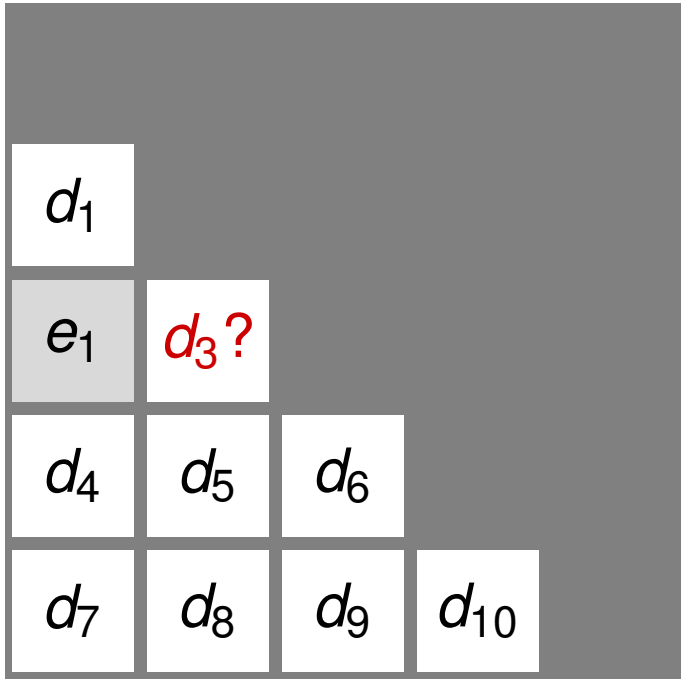
inefficient, when p is small (too many NOs)

Generating graphs from $\mathcal{G}(n, p)$ (naive algorithm).

Naive algorithm: iterate over all dyads $e \in D$

- draw a uniformly distributed random number $r \in [0, 1]$;
- if $r \leq p$ add e to the edge set.

Is d_3 an edge?
(draw a random number. . .)



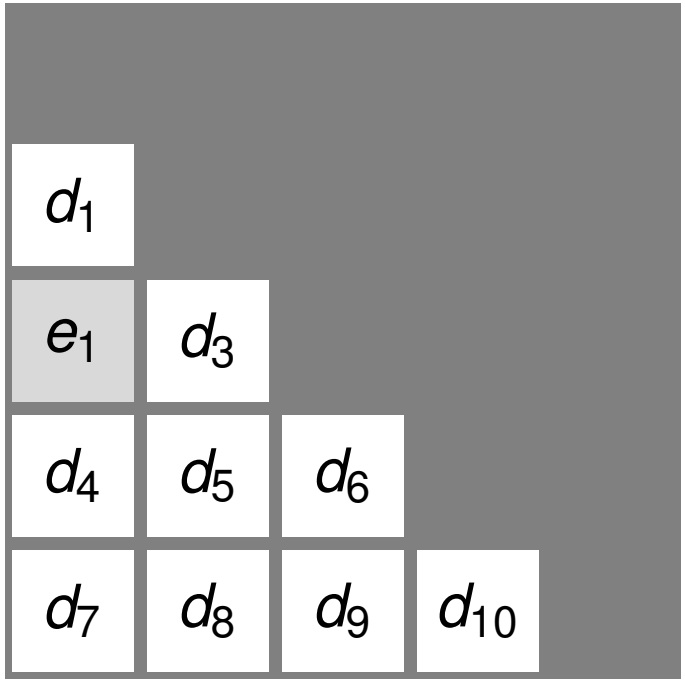
inefficient, when p is small (too many NOs)

Generating graphs from $\mathcal{G}(n, p)$ (naive algorithm).

Naive algorithm: iterate over all dyads $e \in D$

- draw a uniformly distributed random number $r \in [0, 1]$;
- if $r \leq p$ add e to the edge set.

Is d_3 an edge?
$\rightarrow$ NO (for instance)



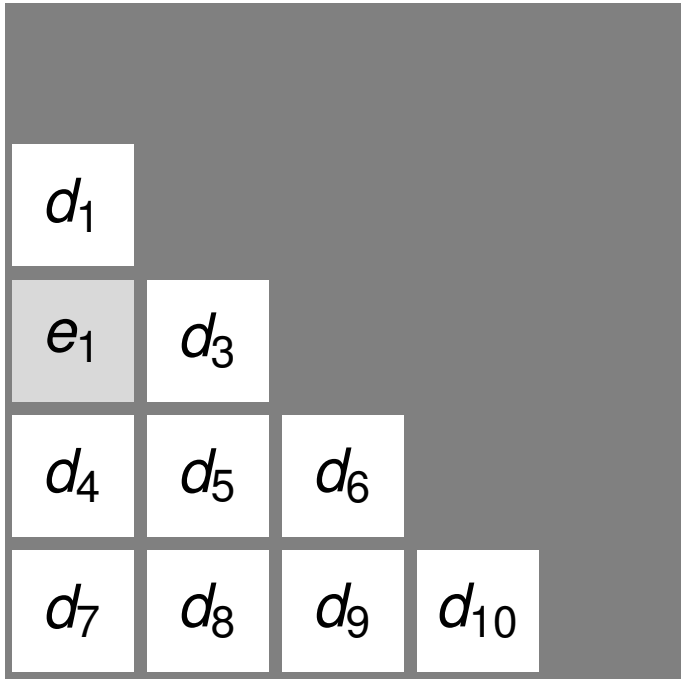
inefficient, when p is small (too many NOs)

Generating graphs from $\mathcal{G}(n, p)$ (naive algorithm).

Naive algorithm: iterate over all dyads $e \in D$

- draw a uniformly distributed random number $r \in [0, 1]$;
- if $r \leq p$ add e to the edge set.

go on . . .



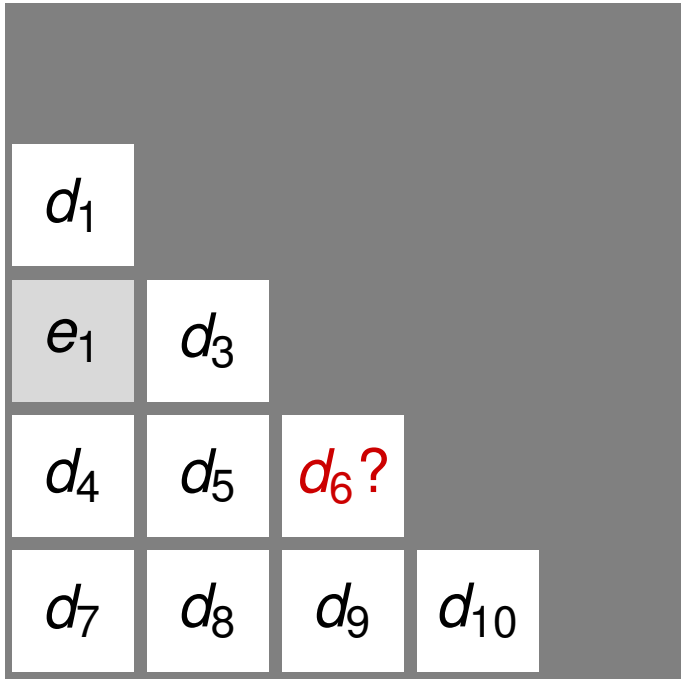
inefficient, when p is small (too many NOs)

Generating graphs from $\mathcal{G}(n, p)$ (naive algorithm).

Naive algorithm: iterate over all dyads $e \in D$

- draw a uniformly distributed random number $r \in [0, 1]$;
- if $r \leq p$ add e to the edge set.

Is d_6 an edge?
(draw a random number. . .)



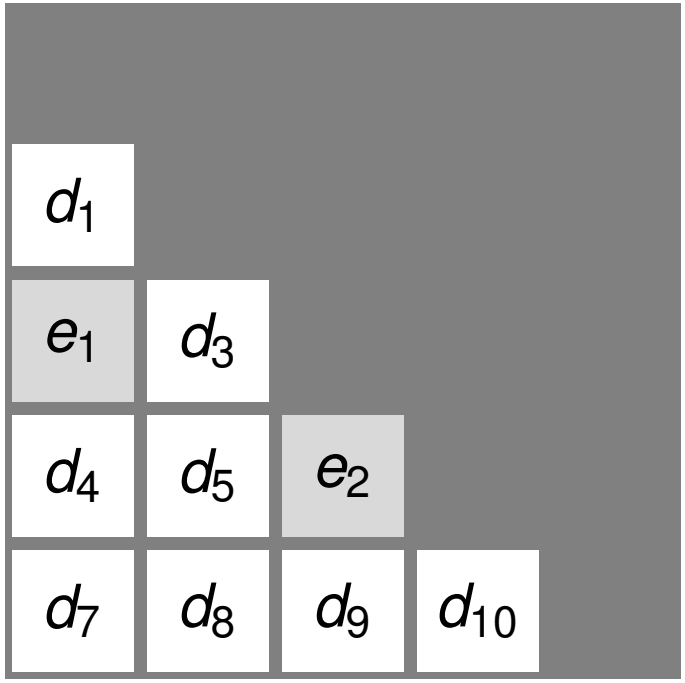
inefficient, when p is small (too many NOs)

Generating graphs from $\mathcal{G}(n, p)$ (naive algorithm).

Naive algorithm: iterate over all dyads $e \in D$

- draw a uniformly distributed random number $r \in [0, 1]$;
- if $r \leq p$ add e to the edge set.

Is d_6 an edge?
$\rightarrow$ YES (for instance)
$\Rightarrow$ turn d_6 into the second edge



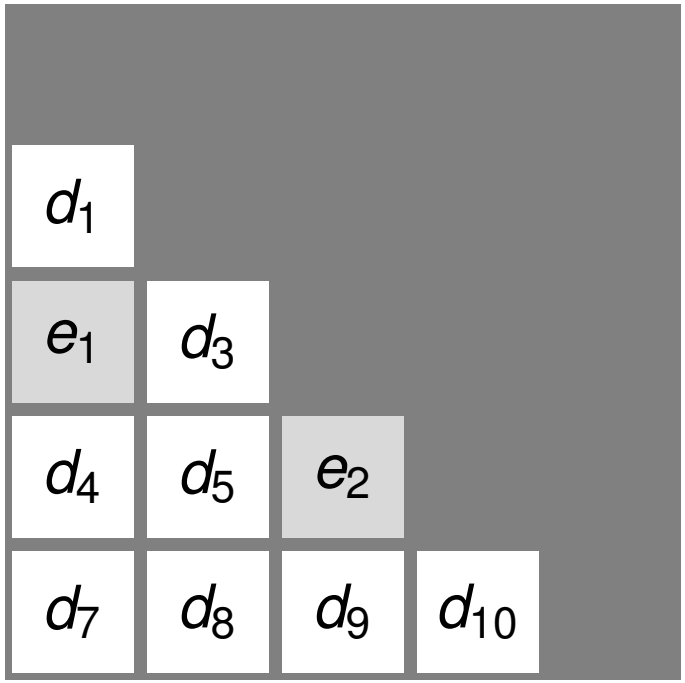
inefficient, when p is small (too many NOs)

Generating graphs from $\mathcal{G}(n, p)$ (naive algorithm).

Naive algorithm: iterate over all dyads $e \in D$

- draw a uniformly distributed random number $r \in [0, 1]$;
- if $r \leq p$ add e to the edge set.

to be continued . . .



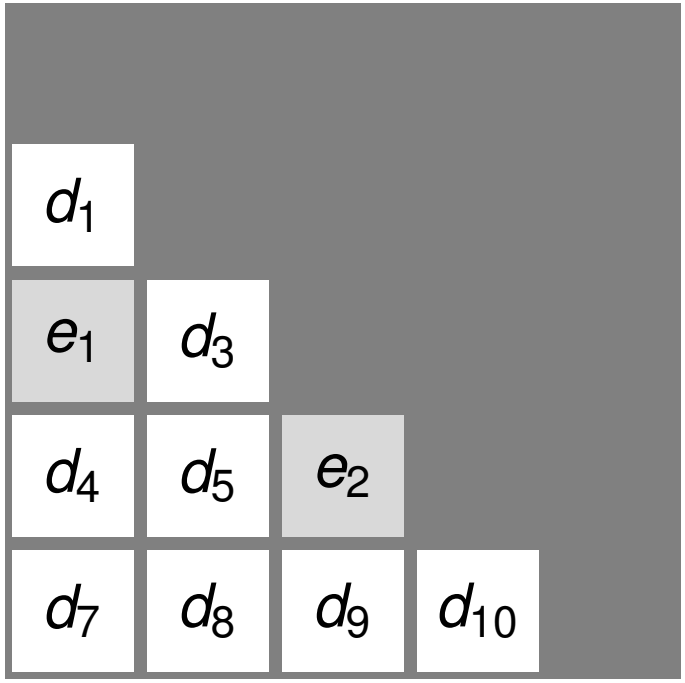
inefficient, when p is small (too many NOs)

Generating graphs from $\mathcal{G}(n, p)$ (naive algorithm).

Naive algorithm: iterate over all dyads $e \in D$

- draw a uniformly distributed random number $r \in [0, 1]$;
- if $r \leq p$ add e to the edge set.

to be continued . . .



inefficient, when p is small (too many NOs)

Background: sparse graphs and dense graphs.

Let n be the number of vertices and m the number of edges.

In undirected, loopless graphs it is
$0 \leq m \leq n(n-1)/2 \in \Theta(n^2)$.

A family of graphs with unbounded $n = 1, 2, 3, \ldots$ is called

- *dense* if $m \in \Theta(n^2)$;
- *sparse* if $m \in \mathcal{O}(n)$;
- (in between: notation depends on context).

Density of sparse graphs tends to zero: $p \in \mathcal{O}(1/n)$.
Average degree of sparse graphs is bounded by constant:
$\overline{d} \in \mathcal{O}(1)$.

Empirical observation: social networks are typically sparse.

Background: sparse graphs and dense graphs.

Let n be the number of vertices and m the number of edges.

In undirected, loopless graphs it is
$0 \leq m \leq n(n-1)/2 \in \Theta(n^2)$.

A family of graphs with unbounded $n = 1, 2, 3, \ldots$ is called

- *dense* if $m \in \Theta(n^2)$;
- *sparse* if $m \in \mathcal{O}(n)$;
- (in between: notation depends on context).

Density of sparse graphs tends to zero: $p \in \mathcal{O}(1/n)$.
Average degree of sparse graphs is bounded by constant:
$\overline{d} \in \mathcal{O}(1)$.

Empirical observation: social networks are typically sparse.

Background: sparse graphs and dense graphs.

Let n be the number of vertices and m the number of edges.

In undirected, loopless graphs it is
$0 \leq m \leq n(n-1)/2 \in \Theta(n^2)$.

A family of graphs with unbounded $n = 1, 2, 3, \ldots$ is called

- *dense* if $m \in \Theta(n^2)$;
- *sparse* if $m \in \mathcal{O}(n)$;
- (in between: notation depends on context).

Density of sparse graphs tends to zero: $p \in \mathcal{O}(1/n)$.
Average degree of sparse graphs is bounded by constant:
$\overline{d} \in \mathcal{O}(1)$.

Empirical observation: social networks are typically sparse.

Background: sparse graphs and dense graphs.

Let n be the number of vertices and m the number of edges.

In undirected, loopless graphs it is
$0 \leq m \leq n(n-1)/2 \in \Theta(n^2)$.

A family of graphs with unbounded $n = 1, 2, 3, \ldots$ is called

- *dense* if $m \in \Theta(n^2)$;
- *sparse* if $m \in \mathcal{O}(n)$;
- (in between: notation depends on context).

Density of sparse graphs tends to zero: $p \in \mathcal{O}(1/n)$.
Average degree of sparse graphs is bounded by constant:
$\overline{d} \in \mathcal{O}(1)$.

Empirical observation: social networks are typically sparse.

$\mathcal{G}(n, p)$ efficient generation.

Naive algorithm: iterate over all dyads $e \in D$

- draw a uniformly distributed random number $r \in [0, 1]$;
- if $r \leq p$ add e to the edge set.

Runtime: is in $\Theta(n^2)$
$\Rightarrow$ asymptotically larger than the expected graph size, $\mathcal{O}(n + m)$, if p is decreasing with n (sparse graphs).

$\mathcal{G}(n, p)$ efficient generation.

Better ask the question:

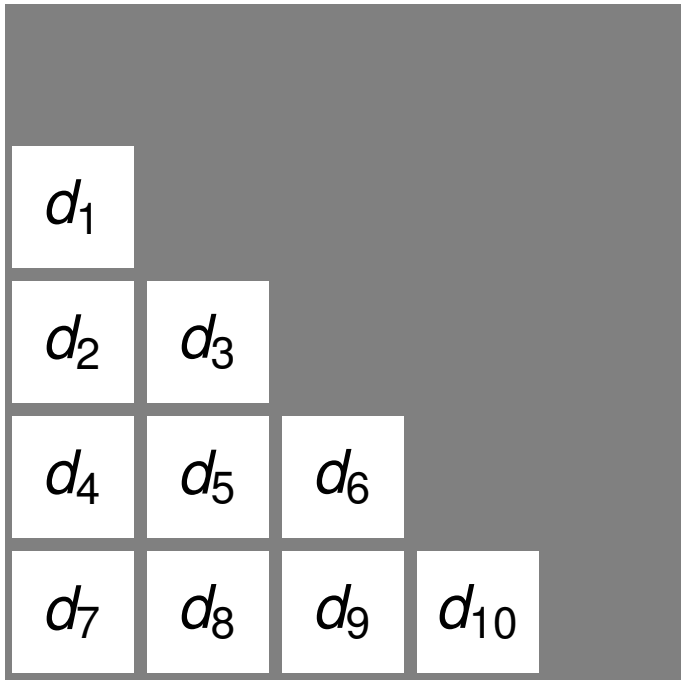
> *How many dyads shall be left out before the next edge?*

$\Rightarrow$ need only $\Theta(m)$ questions.

Randomly draw the number k of non-edges . . .

draw $k = 1 \Rightarrow$ leave out one dyad; turn the second dyad into the first edge

draw $k = 3 \Rightarrow$ leave out the next three dyads (d_3, d_4, d_5); turn d_6 into the second edge



$\mathcal{G}(n, p)$ efficient generation.

Better ask the question:

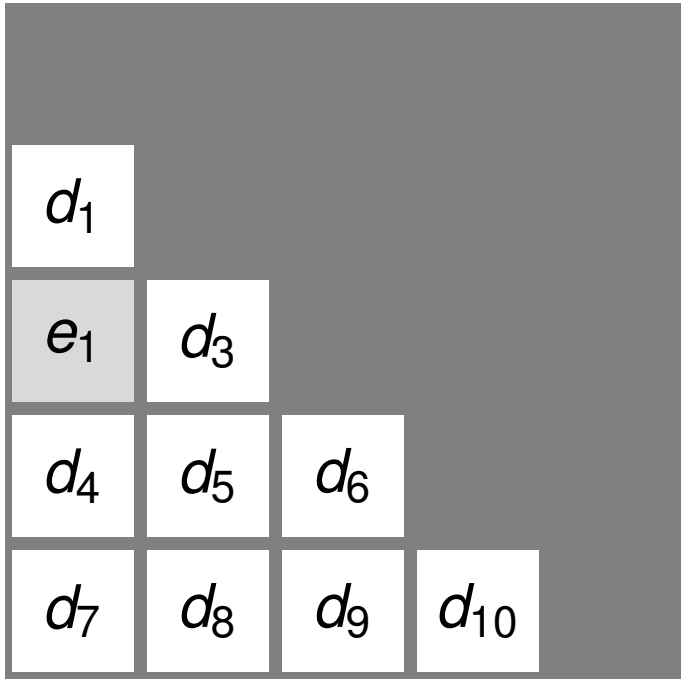
> *How many dyads shall be left out before the next edge?*

$\Rightarrow$ need only $\Theta(m)$ questions.

Randomly draw the number k of non-edges . . .

draw $k = 1 \Rightarrow$ leave out one dyad; turn the second dyad into the first edge

draw $k = 3 \Rightarrow$ leave out the next three dyads (d_3, d_4, d_5); turn d_6 into the second edge



$\mathcal{G}(n, p)$ efficient generation.

Better ask the question:

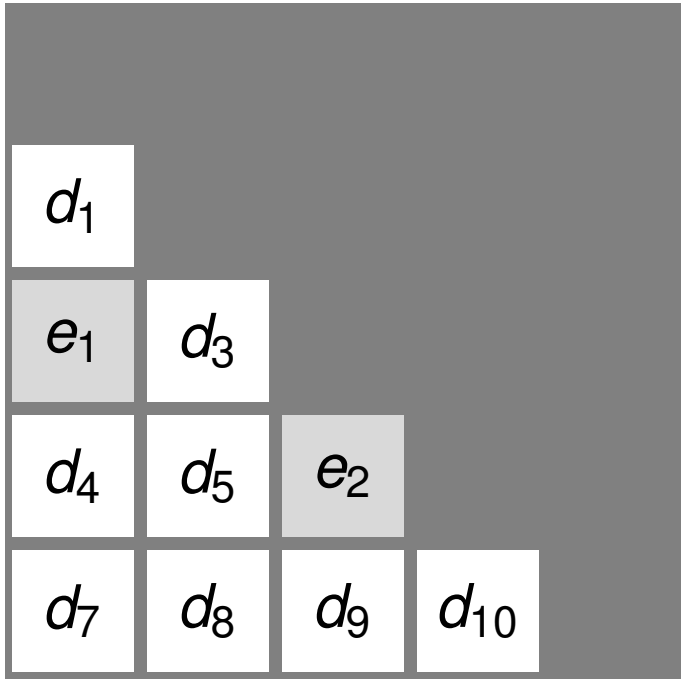
> *How many dyads shall be left out before the next edge?*

$\Rightarrow$ need only $\Theta(m)$ questions.

Randomly draw the number k of non-edges . . .

draw $k = 1 \Rightarrow$ leave out one dyad; turn the second dyad into the first edge

draw $k = 3 \Rightarrow$ leave out the next three dyads (d_3, d_4, d_5); turn d_6 into the second edge



$\mathcal{G}(n, p)$ efficient generation.

How many dyads shall be left out?

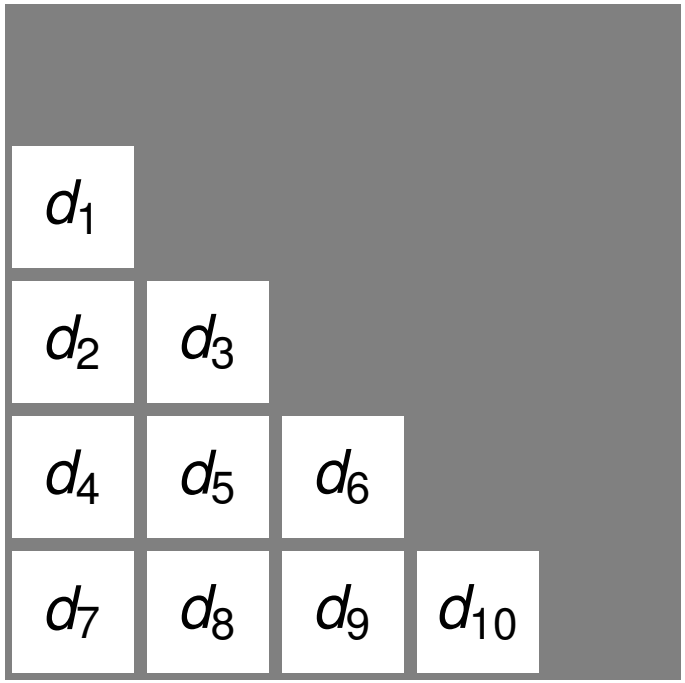
(Notation: $q = 1 - p$ on the following slides.)

Observation: the next dyad that becomes an edge is preceeded by exactly k non-edges with probability $q^k p$.

$\Rightarrow$ randomly draw number k of non-edges (out of $0, 1, \ldots$) with probability $q^k p$ and add the $k + 1$th dyad to the edge set.

draw $k = 1 \Rightarrow$ leave out one dyad

draw $k = 3 \Rightarrow$ leave out three dyads



$\mathcal{G}(n, p)$ efficient generation.

How many dyads shall be left out?

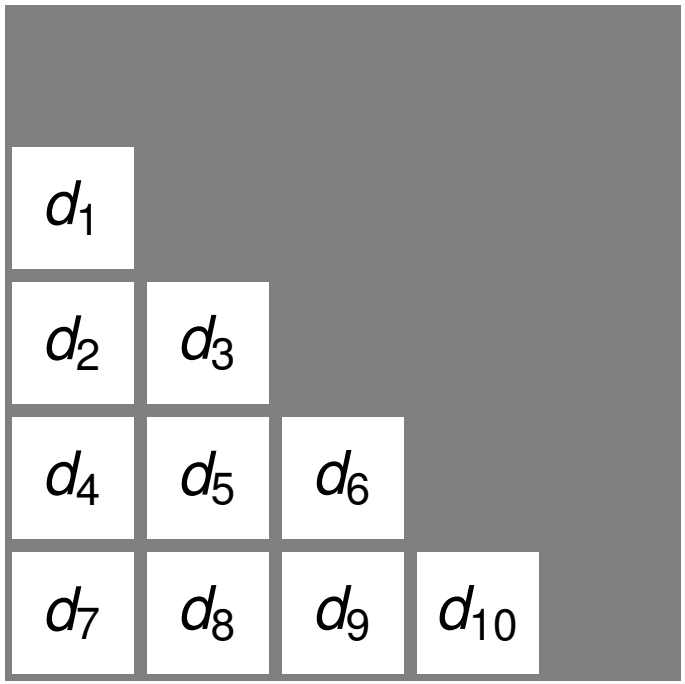
(Notation: $q = 1 - p$ on the following slides.)

Observation: the next dyad that becomes an edge is preceeded by exactly k non-edges with probability $q^k p$.

$\Rightarrow$ randomly draw number k of non-edges (out of $0, 1, \ldots$) with probability $q^k p$ and add the $k + 1$th dyad to the edge set.

draw $k = 1 \Rightarrow$ leave out one dyad

draw $k = 3 \Rightarrow$ leave out three dyads



$\mathcal{G}(n, p)$ efficient generation.

How many dyads shall be left out?

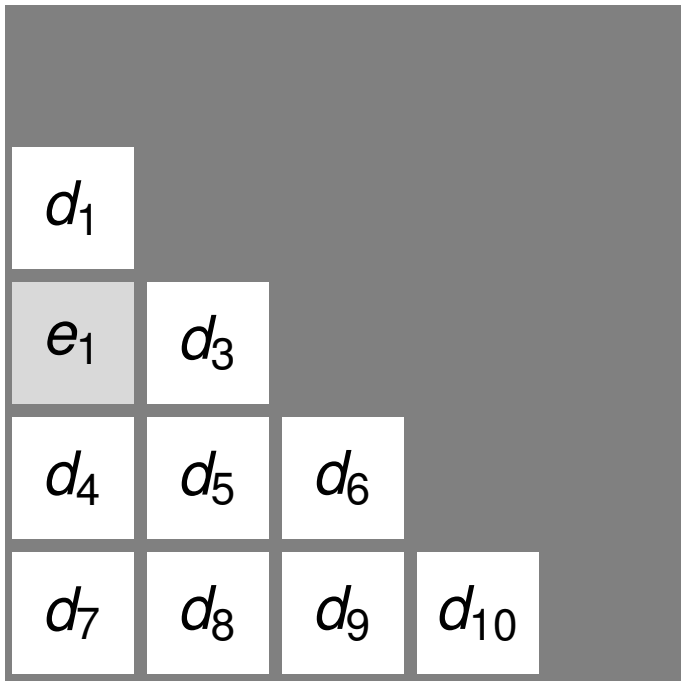
(Notation: $q = 1 - p$ on the following slides.)

Observation: the next dyad that becomes an edge is preceeded by exactly k non-edges with probability $q^k p$.

$\Rightarrow$ randomly draw number k of non-edges (out of $0, 1, \ldots$) with probability $q^k p$ and add the $k + 1$th dyad to the edge set.

draw $k = 1 \Rightarrow$ leave out one dyad

draw $k = 3 \Rightarrow$ leave out three dyads



$\mathcal{G}(n, p)$ efficient generation.

How many dyads shall be left out?

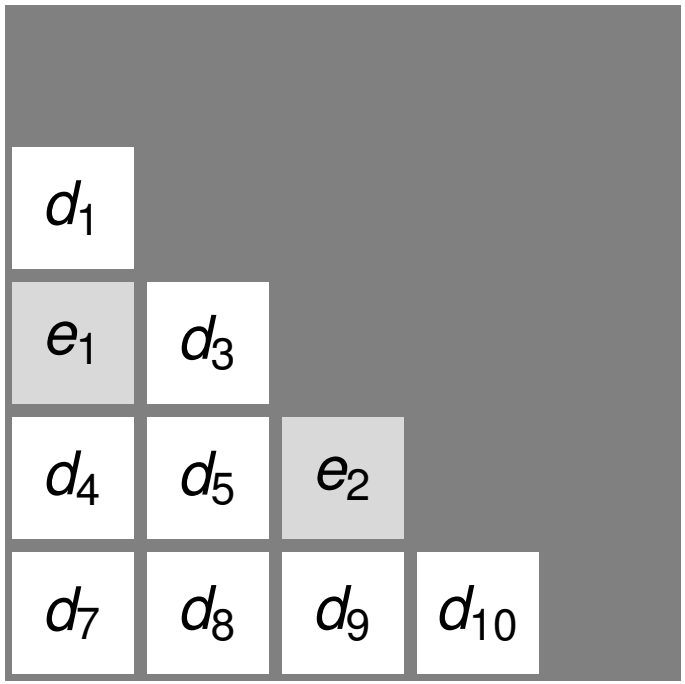
(Notation: $q = 1 - p$ on the following slides.)

Observation: the next dyad that becomes an edge is preceeded by exactly k non-edges with probability $q^k p$.

$\Rightarrow$ randomly draw number k of non-edges (out of $0, 1, \ldots$) with probability $q^k p$ and add the $k + 1$th dyad to the edge set.

draw $k = 1 \Rightarrow$ leave out one dyad

draw $k = 3 \Rightarrow$ leave out three dyads



$\mathcal{G}(n,p)$ efficient generation.

Randomly draw number k of non-edges with probability $q^k p$.

Associate $k = 0, 1, \ldots$ with interval $I_k \subset [0,1]$ of length $q^k p$.

$I_0 = [0, p]$, $I_1 = [p, p + qp]$, $I_2 = [p + qp, p + qp + q^2 p]$, $\ldots$

Note that

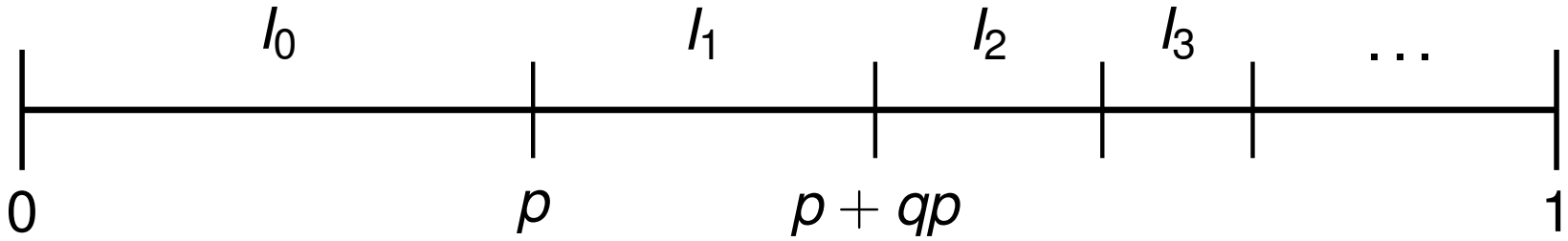
$$\sum_{k=0}^{\infty} q^k p \quad = \quad p \cdot \sum_{k=0}^{\infty} q^k \quad = \quad p \cdot \frac{1}{1-q} \quad = \quad 1 \ .$$

For $r \leftarrow \text{random}([0,1])$ choose k such that r is in I_k.

$\mathcal{G}(n,p)$ efficient generation.

Randomly draw number k of non-edges with probability $q^k p$.

Associate $k = 0, 1, \ldots$ with interval $I_k \subset [0,1]$ of length $q^k p$.



$I_0 = [0, p]$, $I_1 = [p, p+qp]$, $I_2 = [p+qp, p+qp+q^2 p]$, $\ldots$

Note that

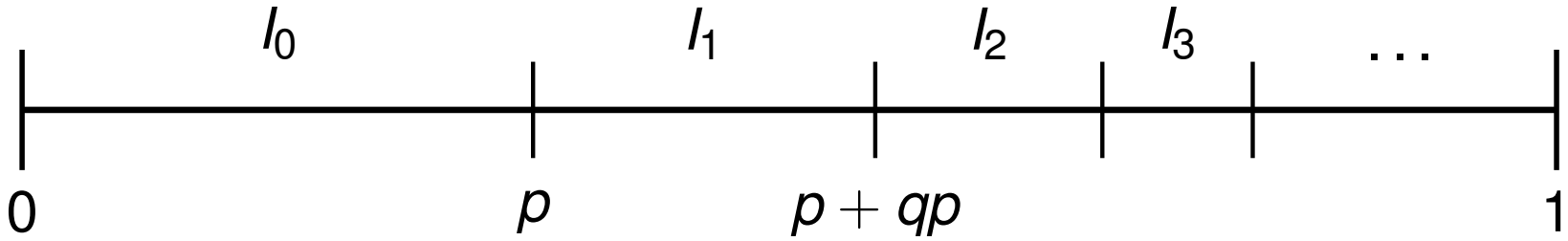
$$\sum_{k=0}^{\infty} q^k p \quad = \quad p \cdot \sum_{k=0}^{\infty} q^k \quad = \quad p \cdot \frac{1}{1-q} \quad = \quad 1 \ .$$

For $r \leftarrow \text{random}([0,1])$ choose k such that r is in I_k.

$\mathcal{G}(n,p)$ efficient generation.

Randomly draw number k of non-edges with probability $q^k p$.

Associate $k = 0, 1, \ldots$ with interval $I_k \subset [0,1]$ of length $q^k p$.



$I_0 = [0, p]$, $I_1 = [p, p + qp]$, $I_2 = [p + qp, p + qp + q^2 p]$, $\ldots$

Note that

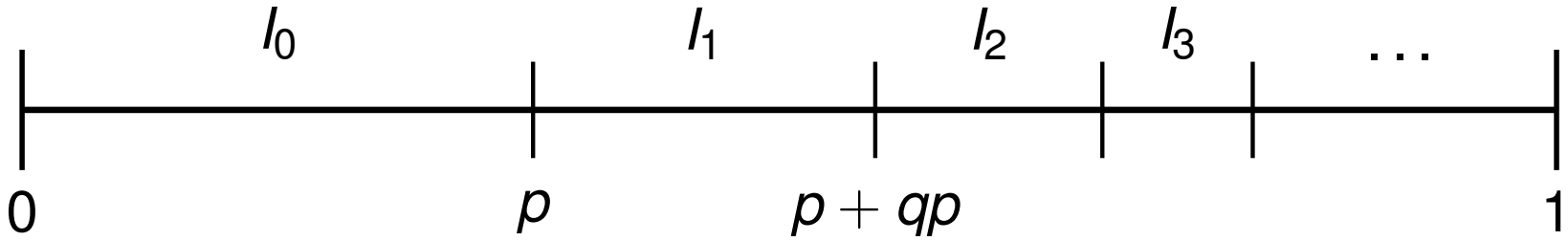
$$\sum_{k=0}^{\infty} q^k p \quad = \quad p \cdot \sum_{k=0}^{\infty} q^k \quad = \quad p \cdot \frac{1}{1-q} \quad = \quad 1 \ .$$

For $r \leftarrow \text{random}([0,1])$ choose k such that r is in I_k.

$\mathcal{G}(n,p)$ efficient generation.

Randomly draw number k of non-edges with probability $q^k p$.

Associate $k = 0, 1, \ldots$ with interval $I_k \subset [0,1]$ of length $q^k p$.



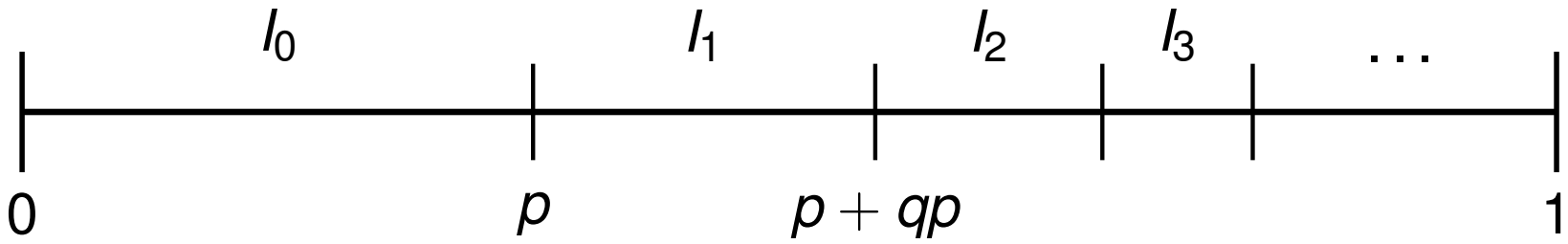
$I_0 = [0, p]$, $I_1 = [p, p + qp]$, $I_2 = [p + qp, p + qp + q^2 p]$, $\ldots$

Note that

$$\sum_{k=0}^{\infty} q^k p \quad = \quad p \cdot \sum_{k=0}^{\infty} q^k \quad = \quad p \cdot \frac{1}{1-q} \quad = \quad 1 \ .$$

For $r \leftarrow \text{random}([0,1])$ choose k such that r is in I_k.

$\mathcal{G}(n,p)$ efficient generation.



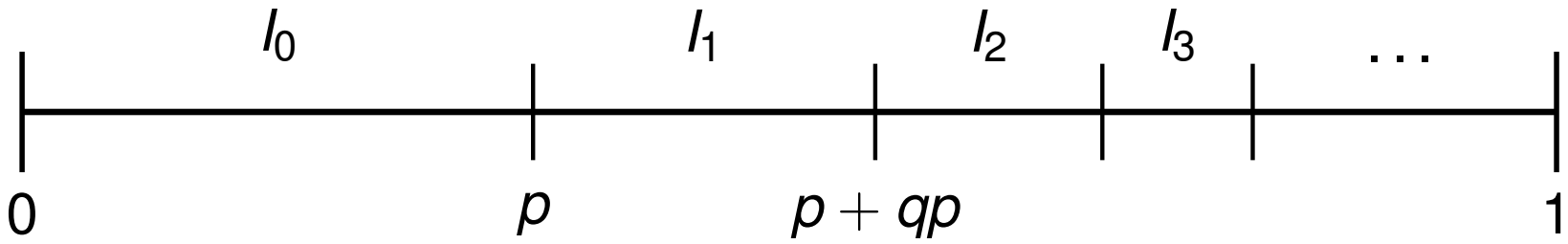
For $r \leftarrow \text{random}([0,1])$ choose k such that r is in I_k.

The interval I_k ends at

$$\sum_{i=0}^{k} q^i p \quad = \quad p \cdot \sum_{i=0}^{k} q^i \quad = \quad p \cdot \frac{1-q^{k+1}}{1-q} \quad = \quad 1-q^{k+1}\,.$$

For $r \leftarrow \text{random}([0,1])$ compute the minimum k such that I_k ends after r, i. e., such that $1-q^{k+1} > r$.

$\mathcal{G}(n,p)$ efficient generation.



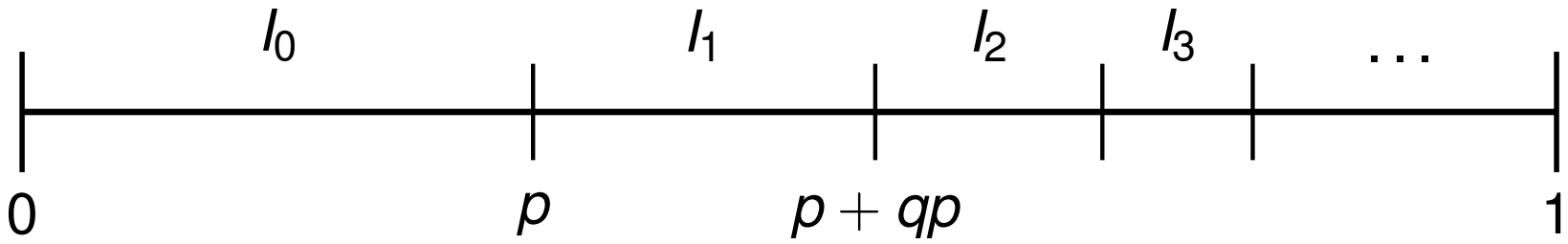
For $r \leftarrow \text{random}([0,1])$ choose k such that r is in I_k.

The interval I_k ends at

$$\sum_{i=0}^{k} q^i p \quad = \quad p \cdot \sum_{i=0}^{k} q^i \quad = \quad p \cdot \frac{1-q^{k+1}}{1-q} \quad = \quad 1-q^{k+1}\,.$$

For $r \leftarrow \text{random}([0,1])$ compute the minimum k such that I_k ends after r, i. e., such that $1-q^{k+1} > r$.

$\mathcal{G}(n,p)$ efficient generation.



For $r \leftarrow \text{random}([0,1])$ choose k such that r is in I_k.

The interval I_k ends at

$$\sum_{i=0}^{k} q^i p \quad = \quad p \cdot \sum_{i=0}^{k} q^i \quad = \quad p \cdot \frac{1-q^{k+1}}{1-q} \quad = \quad 1-q^{k+1}\,.$$

For $r \leftarrow \text{random}([0,1])$ compute the minimum k such that I_k ends after r, i. e., such that $1-q^{k+1} > r$.

$\mathcal{G}(n,p)$ efficient generation.

For $r \leftarrow \text{random}([0,1])$ compute the minimum k such that $1 - q^{k+1} > r$.

The following inequalities are equivalent.

$$\begin{aligned} r &< 1 - q^{k+1} \\ q^{k+1} &< 1 - r \\ (k+1)\log q &< \log(1-r) \\ k &> \frac{\log(1-r)}{\log q} - 1 \end{aligned}$$

Leave out $k := \left\lfloor \frac{\log(1-r)}{\log q} \right\rfloor$ dyads and insert the $k+1$ dyad in the edge set.

$\mathcal{G}(n,p)$ efficient generation.

For $r \leftarrow \text{random}([0,1])$ compute the minimum k such that $1 - q^{k+1} > r$.

The following inequalities are equivalent.

$$\begin{aligned} r &< 1 - q^{k+1} \\ q^{k+1} &< 1 - r \\ (k+1)\log q &< \log(1-r) \\ k &> \frac{\log(1-r)}{\log q} - 1 \end{aligned}$$

Leave out $k := \left\lfloor \frac{\log(1-r)}{\log q} \right\rfloor$ dyads and insert the $k+1$ dyad in the edge set.

$\mathcal{G}(n,p)$ efficient generation.

For $r \leftarrow \text{random}([0,1])$ compute the minimum k such that $1 - q^{k+1} > r$.

The following inequalities are equivalent.

$$\begin{aligned} r &< 1 - q^{k+1} \\ q^{k+1} &< 1 - r \\ (k+1)\log q &< \log(1-r) \\ k &> \frac{\log(1-r)}{\log q} - 1 \end{aligned}$$

Leave out $k := \left\lfloor \frac{\log(1-r)}{\log q} \right\rfloor$ dyads and insert the $k+1$ dyad in the edge set.

$\mathcal{G}(n,p)$ efficient generation.

For $r \leftarrow \text{random}([0,1])$ compute the minimum k such that $1 - q^{k+1} > r$.

The following inequalities are equivalent.

$$\begin{aligned} r &< 1 - q^{k+1} \\ q^{k+1} &< 1 - r \\ (k+1)\log q &< \log(1-r) \\ k &> \frac{\log(1-r)}{\log q} - 1 \end{aligned}$$

Leave out $k := \left\lfloor \frac{\log(1-r)}{\log q} \right\rfloor$ dyads and insert the $k+1$ dyad in the edge set.

$\mathcal{G}(n,p)$ efficient generation.

For $r \leftarrow \text{random}([0,1])$ compute the minimum k such that $1 - q^{k+1} > r$.

The following inequalities are equivalent.

$$\begin{aligned} r &< 1 - q^{k+1} \\ q^{k+1} &< 1 - r \\ (k+1)\log q &< \log(1-r) \\ k &> \frac{\log(1-r)}{\log q} - 1 \end{aligned}$$

Leave out $k := \left\lfloor \frac{\log(1-r)}{\log q} \right\rfloor$ dyads and insert the $k+1$ dyad in the edge set.

$\mathcal{G}(n,p)$ efficient generation.

For $r \leftarrow \text{random}([0,1])$ compute the minimum k such that $1 - q^{k+1} > r$.

The following inequalities are equivalent.

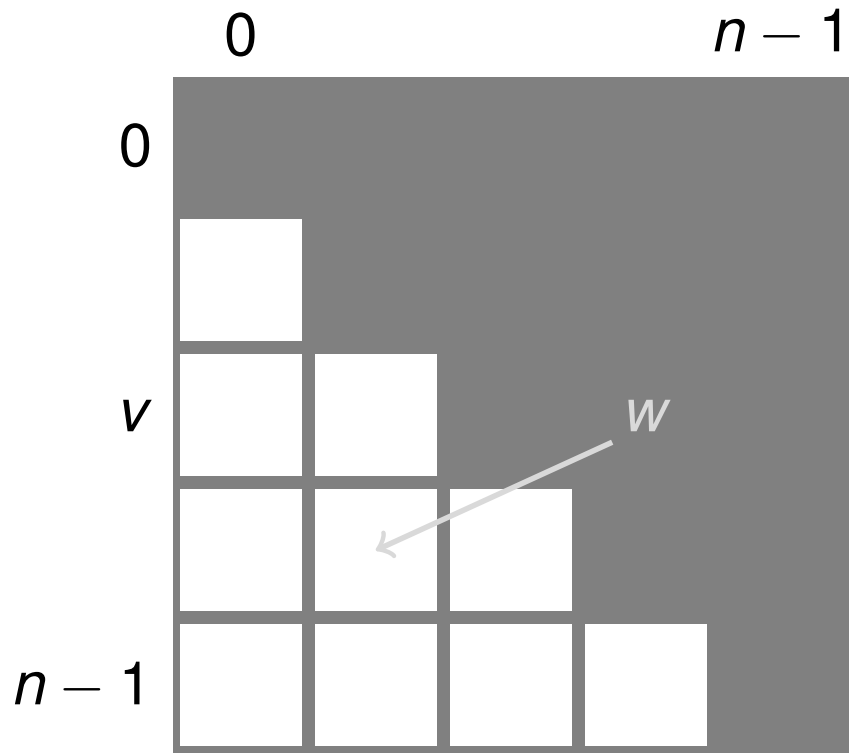
$$\begin{aligned} r &< 1 - q^{k+1} \\ q^{k+1} &< 1 - r \\ (k+1)\log q &< \log(1-r) \\ k &> \frac{\log(1-r)}{\log q} - 1 \end{aligned}$$

Leave out $k := \left\lfloor \frac{\log(1-r)}{\log q} \right\rfloor$ dyads and insert the $k+1$ dyad in the edge set.

$\mathcal{G}(n,p)$ efficient generation (algorithm).

(Insert the $\left\lfloor \frac{\log(1-r)}{\log(1-p)} \right\rfloor + 1$ dyad in the edge set.)

$E \leftarrow \emptyset$
$v \leftarrow 1$ $w \leftarrow -1$
while $v < n$ **do**
  $r \leftarrow \text{random}([0,1])$
  $w \leftarrow w + 1 + \left\lfloor \frac{\log(1-r)}{\log(1-p)} \right\rfloor$
  while $w \geq v$ **and** $v < n$ **do**
    $w \leftarrow w - v$; $v \leftarrow v + 1$
  if $v < n$ **then**
    $E \leftarrow E \cup \{\{v, w\}\}$
return $G = (V, E)$

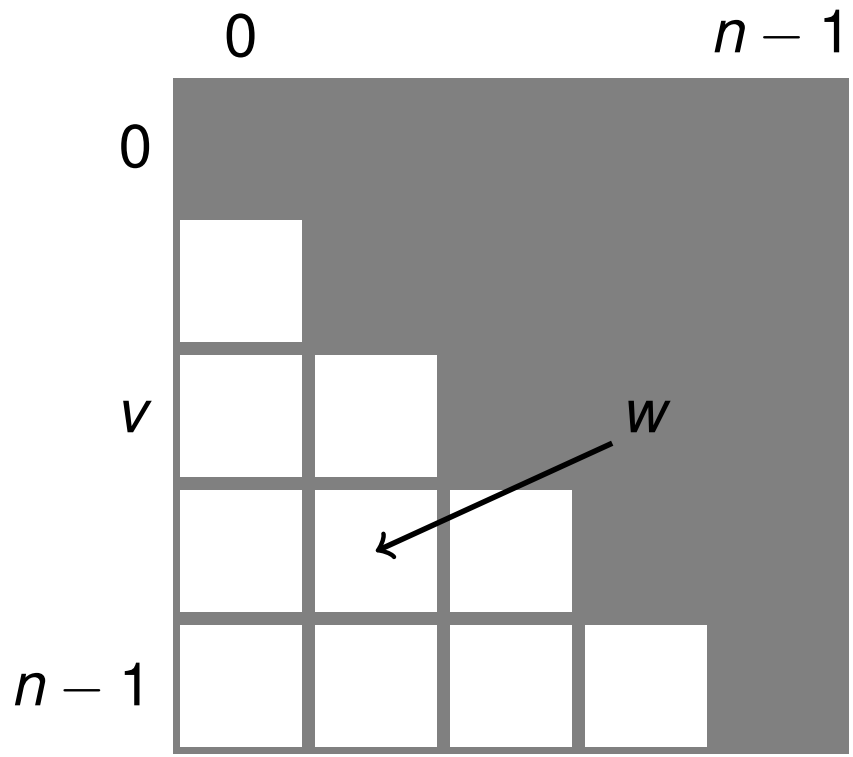


If $w \geq v$ then w is reduced by v and the row index v is incremented by one.

$\mathcal{G}(n,p)$ efficient generation (algorithm).

(Insert the $\left\lfloor \frac{\log(1-r)}{\log(1-p)} \right\rfloor + 1$ dyad in the edge set.)

$E \leftarrow \emptyset$
$v \leftarrow 1$ $w \leftarrow -1$
while $v < n$ **do**
  $r \leftarrow \text{random}([0,1])$
  $w \leftarrow w + 1 + \left\lfloor \frac{\log(1-r)}{\log(1-p)} \right\rfloor$
  while $w \geq v$ **and** $v < n$ **do**
    $w \leftarrow w - v$; $v \leftarrow v + 1$
  if $v < n$ **then**
    $E \leftarrow E \cup \{\{v, w\}\}$
return $G = (V, E)$



If $w \geq v$ then w is reduced by v and the row index v is incremented by one.

$\mathcal{G}(n,p)$ efficient generation (runtime).

$E \leftarrow \emptyset$
$v \leftarrow 1$ $w \leftarrow -1$
while $v < n$ **do**
  $r \leftarrow \text{random}([0,1])$
  $w \leftarrow w + 1 + \left\lfloor \frac{\log(1-r)}{\log(1-p)} \right\rfloor$
  while $w \geq v$ **and** $v < n$ **do**
    $w \leftarrow w - v;\ v \leftarrow v + 1$
  if $v < n$ **then**
    $E \leftarrow E \cup \{\{v, w\}\}$
return $G = (V, E)$

$\Rightarrow$ runtime in $\mathcal{O}(m+n)$.

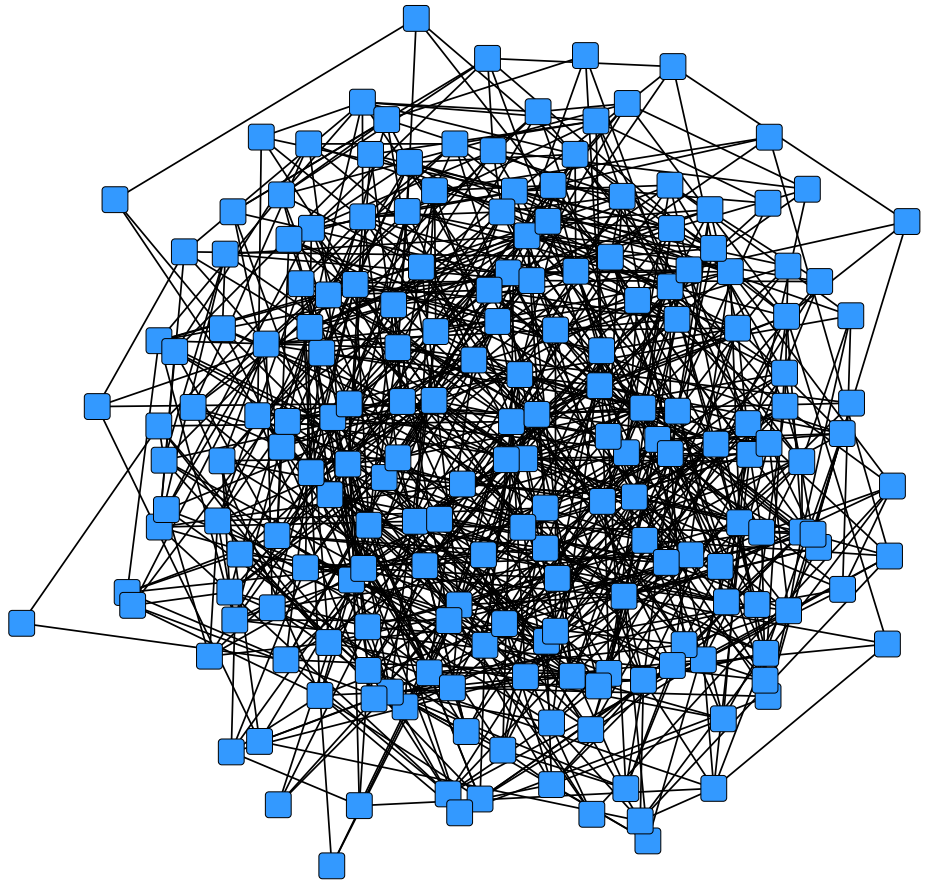
Outer `while` loop is executed $m+1$ times (m is number of edges of G).

Inner `while` loop is executed (in total) n times.

Is $\mathcal{G}(n, p)$ a good model for social networks?

Observed network; assumed to be drawn from a $\mathcal{G}(n, p)$ model.

What is the most likely value for the parameter p?



Graph has $n = 200$ vertices and $m = 915$ edges.

$\mathcal{G}(n,p)$ statistical inference of p.

Problem: given a graph G generated from some $\mathcal{G}(n,p)$ (without knowing the parameter p).

What is the most likely value for p?

Definition (maximum likelihood)

$(\mathcal{G}, P_\theta)$ random graph model parameterized by $\theta \in \Theta \subseteq \mathbb{R}^k$;
$G_{\text{obs}} \in \mathcal{G}$ a graph (observation).
Likelihood function associated with G_{obs}

$$L\colon \Theta \to \mathbb{R};\ \theta \mapsto P_\theta(G_{\text{obs}})$$

A parameter vector $\hat{\theta}$ maximizing L, i. e.,

$$\hat{\theta} = \arg\max_\theta L(\theta)$$

is called a *maximum likelihood estimate (MLE)* for θ.

$\mathcal{G}(n,p)$ statistical inference of p.

Problem: given a graph G generated from some $\mathcal{G}(n,p)$ (without knowing the parameter p).

What is the most likely value for p?

Definition (maximum likelihood)

$(\mathcal{G}, P_\theta)$ random graph model parameterized by $\theta \in \Theta \subseteq \mathbb{R}^k$;
$G_{\text{obs}} \in \mathcal{G}$ a graph (observation).
Likelihood function associated with G_{obs}

$$L\colon \Theta \to \mathbb{R};\ \theta \mapsto P_\theta(G_{\text{obs}})$$

A parameter vector $\hat{\theta}$ maximizing L, i. e.,

$$\hat{\theta} = \arg\max_\theta L(\theta)$$

is called a *maximum likelihood estimate (MLE)* for θ.

$\mathcal{G}(n,p)$ statistical inference of p.

Problem: given a graph G generated from some $\mathcal{G}(n,p)$ (without knowing the parameter p).

What is the most likely value for p?

Definition (maximum likelihood)

$(\mathcal{G}, P_\theta)$ random graph model parameterized by $\theta \in \Theta \subseteq \mathbb{R}^k$; $G_{\text{obs}} \in \mathcal{G}$ a graph (observation).

Likelihood function associated with G_{obs}

$$L\colon \Theta \to \mathbb{R};\ \theta \mapsto P_\theta(G_{\text{obs}})$$

A parameter vector $\hat{\theta}$ maximizing L, i. e.,

$$\hat{\theta} = \arg\max_\theta L(\theta)$$

is called a *maximum likelihood estimate (MLE)* for θ.

$\mathcal{G}(n,p)$ statistical inference of p.

Problem: given a graph G generated from some $\mathcal{G}(n,p)$ (without knowing the parameter p).

What is the most likely value for p?

Definition (maximum likelihood)

$(\mathcal{G}, P_\theta)$ random graph model parameterized by $\theta \in \Theta \subseteq \mathbb{R}^k$;
$G_{\text{obs}} \in \mathcal{G}$ a graph (observation).
Likelihood function associated with G_{obs}

$$L\colon \Theta \to \mathbb{R};\ \theta \mapsto P_\theta(G_{\text{obs}})$$

A parameter vector $\hat{\theta}$ maximizing L, i. e.,

$$\hat{\theta} = \arg\max_\theta L(\theta)$$

is called a *maximum likelihood estimate (MLE)* for θ.

$\mathcal{G}(n,p)$ statistical inference of p.

Problem: given a graph G generated from some $\mathcal{G}(n,p)$ (without knowing the parameter p).

What is the most likely value for p?

Definition (maximum likelihood)

$(\mathcal{G}, P_\theta)$ random graph model parameterized by $\theta \in \Theta \subseteq \mathbb{R}^k$;
$G_{\text{obs}} \in \mathcal{G}$ a graph (observation).
Likelihood function associated with G_{obs}

$$L\colon \Theta \to \mathbb{R};\ \theta \mapsto P_\theta(G_{\text{obs}})$$

A parameter vector $\hat{\theta}$ maximizing L, i. e.,

$$\hat{\theta} = \arg\max_\theta L(\theta)$$

is called a *maximum likelihood estimate (MLE)* for θ.

Maximum likelihood estimate of p (derivation).

Assume that G_{obs} has exactly m edges; let $M = \frac{n(n-1)}{2}$.

$$\begin{aligned} L(p) &= P_p(G_{\text{obs}}) = p^m(1-p)^{M-m} \ . \\ L'(p) &= m \cdot p^{m-1} \cdot (1-p)^{M-m} - p^m \cdot (M-m) \cdot (1-p)^{M-m-1} \ . \end{aligned}$$

Setting $L'(p) = 0$ for $0 < p < 1$ yields

$$\begin{aligned} m \cdot p^{m-1} \cdot (1-p)^{M-m} &= p^m \cdot (M-m) \cdot (1-p)^{M-m-1} \\ m \cdot (1-p) &= p \cdot (M-m) \\ m - pm &= pM - pm \\ \frac{m}{M} &= p \end{aligned}$$

$L(p)$ indeed assumes a maximum at $\hat{p} := \frac{m}{M}$ since [...].

Maximum likelihood estimate of p (derivation).

Assume that G_{obs} has exactly m edges; let $M = \frac{n(n-1)}{2}$.

$$\begin{aligned} L(p) &= P_p(G_{\text{obs}}) = p^m(1-p)^{M-m} \ . \\ L'(p) &= m \cdot p^{m-1} \cdot (1-p)^{M-m} - p^m \cdot (M-m) \cdot (1-p)^{M-m-1} \ . \end{aligned}$$

Setting $L'(p) = 0$ for $0 < p < 1$ yields

$$\begin{aligned} m \cdot p^{m-1} \cdot (1-p)^{M-m} &= p^m \cdot (M-m) \cdot (1-p)^{M-m-1} \\ m \cdot (1-p) &= p \cdot (M-m) \\ m - pm &= pM - pm \\ \frac{m}{M} &= p \end{aligned}$$

$L(p)$ indeed assumes a maximum at $\hat{p} := \frac{m}{M}$ since [...].

Maximum likelihood estimate of p (derivation).

Assume that G_{obs} has exactly m edges; let $M = \frac{n(n-1)}{2}$.

$$\begin{aligned} L(p) &= P_p(G_{\text{obs}}) = p^m(1-p)^{M-m} \ . \\ L'(p) &= m \cdot p^{m-1} \cdot (1-p)^{M-m} - p^m \cdot (M-m) \cdot (1-p)^{M-m-1} \ . \end{aligned}$$

Setting $L'(p) = 0$ for $0 < p < 1$ yields

$$\begin{aligned} m \cdot p^{m-1} \cdot (1-p)^{M-m} &= p^m \cdot (M-m) \cdot (1-p)^{M-m-1} \\ m \cdot (1-p) &= p \cdot (M-m) \\ m - pm &= pM - pm \\ \frac{m}{M} &= p \end{aligned}$$

$L(p)$ indeed assumes a maximum at $\hat{p} := \frac{m}{M}$ since [...].

Maximum likelihood estimate of p (derivation).

Assume that G_{obs} has exactly m edges; let $M = \frac{n(n-1)}{2}$.

$$\begin{aligned} L(p) &= P_p(G_{\text{obs}}) = p^m(1-p)^{M-m} \ . \\ L'(p) &= m \cdot p^{m-1} \cdot (1-p)^{M-m} - p^m \cdot (M-m) \cdot (1-p)^{M-m-1} \ . \end{aligned}$$

Setting $L'(p) = 0$ for $0 < p < 1$ yields

$$\begin{aligned} m \cdot p^{m-1} \cdot (1-p)^{M-m} &= p^m \cdot (M-m) \cdot (1-p)^{M-m-1} \\ m \cdot (1-p) &= p \cdot (M-m) \\ m - pm &= pM - pm \\ \frac{m}{M} &= p \end{aligned}$$

$L(p)$ indeed assumes a maximum at $\hat{p} := \frac{m}{M}$ since [...].

Maximum likelihood estimate of p (derivation).

Assume that G_{obs} has exactly m edges; let $M = \frac{n(n-1)}{2}$.

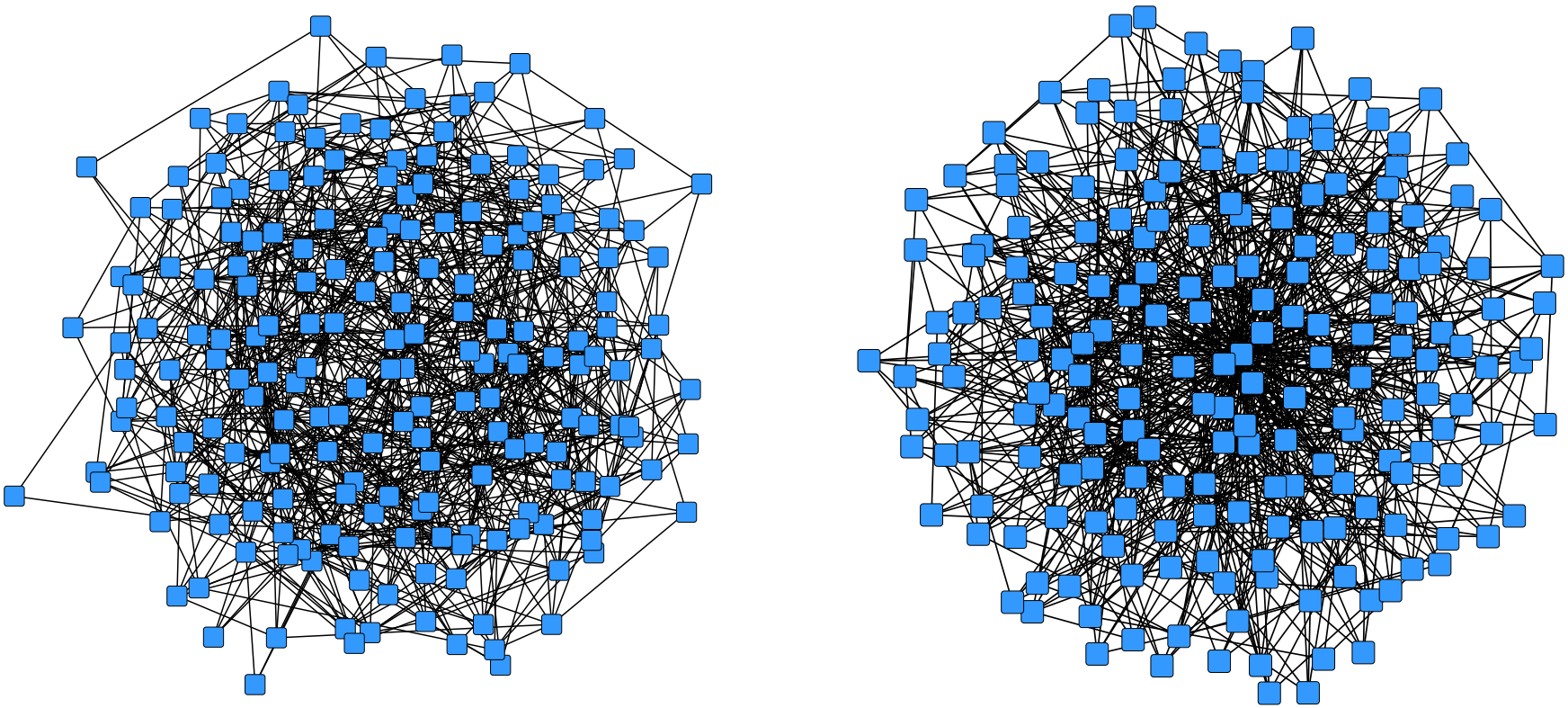
$$\begin{aligned} L(p) &= P_p(G_{\text{obs}}) = p^m(1-p)^{M-m} \ . \\ L'(p) &= m \cdot p^{m-1} \cdot (1-p)^{M-m} - p^m \cdot (M-m) \cdot (1-p)^{M-m-1} \ . \end{aligned}$$

Setting $L'(p) = 0$ for $0 < p < 1$ yields

$$\begin{aligned} m \cdot p^{m-1} \cdot (1-p)^{M-m} &= p^m \cdot (M-m) \cdot (1-p)^{M-m-1} \\ m \cdot (1-p) &= p \cdot (M-m) \\ m - pm &= pM - pm \\ \frac{m}{M} &= p \end{aligned}$$

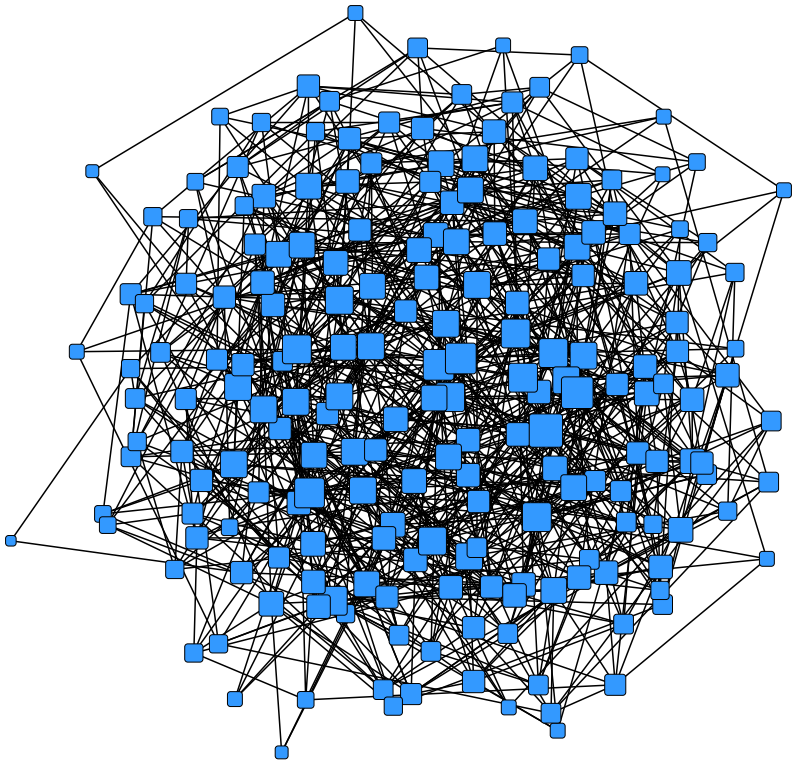
$L(p)$ indeed assumes a maximum at $\hat{p} := \frac{m}{M}$ since [...].

Which graph is drawn from a $\mathcal{G}(n, p)$ model?

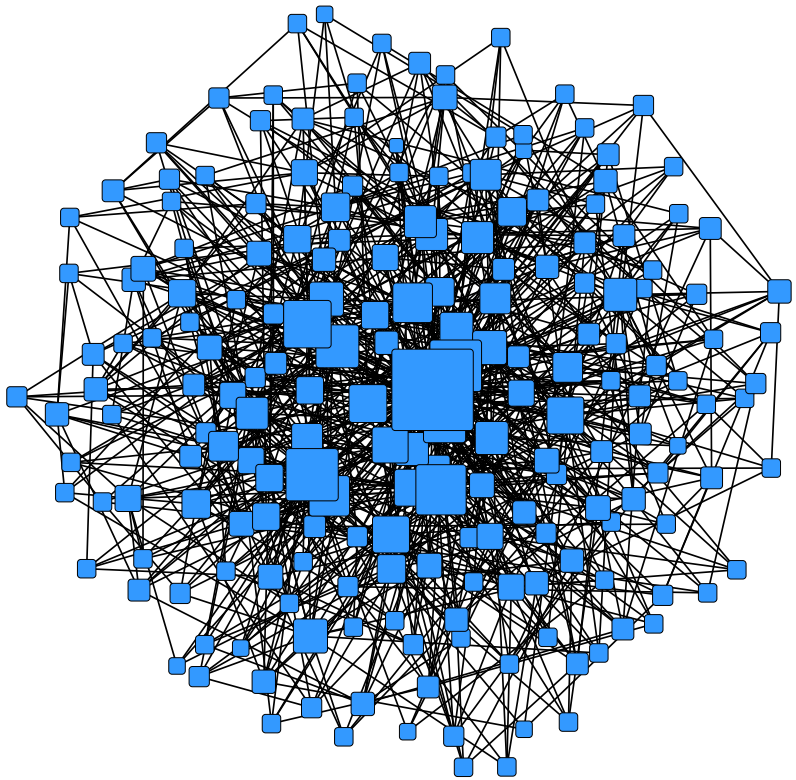


Both have $n = 200$ vertices and $m = 915$ edges.

Vertex size proportional to degree.



Max. degree is 20.



Max. degree is 97.

$\mathcal{G}(n,p)$ probability of degree k.

Lemma

Let $v \in \{1,\ldots,n\}$ be any vertex. The probability that v has degree equal to $k \in \{0,\ldots,n-1\}$ in a graph drawn from $\mathcal{G}(n,p)$ is

$$P(d(v)=k) = \binom{n-1}{k} \cdot p^k q^{n-1-k}$$

Proof.

There are exactly $\binom{n-1}{k}$ different neighborhoods of v that have cardinality k. Each of them has probability $p^k q^{n-1-k}$. □

$\mathcal{G}(n,p)$ probability of degree k.

Lemma

Let $v \in \{1,\ldots,n\}$ be any vertex. The probability that v has degree equal to $k \in \{0,\ldots,n-1\}$ in a graph drawn from $\mathcal{G}(n,p)$ is

$$P(d(v)=k) = \binom{n-1}{k} \cdot p^k q^{n-1-k}$$

Proof.

There are exactly $\binom{n-1}{k}$ different neighborhoods of v that have cardinality k. Each of them has probability $p^k q^{n-1-k}$. □

$\mathcal{G}(n,p)$ probability of degree k.

details on the proof: let

$$\mathcal{N}_k(v) = \{\{v_1, \ldots, v_k\} \subseteq V \setminus \{v\}\}$$

be the set of k-element subsets of $V \setminus \{v\}$ (potential neighborhoods of size k of v).
Define for $U \in \mathcal{N}_k(v)$ the subset

$$\mathcal{G}_U = \{G \in \mathcal{G}\,;\, \forall u \in U\colon \{u,v\} \in E_G \text{ and } \forall u \notin U\colon \{u,v\} \notin E_G\}$$

(all graphs in which the neighborhood of v equals U).
Important fact: $\mathcal{G}_U$ and $\mathcal{G}_{U'}$ are disjoint for $U \neq U'$. Thus

$$P[d(v) = k] = \sum_{U \in \mathcal{N}_k(v)} P(\mathcal{G}_U) = \binom{n-1}{k} \cdot p^k q^{n-1-k}$$

$\mathcal{G}(n,p)$ probability of degree k.

details on the proof: let

$$\mathcal{N}_k(v) = \{\{v_1, \ldots, v_k\} \subseteq V \setminus \{v\}\}$$

be the set of k-element subsets of $V \setminus \{v\}$ (potential neighborhoods of size k of v).
Define for $U \in \mathcal{N}_k(v)$ the subset

$$\mathcal{G}_U = \{G \in \mathcal{G}\,;\ \forall u \in U\colon \{u,v\} \in E_G \text{ and } \forall u \notin U\colon \{u,v\} \notin E_G\}$$

(all graphs in which the neighborhood of v equals U).
Important fact: $\mathcal{G}_U$ and $\mathcal{G}_{U'}$ are disjoint for $U \neq U'$. Thus

$$P[d(v) = k] = \sum_{U \in \mathcal{N}_k(v)} P(\mathcal{G}_U) = \binom{n-1}{k} \cdot p^k q^{n-1-k}$$

$\mathcal{G}(n,p)$ probability of degree k.

details on the proof: let

$$\mathcal{N}_k(v) = \{\{v_1, \ldots, v_k\} \subseteq V \setminus \{v\}\}$$

be the set of k-element subsets of $V \setminus \{v\}$ (potential neighborhoods of size k of v).
Define for $U \in \mathcal{N}_k(v)$ the subset

$$\mathcal{G}_U = \{G \in \mathcal{G}\,;\ \forall u \in U\colon \{u,v\} \in E_G \text{ and } \forall u \notin U\colon \{u,v\} \notin E_G\}$$

(all graphs in which the neighborhood of v equals U).
Important fact: $\mathcal{G}_U$ and $\mathcal{G}_{U'}$ are disjoint for $U \neq U'$. Thus

$$P[d(v)=k] = \sum_{U \in \mathcal{N}_k(v)} P(\mathcal{G}_U) = \binom{n-1}{k} \cdot p^k q^{n-1-k}$$

$\mathcal{G}(n,p)$ probability of degree k.

details on the proof: let

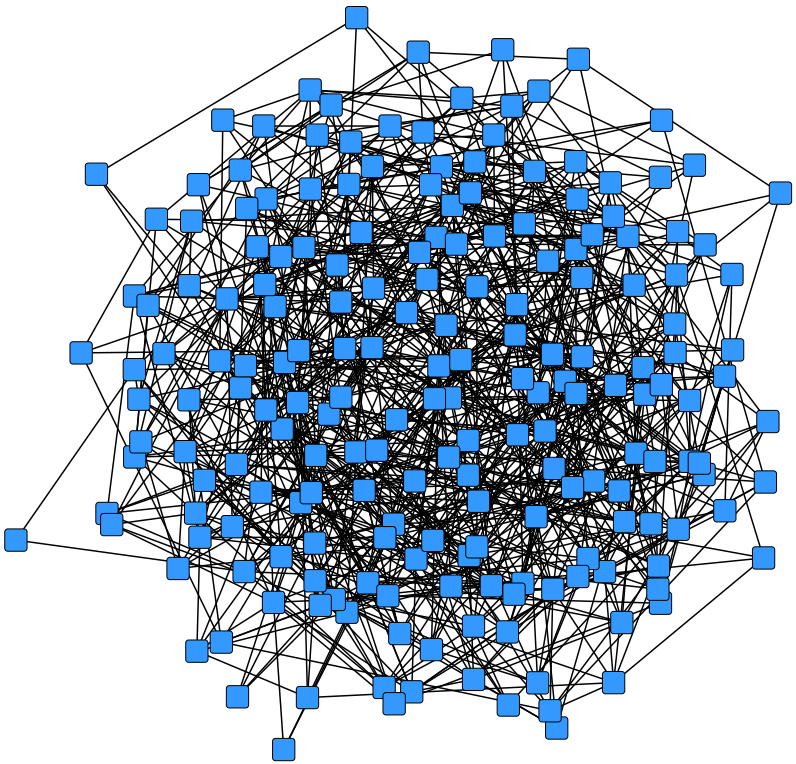
$$\mathcal{N}_k(v) = \{\{v_1, \ldots, v_k\} \subseteq V \setminus \{v\}\}$$

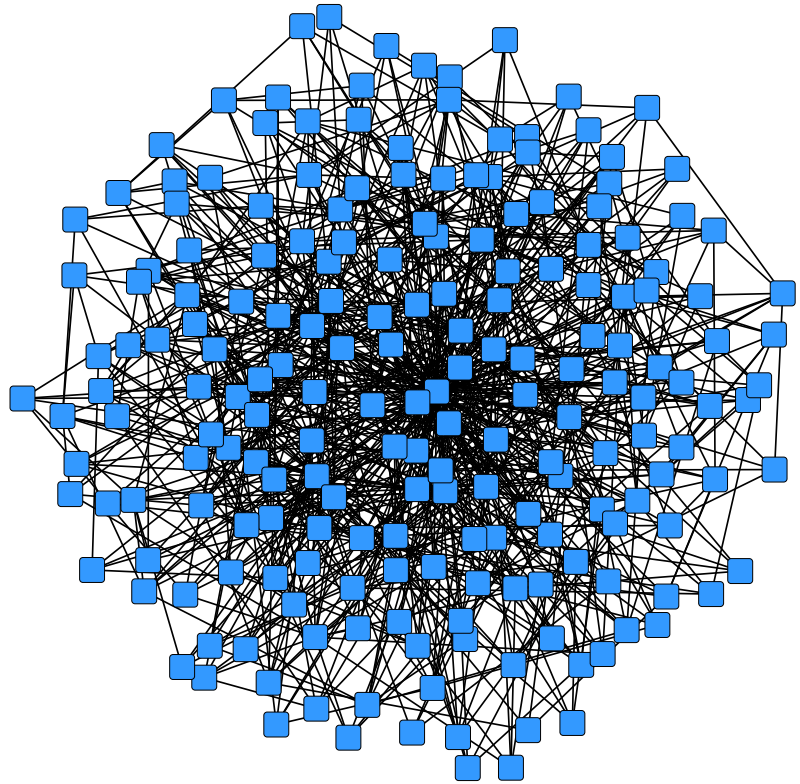
be the set of k-element subsets of $V \setminus \{v\}$ (potential neighborhoods of size k of v).
Define for $U \in \mathcal{N}_k(v)$ the subset

$$\mathcal{G}_U = \{G \in \mathcal{G}\,;\ \forall u \in U\colon \{u,v\} \in E_G \text{ and } \forall u \notin U\colon \{u,v\} \notin E_G\}$$

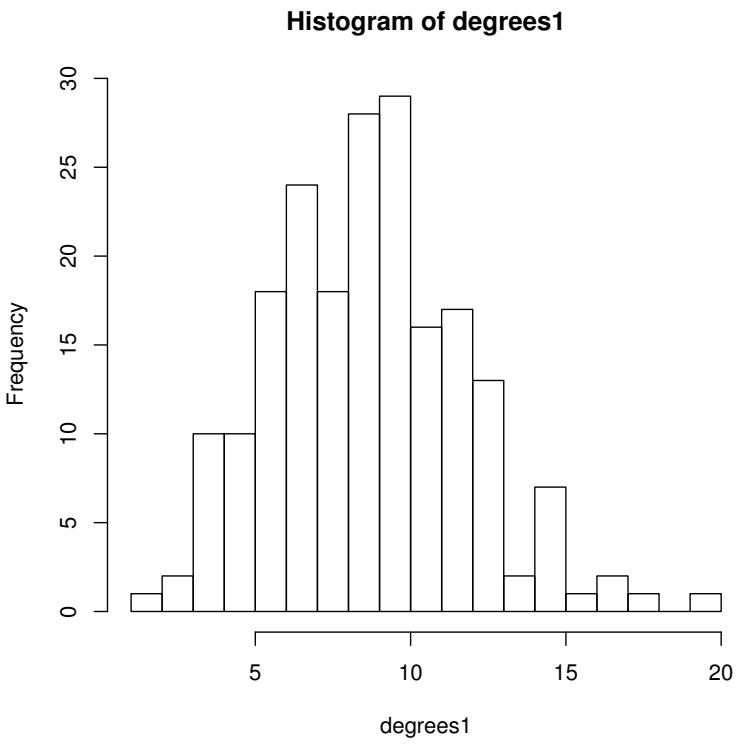
(all graphs in which the neighborhood of v equals U).
Important fact: $\mathcal{G}_U$ and $\mathcal{G}_{U'}$ are disjoint for $U \neq U'$. Thus

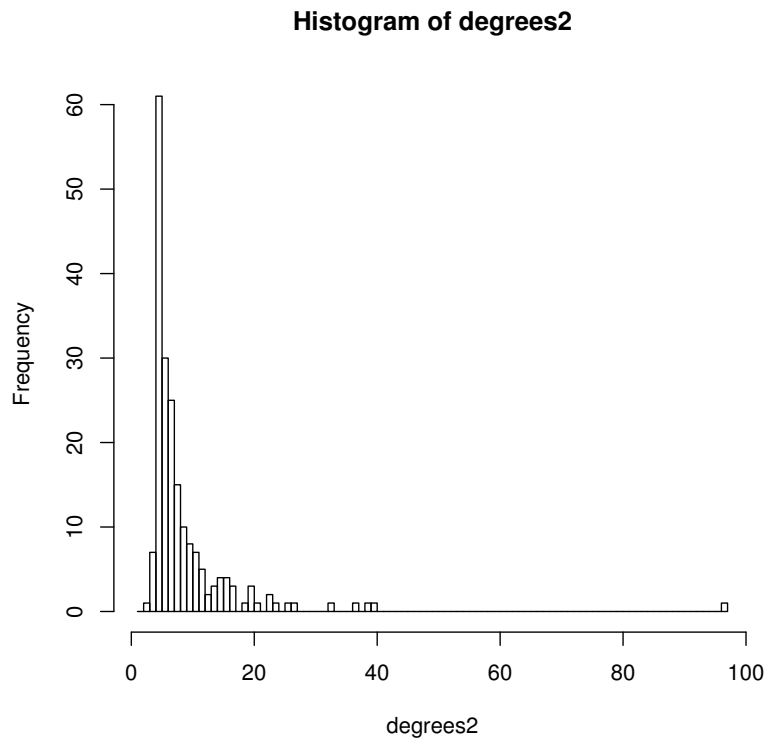
$$P[d(v) = k] = \sum_{U \in \mathcal{N}_k(v)} P(\mathcal{G}_U) = \binom{n-1}{k} \cdot p^k q^{n-1-k}$$





Let's look at the degree distributions of these graphs.


Histogram of degrees1
Frequency
0
5
10
15
20
25
30
5
10
15
20
degrees1


Histogram of degrees2
Frequency
0
10
20
30
40
50
60
0
20
40
60
80
100
degrees2

$\mathcal{G}(n,p)$ distribution of degrees (limit $n \to \infty$).

Theorem

Let $\lambda \in \mathbb{R}_{>0}$, $p_n := \lambda/(n-1)$ a sequence of edge probabilities, defined for $n \geq \lambda + 1$,
$k \in \mathbb{N}_0$, $P_n[d(v) = k]$ probability that $d(v) = k$ in $\mathcal{G}(n, p_n)$ for fixed v.

Then it is

$$\lim_{n\to\infty} P_n[d(v) = k] = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \ .$$

*(Is called **Poisson distribution**.)*

$\mathcal{G}(n,p)$ distribution of degrees (limit $n \to \infty$).

Theorem

Let $\lambda \in \mathbb{R}_{>0}$, $p_n := \lambda/(n-1)$ a sequence of edge probabilities, defined for $n \geq \lambda + 1$,
$k \in \mathbb{N}_0$, $P_n[d(v) = k]$ probability that $d(v) = k$ in $\mathcal{G}(n, p_n)$ for fixed v.

Then it is

$$\lim_{n\to\infty} P_n[d(v) = k] = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \ .$$

*(Is called **Poisson distribution**.)*

$\mathcal{G}(n,p)$ distribution of degrees (limit $n \to \infty$).

Theorem

Let $\lambda \in \mathbb{R}_{>0}$, $p_n := \lambda/(n-1)$ a sequence of edge probabilities, defined for $n \geq \lambda + 1$,
$k \in \mathbb{N}_0$, $P_n[d(v) = k]$ probability that $d(v) = k$ in $\mathcal{G}(n, p_n)$ for fixed v.

Then it is

$$\lim_{n\to\infty} P_n[d(v) = k] = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \ .$$

*(Is called **Poisson distribution**.)*

$\mathcal{G}(n,p)$ distribution of degrees (limit $n \to \infty$).

Theorem

Let $\lambda \in \mathbb{R}_{>0}$, $p_n := \lambda/(n-1)$ a sequence of edge probabilities, defined for $n \geq \lambda + 1$,
$k \in \mathbb{N}_0$, $P_n[d(v) = k]$ probability that $d(v) = k$ in $\mathcal{G}(n, p_n)$ for fixed v.

Then it is

$$\lim_{n\to\infty} P_n[d(v) = k] = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \ .$$

*(Is called **Poisson distribution**.)*

$\mathcal{G}(n,p)$ distribution of degrees (limit $n \to \infty$).

Theorem

Let $\lambda \in \mathbb{R}_{>0}$, $p_n := \lambda/(n-1)$ a sequence of edge probabilities, defined for $n \geq \lambda + 1$,
$k \in \mathbb{N}_0$, $P_n[d(v) = k]$ probability that $d(v) = k$ in $\mathcal{G}(n, p_n)$ for fixed v.

Then it is

$$\lim_{n\to\infty} P_n[d(v) = k] = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \ .$$

*(Is called **Poisson distribution**.)*

$\mathcal{G}(n,p)$ distribution of degrees (limit $n \to \infty$).

Proof.

$P_n(d(v) = k)$ equals

$$\begin{aligned}
\binom{n-1}{k} \cdot p_n^k q_n^{n-1-k} &= \frac{(n-1)!}{k!(n-1-k)!} \cdot p_n^k (1-p_n)^{n-1-k} \\
&= \frac{(n-1)!}{(n-1-k)!(1-p_n)^k} \cdot \frac{p_n^k}{k!} \cdot (1-p_n)^{n-1} \\
&= \frac{(n-1)!}{(n-1-k)!(1-p_n)^k (n-1)^k} \cdot \\
&\quad \frac{(n-1)^k p_n^k}{k!} \cdot (1-p_n)^{n-1}
\end{aligned}$$

The first factor converges for $n \to \infty$ to one, the second to $\frac{\lambda^k}{k!}$, and the third to $e^{-\lambda}$. □

$\mathcal{G}(n,p)$ distribution of degrees (limit $n \to \infty$).

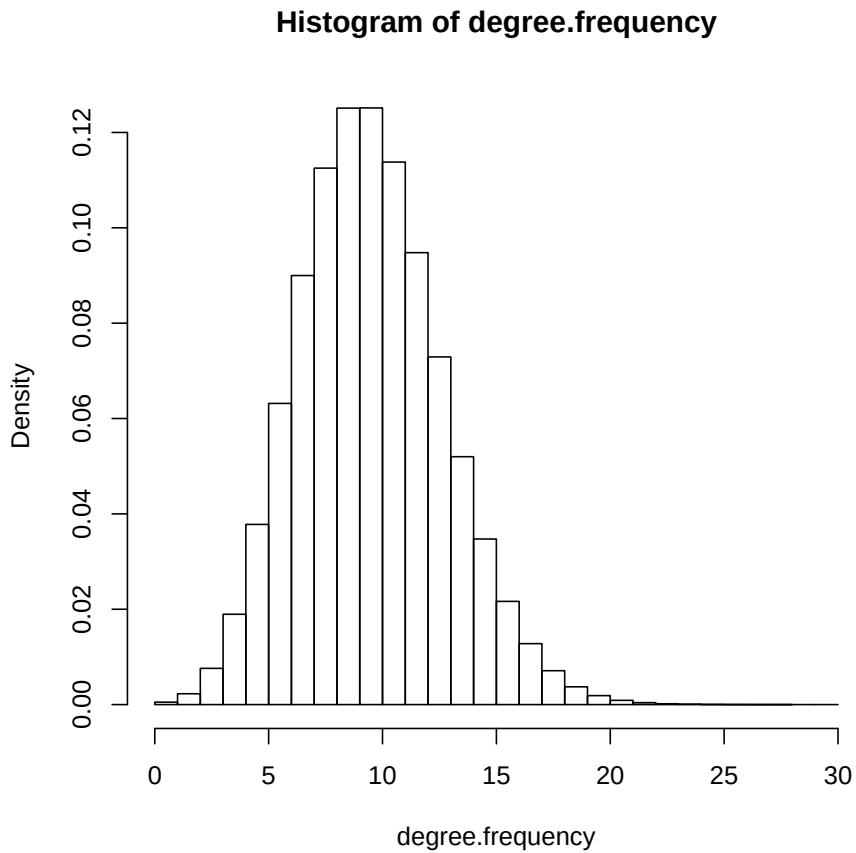
Proof.
$P_n(d(v) = k)$ equals

$$\begin{aligned}
\binom{n-1}{k} \cdot p_n^k q_n^{n-1-k} &= \frac{(n-1)!}{k!(n-1-k)!} \cdot p_n^k (1-p_n)^{n-1-k} \\
&= \frac{(n-1)!}{(n-1-k)!(1-p_n)^k} \cdot \frac{p_n^k}{k!} \cdot (1-p_n)^{n-1} \\
&= \frac{(n-1)!}{(n-1-k)!(1-p_n)^k (n-1)^k} \cdot \\
&\quad \frac{(n-1)^k p_n^k}{k!} \cdot (1-p_n)^{n-1}
\end{aligned}$$

The first factor converges for $n \to \infty$ to one, the second to $\frac{\lambda^k}{k!}$, and the third to $e^{-\lambda}$. □

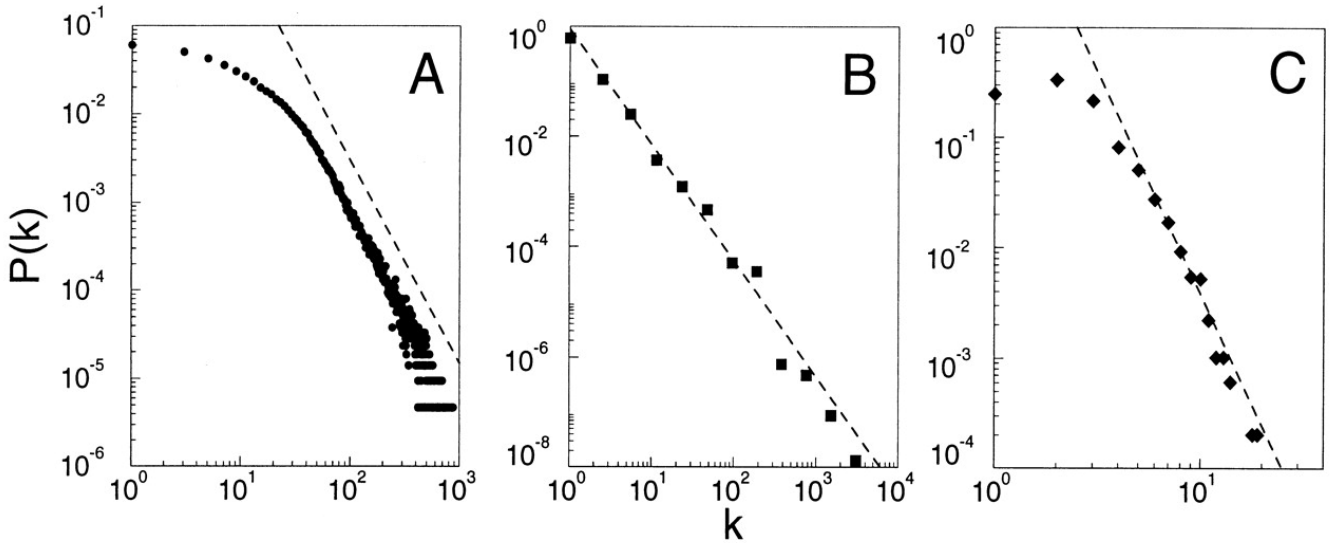
$\mathcal{G}(n, p)$ distribution of degrees (limit $n \to \infty$).

Degree distribution of a graph drawn from $\mathcal{G}(n, p)$ with $n = 10^7$ and $p = 10/(n-1)$; maximum observed degree is 30.



Empirical distributions (Barabasi and Albert, 1999).

Note: logarithmic scaling of axes.



A Actor collaboration network $n = 212,250$ and $\overline{d} = 28.78$

B WWW $n = 325,729$ and $\overline{d} = 5.46$

C Power grid $n = 4,941$ and $\overline{d} = 2.67$

Outline.

Preferential attachment: motivation and history.

Empirical observation: often a few nodes have very high degrees; degree-distribution resembles a *power-law*:

$$P(d(v) = k) \approx c \cdot \frac{1}{k^{\gamma}}$$

Model idea (Barabási and Albert, 1999):

1. vertices are successively added to the network;
2. new vertices create a fixed number of edges to already existing vertices;
3. probability of edge to vertex v is proportional to v's degree.

Interpretation high-degree vertices are more *popular*.

Experimental evidence for power-law distribution with $\gamma \approx 3$.

Preferential attachment: motivation and history.

Empirical observation: often a few nodes have very high degrees; degree-distribution resembles a *power-law*:

$$P(d(v) = k) \approx c \cdot \frac{1}{k^{\gamma}}$$

Model idea (Barabási and Albert, 1999):

1. vertices are successively added to the network;
2. new vertices create a fixed number of edges to already existing vertices;
3. probability of edge to vertex v is proportional to v's degree.

Interpretation high-degree vertices are more *popular*.

Experimental evidence for power-law distribution with $\gamma \approx 3$.

Preferential attachment: motivation and history.

Empirical observation: often a few nodes have very high degrees; degree-distribution resembles a *power-law*:

$$P(d(v) = k) \approx c \cdot \frac{1}{k^{\gamma}}$$

Model idea (Barabási and Albert, 1999):

1. vertices are successively added to the network;
2. new vertices create a fixed number of edges to already existing vertices;
3. probability of edge to vertex v is proportional to v's degree.

Interpretation high-degree vertices are more *popular*.

Experimental evidence for power-law distribution with $\gamma \approx 3$.

Preferential attachment: motivation and history.

Empirical observation: often a few nodes have very high degrees; degree-distribution resembles a *power-law*:

$$P(d(v) = k) \approx c \cdot \frac{1}{k^{\gamma}}$$

Model idea (Barabási and Albert, 1999):

1. vertices are successively added to the network;
2. new vertices create a fixed number of edges to already existing vertices;
3. probability of edge to vertex v is proportional to v's degree.

Interpretation high-degree vertices are more *popular*.

Experimental evidence for power-law distribution with $\gamma \approx 3$.

Preferential attachment: motivation and history.

Empirical observation: often a few nodes have very high degrees; degree-distribution resembles a *power-law*:

$$P(d(v) = k) \approx c \cdot \frac{1}{k^{\gamma}}$$

Model idea (Barabási and Albert, 1999):

1. vertices are successively added to the network;
2. new vertices create a fixed number of edges to already existing vertices;
3. probability of edge to vertex v is proportional to v's degree.

Interpretation high-degree vertices are more *popular*.

Experimental evidence for power-law distribution with $\gamma \approx 3$.

Preferential attachment model.

Definition (Bollobás, Riordan, Spencer, and Tusnády)

Directed multi-graphs, including loops, with $n \geq 1$ vertices and constant outdegree equal to $b \geq 1$.

Iterative definition:

start with empty graph $G = (V, E)$, $V = E = \emptyset$

foreach $v = 0, \ldots, n-1$ **do**
- put v into V
- **foreach** $j = 0, \ldots, b-1$ **do**
  - attach an outgoing edge $e = (v, \cdot)$ to v;
  - randomly select target w of e with probability

$$\frac{d_G(w)}{\sum_{w' \in V} d_G(w')} \; ;$$

  - put $e = (v, w)$ into E;

Preferential attachment model.

Definition (Bollobás, Riordan, Spencer, and Tusnády)

Directed multi-graphs, including loops, with $n \geq 1$ vertices and constant outdegree equal to $b \geq 1$.

Iterative definition:

start with empty graph $G = (V, E)$, $V = E = \emptyset$

foreach $v = 0, \ldots, n-1$ **do**
- put v into V
- **foreach** $j = 0, \ldots, b-1$ **do**
  - attach an outgoing edge $e = (v, \cdot)$ to v;
  - randomly select target w of e with probability

$$\frac{d_G(w)}{\sum_{w' \in V} d_G(w')} \; ;$$

  - put $e = (v, w)$ into E;

Preferential attachment (algorithm).

uses: uniform random sampling of integer from $\{0, \ldots, k\}$

input : number of nodes $n \in \mathbb{N}_{\geq 1}$, out-degree $b \in \mathbb{N}_{\geq 1}$
data : array $A[0 \ldots 2nb-1]$ *//collects endpoints of edges*
output multi-graph $G = (\{0, \ldots, n-1\}, E)$
:
$E \leftarrow \emptyset$; $m \leftarrow 0$ *//edge set and edge counter*

foreach $v = 0, \ldots, n-1$ **do**
  foreach $j = 0, \ldots, b-1$ **do**
    $A[2m] \leftarrow v$ *//v is source of next edge*
    $w \leftarrow A[\text{random}(\{0, \ldots, 2m\})]$ *//randomly select target*
    $A[2m+1] \leftarrow w$; *//put target in A*
    $E \leftarrow E \cup \{(v, w)\}$; $m \leftarrow m+1$ *//update edges*

Note: number of occurences of v in A equals degree of v
$\Rightarrow$ correct probability in selecting targets.

Preferential attachment (algorithm).

uses: uniform random sampling of integer from $\{0, \ldots, k\}$

input : number of nodes $n \in \mathbb{N}_{\geq 1}$, out-degree $b \in \mathbb{N}_{\geq 1}$
data : array $A[0 \ldots 2nb-1]$ *//collects endpoints of edges*
output multi-graph $G = (\{0, \ldots, n-1\}, E)$
:

$E \leftarrow \emptyset$; $m \leftarrow 0$ *//edge set and edge counter*

foreach $v = 0, \ldots, n-1$ **do**
  foreach $j = 0, \ldots, b-1$ **do**
    $A[2m] \leftarrow v$ *//v is source of next edge*
    $w \leftarrow A[\text{random}(\{0, \ldots, 2m\})]$ *//randomly select target*
    $A[2m+1] \leftarrow w$; *//put target in A*
    $E \leftarrow E \cup \{(v, w)\}$; $m \leftarrow m+1$ *//update edges*

Note: number of occurences of v in A equals degree of v
$\Rightarrow$ correct probability in selecting targets.

Preferential attachment (algorithm).

uses: uniform random sampling of integer from $\{0, \ldots, k\}$

input : number of nodes $n \in \mathbb{N}_{\geq 1}$, out-degree $b \in \mathbb{N}_{\geq 1}$
data : array $A[0 \ldots 2nb-1]$ *//collects endpoints of edges*
output multi-graph $G = (\{0, \ldots, n-1\}, E)$
:
$E \leftarrow \emptyset$; $m \leftarrow 0$ *//edge set and edge counter*

foreach $v = 0, \ldots, n-1$ **do**
  foreach $j = 0, \ldots, b-1$ **do**
    $A[2m] \leftarrow v$ *//v is source of next edge*
    $w \leftarrow A[\text{random}(\{0, \ldots, 2m\})]$ *//randomly select target*
    $A[2m+1] \leftarrow w$; *//put target in A*
    $E \leftarrow E \cup \{(v, w)\}$; $m \leftarrow m+1$ *//update edges*

Note: number of occurences of v in A equals degree of v
$\Rightarrow$ correct probability in selecting targets.

Preferential attachment (algorithm).

uses: uniform random sampling of integer from $\{0, \ldots, k\}$

input : number of nodes $n \in \mathbb{N}_{\geq 1}$, out-degree $b \in \mathbb{N}_{\geq 1}$
data : array $A[0 \ldots 2nb-1]$ *//collects endpoints of edges*
output multi-graph $G = (\{0, \ldots, n-1\}, E)$
:
$E \leftarrow \emptyset$; $m \leftarrow 0$ *//edge set and edge counter*

foreach $v = 0, \ldots, n-1$ **do**
  foreach $j = 0, \ldots, b-1$ **do**
    $A[2m] \leftarrow v$ *//v is source of next edge*
    $w \leftarrow A[\text{random}(\{0, \ldots, 2m\})]$ *//randomly select target*
    $A[2m+1] \leftarrow w$; *//put target in A*
    $E \leftarrow E \cup \{(v, w)\}$; $m \leftarrow m+1$ *//update edges*

Note: number of occurences of v in A equals degree of v
$\Rightarrow$ correct probability in selecting targets.

Preferential attachment (algorithm).

uses: uniform random sampling of integer from $\{0, \ldots, k\}$

input : number of nodes $n \in \mathbb{N}_{\geq 1}$, out-degree $b \in \mathbb{N}_{\geq 1}$
data : array $A[0 \ldots 2nb-1]$ *//collects endpoints of edges*
output multi-graph $G = (\{0, \ldots, n-1\}, E)$
:
$E \leftarrow \emptyset$; $m \leftarrow 0$ *//edge set and edge counter*

foreach $v = 0, \ldots, n-1$ **do**
  foreach $j = 0, \ldots, b-1$ **do**
    $A[2m] \leftarrow v$ *//v is source of next edge*
    $w \leftarrow A[\text{random}(\{0, \ldots, 2m\})]$ *//randomly select target*
    $A[2m+1] \leftarrow w$; *//put target in A*
    $E \leftarrow E \cup \{(v, w)\}$; $m \leftarrow m+1$ *//update edges*

Note: number of occurences of v in A equals degree of v
$\Rightarrow$ correct probability in selecting targets.

Preferential attachment (algorithm).

uses: uniform random sampling of integer from $\{0, \ldots, k\}$

input : number of nodes $n \in \mathbb{N}_{\geq 1}$, out-degree $b \in \mathbb{N}_{\geq 1}$
data : array $A[0 \ldots 2nb-1]$ *//collects endpoints of edges*
output multi-graph $G = (\{0, \ldots, n-1\}, E)$
:
$E \leftarrow \emptyset$; $m \leftarrow 0$ *//edge set and edge counter*

foreach $v = 0, \ldots, n-1$ **do**
  foreach $j = 0, \ldots, b-1$ **do**
    $A[2m] \leftarrow v$ *//v is source of next edge*
    $w \leftarrow A[\text{random}(\{0, \ldots, 2m\})]$ *//randomly select target*
    $A[2m+1] \leftarrow w$; *//put target in A*
    $E \leftarrow E \cup \{(v, w)\}$; $m \leftarrow m+1$ *//update edges*

Note: number of occurences of v in A equals degree of v
$\Rightarrow$ correct probability in selecting targets.

Preferential attachment (algorithm).

uses: uniform random sampling of integer from $\{0, \ldots, k\}$

input : number of nodes $n \in \mathbb{N}_{\geq 1}$, out-degree $b \in \mathbb{N}_{\geq 1}$
data : array $A[0 \ldots 2nb-1]$ *//collects endpoints of edges*
output multi-graph $G = (\{0, \ldots, n-1\}, E)$
:
$E \leftarrow \emptyset$; $m \leftarrow 0$ *//edge set and edge counter*

foreach $v = 0, \ldots, n-1$ **do**
  foreach $j = 0, \ldots, b-1$ **do**
    $A[2m] \leftarrow v$ *//v is source of next edge*
    $w \leftarrow A[\text{random}(\{0, \ldots, 2m\})]$ *//randomly select target*
    $A[2m+1] \leftarrow w$; *//put target in A*
    $E \leftarrow E \cup \{(v, w)\}$; $m \leftarrow m+1$ *//update edges*

Note: number of occurences of v in A equals degree of v
$\Rightarrow$ correct probability in selecting targets.

Preferential attachment (algorithm).

uses: uniform random sampling of integer from $\{0, \ldots, k\}$

input : number of nodes $n \in \mathbb{N}_{\geq 1}$, out-degree $b \in \mathbb{N}_{\geq 1}$
data : array $A[0 \ldots 2nb-1]$ *//collects endpoints of edges*
output multi-graph $G = (\{0, \ldots, n-1\}, E)$
:
$E \leftarrow \emptyset$; $m \leftarrow 0$ *//edge set and edge counter*

foreach $v = 0, \ldots, n-1$ **do**
  foreach $j = 0, \ldots, b-1$ **do**
    $A[2m] \leftarrow v$ *//v is source of next edge*
    $w \leftarrow A[\text{random}(\{0, \ldots, 2m\})]$ *//randomly select target*
    $A[2m+1] \leftarrow w$; *//put target in A*
    $E \leftarrow E \cup \{(v, w)\}$; $m \leftarrow m+1$ *//update edges*

Note: number of occurences of v in A equals degree of v
$\Rightarrow$ correct probability in selecting targets.

Preferential attachment leads to power law for low degree vertices.

Theorem (Bollobás, Riordan, Spencer, and Tusnády)

For $a, b \in \mathbb{N}$ let δ be defined by

$$\delta(a,b) = \frac{2b(b+1)}{(a+b)(a+b+1)(a+b+2)} \quad .$$

For $n, a, b \in \mathbb{N}$ with $0 \leq a \leq n^{\frac{1}{15}}$ and $\varepsilon \in \mathbb{R}_{>0}$ it holds in the preferential attachment model that

$$P\left[(1-\varepsilon)\cdot\delta(a,b) \leq \frac{|\{v \in V_n : d_G^-(v) = a\}|}{n} \leq (1+\varepsilon)\cdot\delta(a,b)\right] \xrightarrow[n\to\infty]{} 1$$

Outline.

Outline.

Exponential random graph models (informal).

Exponential random graph models (ERGMs) are a **class** of random graph models.

Concrete ERG-model is specified by two components:

1. A set of **network characteristics** (*statistics*) that determine the probability of a graph.
2. A set of **parameters** (associated with statistics) that determine which network properties increase or decrease the probabilities of graphs.

Choice of statistics often motivated by social science theory.

Parameters can be fitted to an observed network $\Rightarrow$ hypothesis testing.

Exponential random graph models (informal).

Exponential random graph models (ERGMs) are a **class** of random graph models.

Concrete ERG-model is specified by two components:

1. A set of **network characteristics** (*statistics*) that determine the probability of a graph.
2. A set of **parameters** (associated with statistics) that determine which network properties increase or decrease the probabilities of graphs.

Choice of statistics often motivated by social science theory.

Parameters can be fitted to an observed network $\Rightarrow$ hypothesis testing.

Exponential random graph models (informal).

Exponential random graph models (ERGMs) are a **class** of random graph models.

Concrete ERG-model is specified by two components:

1. A set of **network characteristics** (*statistics*) that determine the probability of a graph.
2. A set of **parameters** (associated with statistics) that determine which network properties increase or decrease the probabilities of graphs.

Choice of statistics often motivated by social science theory.

Parameters can be fitted to an observed network $\Rightarrow$ hypothesis testing.

Exponential random graph models (informal).

Exponential random graph models (ERGMs) are a **class** of random graph models.

Concrete ERG-model is specified by two components:

1. A set of **network characteristics** (*statistics*) that determine the probability of a graph.
2. A set of **parameters** (associated with statistics) that determine which network properties increase or decrease the probabilities of graphs.

Choice of statistics often motivated by social science theory.

Parameters can be fitted to an observed network $\Rightarrow$ hypothesis testing.

Exponential random graph models (ERGM).

Definition

The *ERGM class* consists of random graph models $(\mathcal{G}, P_\theta)$ whose probability can be written as

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

with

- $g_i : \mathcal{G} \to \mathbb{R}$ for $i = 1, \ldots, k$ (*statistics*);
- $\theta_i \in \mathbb{R}$ for $i = 1, \ldots, k$ (*parameters*); $\theta = (\theta_1, \ldots, \theta_k)$;
- *normalizing constant* κ defined by

$$\kappa(\theta) = \sum_{G' \in \mathcal{G}} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G')\right) \quad .$$

ERGM: commonly used statistics $g_i(G)$.

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

- number of edges (models density of graphs);
- number of edges connecting actors of the same type, e. g., gender (models homophily / heterophily);
- number of triangles (models tendency for / against transitive closure "*the friend of a friend is a friend*");
- number of k-stars (models tendency for / against connecting to high degree vertices);
- only in directed networks: number of reciprocated edges (models tendency for / against reciprocity).

ERGM: typical parameter signs.

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

Parameters θ associated with

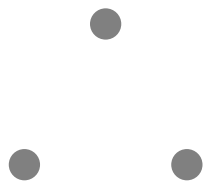
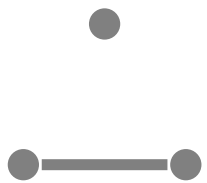
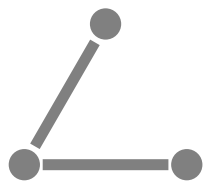
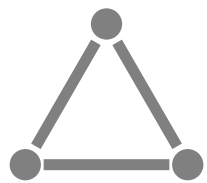
- number of edges: often **negative** (sparsity of networks; edges cost some effort);
- number of edges connecting actors of the same type: often **positive** (homophily);
- number of triangles: often **positive** ("*the friend of a friend is a friend*");
- number of k-stars: often **positive** (preferential attachment);
- number of reciprocated edges: often **positive.**

Note that adding / deleting a single edge can change the values of several statistics.

ERGM (example).

Consider undirected graphs with 3 vertices.

$$P(G) = \frac{1}{\kappa} \exp\left[-\log(2) \cdot m(G) + \log(16) \cdot \text{triangles}(G)\right]$$

				
$m(G)$	0	1	2	3
$\text{triangles}(G)$	0	0	0	1
$P(G) \cdot \kappa$	1	$\frac{1}{2}$	$\frac{1}{2^2}$	$\frac{16}{2^3}$
# isomorphic graphs	1	3	3	1

$$\Rightarrow \kappa = 1 + 3 \cdot 1/2 + 3 \cdot 1/4 + 2 = 21/4$$

ERGM: relation between statistics and probability.

Probability function

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right) \quad .$$

Isolating the effect of one specific statistic g_{i_0}:

$$P_\theta(G) = \exp[\theta_{i_0} \cdot g_{i_0}(G)] \cdot \frac{1}{\kappa(\theta)} \exp\left(\sum_{i \neq i_0} \theta_i \cdot g_i(G)\right) \quad .$$

$\Rightarrow$ if $g_{i_0}(G') = g_{i_0}(G) + c$, then $P(G') = \exp(\theta_{i_0})^c \cdot P(G)$
(assuming that $g_i(G') = g_i(G)$ for all $i \neq i_0$).

It is $\exp(\theta_{i_0}) > 1 \Leftrightarrow \theta_{i_0} > 0$ and $\exp(\theta_{i_0}) < 1 \Leftrightarrow \theta_{i_0} < 0$.

ERGM: relation between statistics and probability.

Probability function

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right) \quad .$$

Isolating the effect of one specific statistic g_{i_0}:

$$P_\theta(G) = \exp[\theta_{i_0} \cdot g_{i_0}(G)] \cdot \frac{1}{\kappa(\theta)} \exp\left(\sum_{i \neq i_0} \theta_i \cdot g_i(G)\right) \quad .$$

$\Rightarrow$ if $g_{i_0}(G') = g_{i_0}(G) + c$, then $P(G') = \exp(\theta_{i_0})^c \cdot P(G)$ (assuming that $g_i(G') = g_i(G)$ for all $i \neq i_0$).

It is $\exp(\theta_{i_0}) > 1 \Leftrightarrow \theta_{i_0} > 0$ and $\exp(\theta_{i_0}) < 1 \Leftrightarrow \theta_{i_0} < 0$.

ERGM: relation between statistics and probability.

Probability function

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right) \quad .$$

Isolating the effect of one specific statistic g_{i_0}:

$$P_\theta(G) = \exp[\theta_{i_0} \cdot g_{i_0}(G)] \cdot \frac{1}{\kappa(\theta)} \exp\left(\sum_{i \neq i_0} \theta_i \cdot g_i(G)\right) \quad .$$

$\Rightarrow$ if $g_{i_0}(G') = g_{i_0}(G) + c$, then $P(G') = \exp(\theta_{i_0})^c \cdot P(G)$
(assuming that $g_i(G') = g_i(G)$ for all $i \neq i_0$).

It is $\exp(\theta_{i_0}) > 1 \Leftrightarrow \theta_{i_0} > 0$ and $\exp(\theta_{i_0}) < 1 \Leftrightarrow \theta_{i_0} < 0$.

ERGM: relation between statistics and probability.

Probability function

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right) \quad .$$

Isolating the effect of one specific statistic g_{i_0}:

$$P_\theta(G) = \exp[\theta_{i_0} \cdot g_{i_0}(G)] \cdot \frac{1}{\kappa(\theta)} \exp\left(\sum_{i \neq i_0} \theta_i \cdot g_i(G)\right) \quad .$$

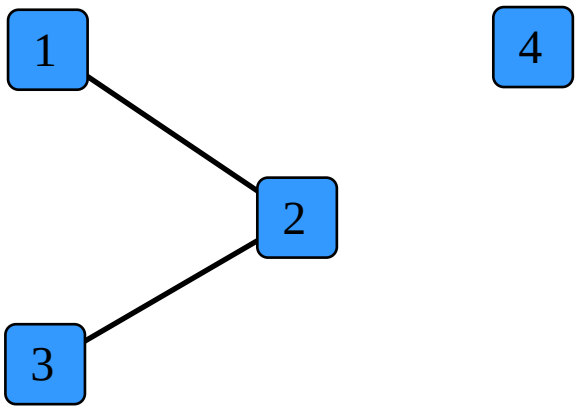
$\Rightarrow$ if $g_{i_0}(G') = g_{i_0}(G) + c$, then $P(G') = \exp(\theta_{i_0})^c \cdot P(G)$
(assuming that $g_i(G') = g_i(G)$ for all $i \neq i_0$).

It is $\exp(\theta_{i_0}) > 1 \Leftrightarrow \theta_{i_0} > 0$ and $\exp(\theta_{i_0}) < 1 \Leftrightarrow \theta_{i_0} < 0$.

Relation between statistics and probability (example).

Let g_{i_0} count the number of triangles in G.

$$P_\theta(G) = \exp[\theta_{i_0} \cdot g_{i_0}(G)] \cdot \frac{1}{\kappa(\theta)} \exp\left(\sum_{i \neq i_0} \theta_i \cdot g_i(G)\right) \quad .$$



Edge between 1 and 3 is $\exp(\theta_{i_0})$-times as likely as between 1 and 4.

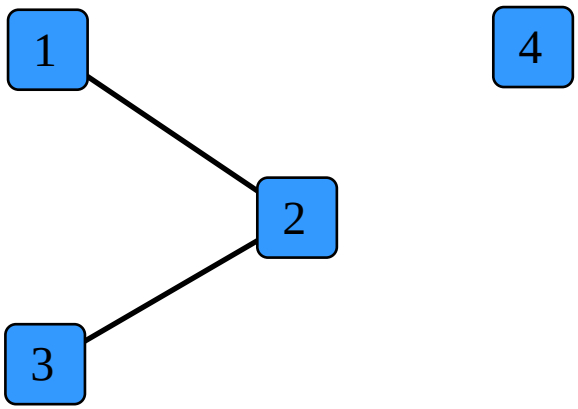
If other statistics change identically!

Positive $\theta_{i_0} \Rightarrow$ more likely; negative $\theta_{i_0} \Rightarrow$ less likely.

Relation between statistics and probability (example).

Let g_{i_0} count the number of triangles in G.

$$P_\theta(G) = \exp[\theta_{i_0} \cdot g_{i_0}(G)] \cdot \frac{1}{\kappa(\theta)} \exp\left(\sum_{i \neq i_0} \theta_i \cdot g_i(G)\right) \quad .$$



Edge between 1 and 3 is $\exp(\theta_{i_0})$-times as likely as between 1 and 4.

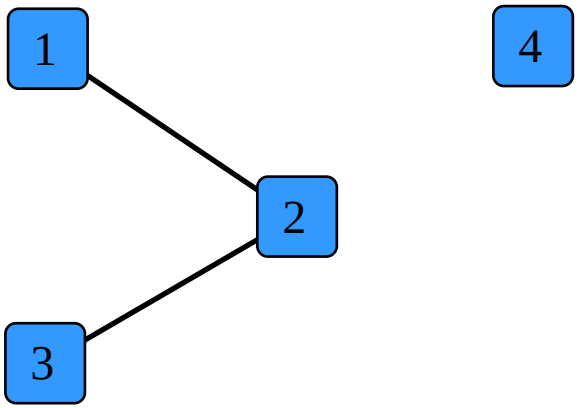
If other statistics change identically!

Positive $\theta_{i_0} \Rightarrow$ more likely; negative $\theta_{i_0} \Rightarrow$ less likely.

Relation between statistics and probability (example).

Let g_{i_0} count the number of triangles in G.

$$P_\theta(G) = \exp[\theta_{i_0} \cdot g_{i_0}(G)] \cdot \frac{1}{\kappa(\theta)} \exp\left(\sum_{i \neq i_0} \theta_i \cdot g_i(G)\right) \quad .$$



Edge between 1 and 3 is $\exp(\theta_{i_0})$-times as likely as between 1 and 4.

If other statistics change identically!

Positive $\theta_{i_0} \Rightarrow$ more likely; negative $\theta_{i_0} \Rightarrow$ less likely.

Remark.

In this lecture we consider only ERGMs $(\mathcal{G}, P)$ where $\mathcal{G}$ is the set of all undirected, loopless graphs with vertex set $V = \{1, \ldots, n\}$.

Example: $\mathcal{G}(n,p)$ belongs to the ERGM class.

Lemma

If $p \neq 0, 1$, then $\mathcal{G}(n,p)$ equals the ERGM defined by

$$P_1(G) = \exp\left[\theta \cdot m(G)\right] \cdot \kappa(\theta)^{-1}$$

where $\theta = \log\left(\frac{p}{1-p}\right)$ and $m(G)$ is the number of edges.

Proof.

Let P_2 denote the probability function of the $\mathcal{G}(n,p)$ model.

$$\begin{aligned} P_2(G) &= p^{m(G)} \cdot (1-p)^{\binom{n}{2} - m(G)} \\ &= \left(\frac{p}{1-p}\right)^{m(G)} \cdot (1-p)^{\binom{n}{2}} \\ &= \exp\left[\theta \cdot m(G)\right] \cdot (1-p)^{\binom{n}{2}} \end{aligned}$$

Thus, $\frac{P_1(G)}{P_1(G')} = \frac{P_2(G)}{P_2(G')}$ for any two graphs G, G'. $\square$

Example: $\mathcal{G}(n,p)$ belongs to the ERGM class.

Lemma

If $p \neq 0, 1$, then $\mathcal{G}(n,p)$ equals the ERGM defined by

$$P_1(G) = \exp\left[\theta \cdot m(G)\right] \cdot \kappa(\theta)^{-1}$$

where $\theta = \log\left(\frac{p}{1-p}\right)$ and $m(G)$ is the number of edges.

Proof.

Let P_2 denote the probability function of the $\mathcal{G}(n,p)$ model.

$$\begin{aligned} P_2(G) &= p^{m(G)} \cdot (1-p)^{\binom{n}{2} - m(G)} \\ &= \left(\frac{p}{1-p}\right)^{m(G)} \cdot (1-p)^{\binom{n}{2}} \\ &= \exp\left[\theta \cdot m(G)\right] \cdot (1-p)^{\binom{n}{2}} \end{aligned}$$

Thus, $\frac{P_1(G)}{P_1(G')} = \frac{P_2(G)}{P_2(G')}$ for any two graphs G, G'. □

Example: $\mathcal{G}(n,p)$ belongs to the ERGM class.

Lemma

If $p \neq 0, 1$, then $\mathcal{G}(n,p)$ equals the ERGM defined by

$$P_1(G) = \exp\left[\theta \cdot m(G)\right] \cdot \kappa(\theta)^{-1}$$

where $\theta = \log\left(\frac{p}{1-p}\right)$ and $m(G)$ is the number of edges.

Proof.

Let P_2 denote the probability function of the $\mathcal{G}(n,p)$ model.

$$\begin{aligned} P_2(G) &= p^{m(G)} \cdot (1-p)^{\binom{n}{2} - m(G)} \\ &= \left(\frac{p}{1-p}\right)^{m(G)} \cdot (1-p)^{\binom{n}{2}} \\ &= \exp\left[\theta \cdot m(G)\right] \cdot (1-p)^{\binom{n}{2}} \end{aligned}$$

Thus, $\frac{P_1(G)}{P_1(G')} = \frac{P_2(G)}{P_2(G')}$ for any two graphs G, G'. □

Example: $\mathcal{G}(n,p)$ belongs to the ERGM class.

Lemma

If $p \neq 0, 1$, then $\mathcal{G}(n,p)$ equals the ERGM defined by

$$P_1(G) = \exp\left[\theta \cdot m(G)\right] \cdot \kappa(\theta)^{-1}$$

where $\theta = \log\left(\frac{p}{1-p}\right)$ and $m(G)$ is the number of edges.

Proof.

Let P_2 denote the probability function of the $\mathcal{G}(n,p)$ model.

$$\begin{aligned} P_2(G) &= p^{m(G)} \cdot (1-p)^{\binom{n}{2} - m(G)} \\ &= \left(\frac{p}{1-p}\right)^{m(G)} \cdot (1-p)^{\binom{n}{2}} \\ &= \exp\left[\theta \cdot m(G)\right] \cdot (1-p)^{\binom{n}{2}} \end{aligned}$$

Thus, $\frac{P_1(G)}{P_1(G')} = \frac{P_2(G)}{P_2(G')}$ for any two graphs G, G'. □

Example: $\mathcal{G}(n,p)$ belongs to the ERGM class.

Lemma

If $p \neq 0, 1$, then $\mathcal{G}(n,p)$ equals the ERGM defined by

$$P_1(G) = \exp\left[\theta \cdot m(G)\right] \cdot \kappa(\theta)^{-1}$$

where $\theta = \log\left(\frac{p}{1-p}\right)$ and $m(G)$ is the number of edges.

Proof.

Let P_2 denote the probability function of the $\mathcal{G}(n,p)$ model.

$$\begin{aligned} P_2(G) &= p^{m(G)} \cdot (1-p)^{\binom{n}{2} - m(G)} \\ &= \left(\frac{p}{1-p}\right)^{m(G)} \cdot (1-p)^{\binom{n}{2}} \\ &= \exp\left[\theta \cdot m(G)\right] \cdot (1-p)^{\binom{n}{2}} \end{aligned}$$

Thus, $\frac{P_1(G)}{P_1(G')} = \frac{P_2(G)}{P_2(G')}$ for any two graphs G, G'. □

Example: $\mathcal{G}(n,p)$ belongs to the ERGM class.

Proof.
$P_1(G) = P_2(G)\frac{P_1(G')}{P_2(G')}$ for any two graphs G, G' implies that for arbitrary but fixed G' it is

$$\begin{aligned} \sum_{G\in\mathcal{G}} P_1(G) &= \sum_{G\in\mathcal{G}} P_2(G)\cdot\frac{P_1(G')}{P_2(G')} \\ 1 &= 1\cdot\frac{P_1(G')}{P_2(G')} \end{aligned}$$

Hence, $P_1(G') = P_2(G')$. □

Interpretation of $\theta = \log\left(\frac{p}{1-p}\right)$.

Lemma

If $p \neq 0, 1$, then $\mathcal{G}(n, p)$ equals the ERGM defined by

$$P_1(G) = \exp\left[\theta \cdot m(G)\right] \cdot \kappa(\theta)^{-1}$$

where $\theta = \log\left(\frac{p}{1-p}\right)$ and $m(G)$ is the number of edges.

Connection between θ and p

- $\theta < 0 \iff$ expected density $p < 1/2$;
- $\theta = 0 \iff$ expected density $p = 1/2$;
- $\theta > 0 \iff$ expected density $p > 1/2$.

Does not hold in general (if the ERGM contains other statistics).

Interpretation of $\theta = \log\left(\frac{p}{1-p}\right)$.

Lemma

If $p \neq 0, 1$, then $\mathcal{G}(n, p)$ equals the ERGM defined by

$$P_1(G) = \exp\left[\theta \cdot m(G)\right] \cdot \kappa(\theta)^{-1}$$

where $\theta = \log\left(\frac{p}{1-p}\right)$ and $m(G)$ is the number of edges.

Connection between θ and p

- $\theta < 0 \iff$ expected density $p < 1/2$;
- $\theta = 0 \iff$ expected density $p = 1/2$;
- $\theta > 0 \iff$ expected density $p > 1/2$.

Does not hold in general (if the ERGM contains other statistics).

Interpretation of $\theta = \log\left(\frac{p}{1-p}\right)$.

Lemma

If $p \neq 0, 1$, then $\mathcal{G}(n, p)$ equals the ERGM defined by

$$P_1(G) = \exp\left[\theta \cdot m(G)\right] \cdot \kappa(\theta)^{-1}$$

where $\theta = \log\left(\frac{p}{1-p}\right)$ and $m(G)$ is the number of edges.

Connection between θ and p

- $\theta < 0 \iff$ expected density $p < 1/2$;
- $\theta = 0 \iff$ expected density $p = 1/2$;
- $\theta > 0 \iff$ expected density $p > 1/2$.

Does not hold in general (if the ERGM contains other statistics).

Commonly used network statistics (I).

Statistics g_i counting specific subgraphs (*configurations*).

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

If a subgraph count is associated with a positive (negative) parameter, then those subgraphs are more (less) likely.

Example

- $m(G)$ defined as the number of **edges**

models preference for edges over non-edges or vice versa.

Commonly used network statistics (I).

Statistics g_i counting specific subgraphs (*configurations*).

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

If a subgraph count is associated with a positive (negative) parameter, then those subgraphs are more (less) likely.

Example

- $m(G)$ defined as the number of **edges**

models preference for edges over non-edges or vice versa.

Commonly used network statistics (I).

Statistics g_i counting specific subgraphs (*configurations*).

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

If a subgraph count is associated with a positive (negative) parameter, then those subgraphs are more (less) likely.

Example

- $m(G)$ defined as the number of **edges**

models preference for edges over non-edges or vice versa.

Commonly used network statistics (II).

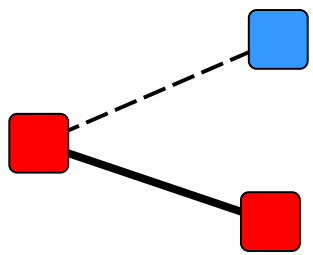
$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

Example

assuming that actors have attribute values $a\colon V \to \{1, \ldots, c\}$, such as age, gender, ...

- $m_a(G) = |\{\{u, v\} \in E\,;\ a(u) = a(v)\}|$, i. e., the number of edges connecting actors with the same attribute value

models tendency for (against) creating edges to similar actors
homophily (heterophily)



Commonly used network statistics (III).

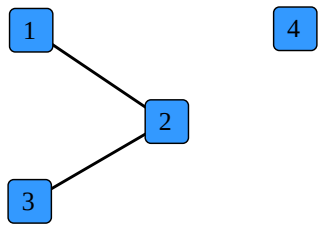
Statistics g_i counting specific subgraphs (*configurations*).

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

Example

- $t(G)$ defined as the number of **triangles**

models preference (reluctance) to close triangles (transitivity).



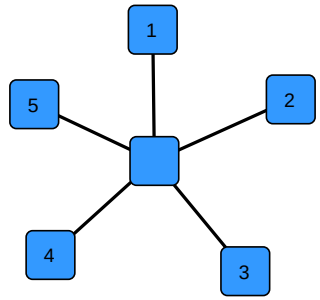
Commonly used network statistics (IV).

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

Example

- $s_\ell(G)$ defined as the number of ℓ**-stars**, $\ell = 2, \ldots, n-1$

models tendency for (against) connecting to high-degree vertices.



Note: a vertex of degree d contributes $\binom{d}{\ell}$ to the ℓ-star count.

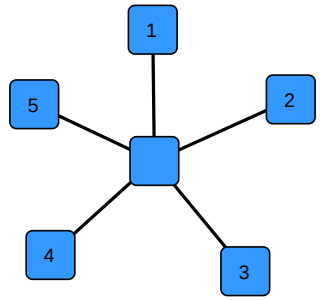
Commonly used network statistics (IV).

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

Example

- $s_\ell(G)$ defined as the number of ℓ**-stars**, $\ell = 2, \ldots, n-1$

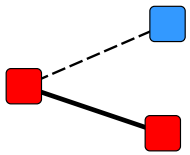
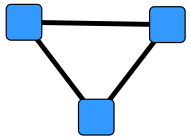
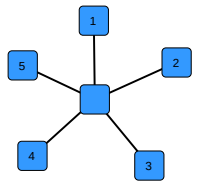
models tendency for (against) connecting to high-degree vertices.



Note: a vertex of degree d contributes $\binom{d}{\ell}$ to the ℓ-star count.

Implication on dyad dependency.

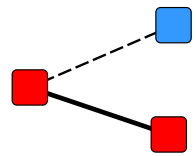
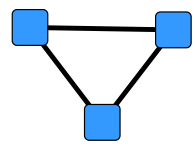
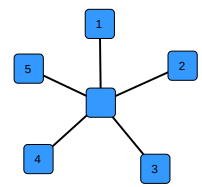
$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

number of edges	
edges connecting same attribute	
number of triangles	
number of ℓ-stars	

Using some of these statistics make edge probabilities **dependent**.

Implication on dyad dependency.

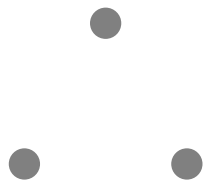
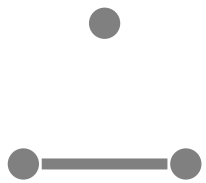
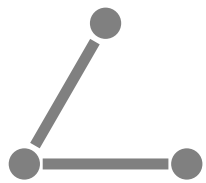
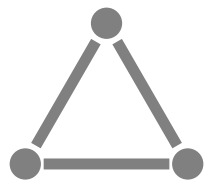
$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

$g_i(G)$		**edge prob.**
number of edges		independent
edges connecting same attribute		independent
number of triangles		dependent
number of ℓ-stars		dependent

Edge dependency (example).

Consider undirected graphs with 3 vertices; 2-star count s_2.

$$P(G) = \frac{1}{\kappa} \exp\left[\log(2) \cdot s_2(G)\right]$$

				
$s_2(G)$	0	0	1	3
$P(G) \cdot \kappa$	1	1	2	$2^3 = 8$
# isomorphic graphs	1	3	3	1

Let e, e' be two different dyads.

$$\begin{aligned} P(\mathcal{G}_e | \mathcal{G}_{e'}) &= (2+8)/(1+2\cdot 2+8) = 10/13 \\ P(\mathcal{G}_e) &= (1+2\cdot 2+8)/(1+3\cdot 1+3\cdot 2+8) = 13/18 \end{aligned}$$

Thus, dyads e and e' are statistically dependent.

Outline.

Sampling from an ERGM: why is this difficult?

We want to design a probabilistic algorithm that

- returns at each call a graph G from $\mathcal{G}$;
- with probability equal to $P(G)$.

So far, algorithms for sampling from $\mathcal{G}(n,p)$ or the preferential attachment model decided about the inclusion/exclusion of edges one after the other.

(In general) this is not feasible for ERGMs

- we cannot compute edge probabilities

$$P(e \in E) = P(\mathcal{G}_e) = P(\{G \in \mathcal{G}\,;\ e \in E_G\})$$

  in an efficient way;
- we cannot even compute $P(G)$ for a single graph G since the normalizing constant has $2^{\binom{n}{2}}$ terms.

Sampling from an ERGM: why is this difficult?

We want to design a probabilistic algorithm that

- returns at each call a graph G from $\mathcal{G}$;
- with probability equal to $P(G)$.

So far, algorithms for sampling from $\mathcal{G}(n,p)$ or the preferential attachment model decided about the inclusion/exclusion of edges one after the other.

(In general) this is not feasible for ERGMs

- we cannot compute edge probabilities

$$P(e \in E) = P(\mathcal{G}_e) = P(\{G \in \mathcal{G}\,;\ e \in E_G\})$$

  in an efficient way;
- we cannot even compute $P(G)$ for a single graph G since the normalizing constant has $2^{\binom{n}{2}}$ terms.

Sampling from an ERGM: why is this difficult?

We want to design a probabilistic algorithm that

- returns at each call a graph G from $\mathcal{G}$;
- with probability equal to $P(G)$.

So far, algorithms for sampling from $\mathcal{G}(n,p)$ or the preferential attachment model decided about the inclusion/exclusion of edges one after the other.

(In general) this is not feasible for ERGMs

- we cannot compute edge probabilities

$$P(e \in E) = P(\mathcal{G}_e) = P(\{G \in \mathcal{G}\,;\ e \in E_G\})$$

  in an efficient way;
- we cannot even compute $P(G)$ for a single graph G since the normalizing constant has $2^{\binom{n}{2}}$ terms.

Markov chain simulation (informal).

A Markov chain consists of a set of states and transition probabilities to jump from one state to another.

Here, given an ERGM $(\mathcal{G}, P)$

- the set of states is $\mathcal{G}$ (all graphs);
- transition probabilities π are a function of P
- in such a way that
  - the probability to **be on a graph** G converges to $P(G)$, when the number of simulation steps tends to ∞.

$\Rightarrow$ Simulate many steps and return the current graph.

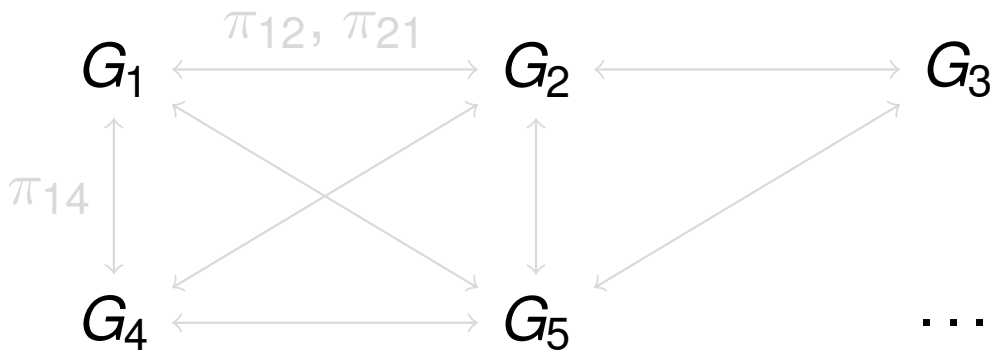
Markov chain simulation (informal).

A Markov chain consists of a set of states and transition probabilities to jump from one state to another.

Here, given an ERGM $(\mathcal{G}, P)$

- the set of states is $\mathcal{G}$ (all graphs);
- transition probabilities π are a function of P
- in such a way that
  - the probability to **be on a graph** G converges to $P(G)$, when the number of simulation steps tends to ∞.

$\Rightarrow$ Simulate many steps and return the current graph.



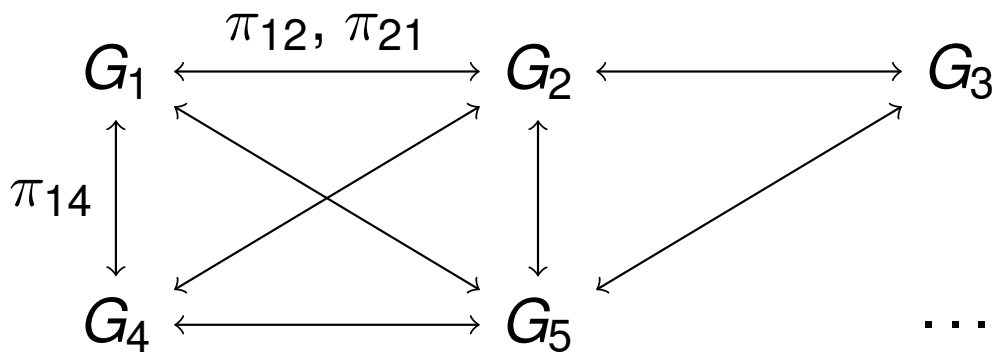
Markov chain simulation (informal).

A Markov chain consists of a set of states and transition probabilities to jump from one state to another.

Here, given an ERGM $(\mathcal{G}, P)$

- the set of states is $\mathcal{G}$ (all graphs);
- transition probabilities π are a function of P
- in such a way that
  - the probability to **be on a graph** G converges to $P(G)$, when the number of simulation steps tends to ∞.

$\Rightarrow$ Simulate many steps and return the current graph.



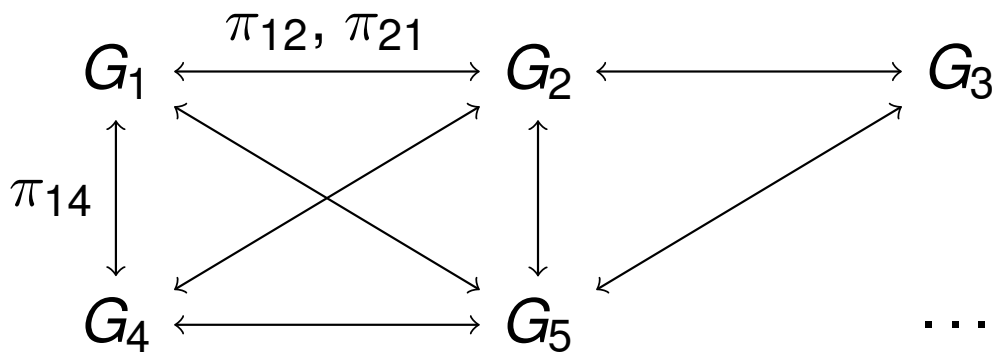
Markov chain simulation (informal).

A Markov chain consists of a set of states and transition probabilities to jump from one state to another.

Here, given an ERGM $(\mathcal{G}, P)$

- the set of states is $\mathcal{G}$ (all graphs);
- transition probabilities π are a function of P
- in such a way that
  - the probability to **be on a graph** G converges to $P(G)$, when the number of simulation steps tends to ∞.

$\Rightarrow$ Simulate many steps and return the current graph.



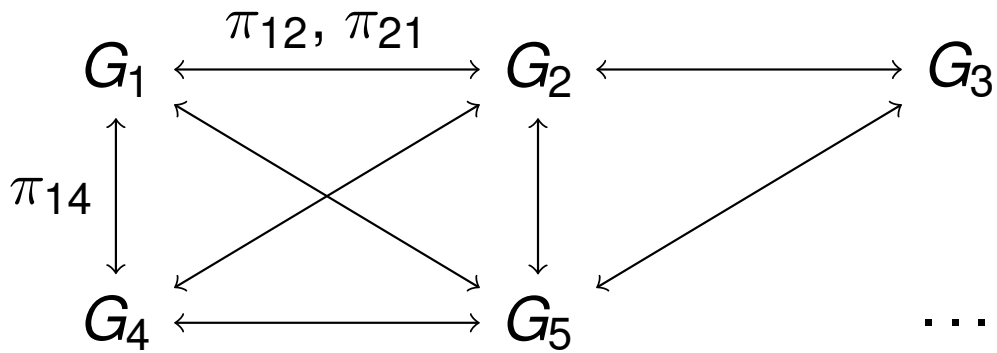
Markov chain simulation (informal).

A Markov chain consists of a set of states and transition probabilities to jump from one state to another.

Here, given an ERGM $(\mathcal{G}, P)$

- the set of states is $\mathcal{G}$ (all graphs);
- transition probabilities π are a function of P
- in such a way that
  - the probability to **be on a graph** G converges to $P(G)$, when the number of simulation steps tends to ∞.

$\Rightarrow$ Simulate many steps and return the current graph.



Finite stationary Markov chain (simplified definition).

Note: Markov chains are usually defined as *random processes* that satisfy certain properties. The following is a more intuitive definition for stationary Markov chains.

Definition

A *(finite stationary) Markov chain* is a pair $(\mathcal{G}, \pi)$, where

- $\mathcal{G}$ is a finite set $\mathcal{G} = \{G_1, \ldots, G_N\}$ (*state space*);
- π is a matrix $\pi \in \mathbb{R}^{N \times N}$ (*transition matrix*) satisfying
  - for all i, j it is $\pi_{ij} \in [0, 1]$;
  - for all i it is $\sum_{j=1}^{N} \pi_{ij} = 1$.

π_{ij} interpreted as the probability to jump from state G_i to G_j.

Finite stationary Markov chain (simplified definition).

Note: Markov chains are usually defined as *random processes* that satisfy certain properties. The following is a more intuitive definition for stationary Markov chains.

Definition

A *(finite stationary) Markov chain* is a pair $(\mathcal{G}, \pi)$, where

- $\mathcal{G}$ is a finite set $\mathcal{G} = \{G_1, \ldots, G_N\}$ (*state space*);
- π is a matrix $\pi \in \mathbb{R}^{N \times N}$ (*transition matrix*) satisfying
  - for all i, j it is $\pi_{ij} \in [0, 1]$;
  - for all i it is $\sum_{j=1}^{N} \pi_{ij} = 1$.

π_{ij} interpreted as the probability to jump from state G_i to G_j.

Finite stationary Markov chain (simplified definition).

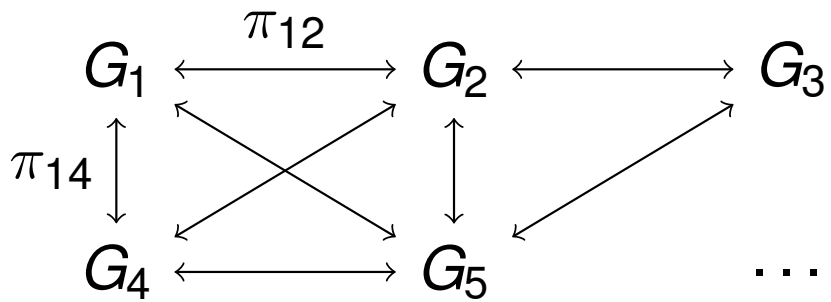
Note: Markov chains are usually defined as *random processes* that satisfy certain properties. The following is a more intuitive definition for stationary Markov chains.

Definition

A *(finite stationary) Markov chain* is a pair $(\mathcal{G}, \pi)$, where

- $\mathcal{G}$ is a finite set $\mathcal{G} = \{G_1, \ldots, G_N\}$ (*state space*);
- π is a matrix $\pi \in \mathbb{R}^{N \times N}$ (*transition matrix*) satisfying
  - for all i, j it is $\pi_{ij} \in [0, 1]$;
  - for all i it is $\sum_{j=1}^{N} \pi_{ij} = 1$.

π_{ij} interpreted as the probability to jump from state G_i to G_j.



Background: eigenvectors and eigenvalues.

Let $A \in \mathbb{R}^{n \times n}$ be a matrix and $x \in \mathbb{C}^n$ be a vector.

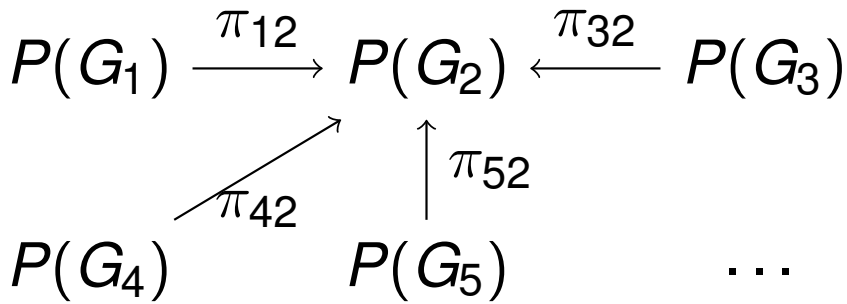
If there is a $\lambda \in \mathbb{C}$ such that

$$A \cdot x = \lambda \cdot x \ ,$$

then x is called an *eigenvector* of A and (if $x \neq 0$) λ is called an *eigenvalue* of A.

Stationary state space distributions.

A *Markov chain* is a pair $(\mathcal{G}, \pi)$, where π_{ij} is the probability to jump from state G_i to state G_j.



A probability distribution P on $\mathcal{G}$ is called *stationary* if for all j it is $P(G_j) = \sum_{i=1}^{N} P(G_i)\pi_{ij}$.

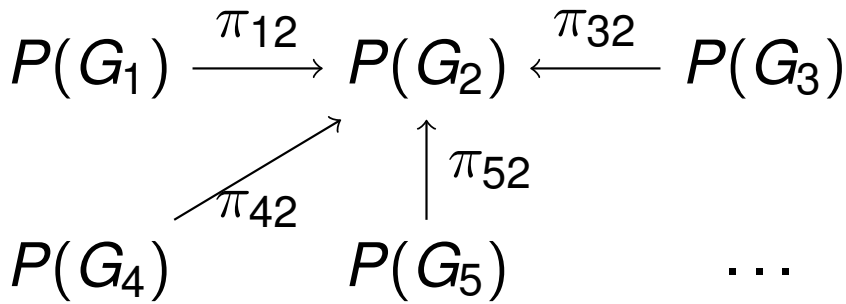
Satisfied if and only if (with $P = [P(G_1) \dots P(G_N)] \in \mathbb{R}^N$ written as a row vector) it is

$$P = P\pi \ ,$$

i. e., P is an eigenvector of π with eigenvalue one.

Stationary state space distributions.

A *Markov chain* is a pair $(\mathcal{G}, \pi)$, where π_{ij} is the probability to jump from state G_i to state G_j.



A probability distribution P on $\mathcal{G}$ is called *stationary* if for all j it is $P(G_j) = \sum_{i=1}^{N} P(G_i)\pi_{ij}$.

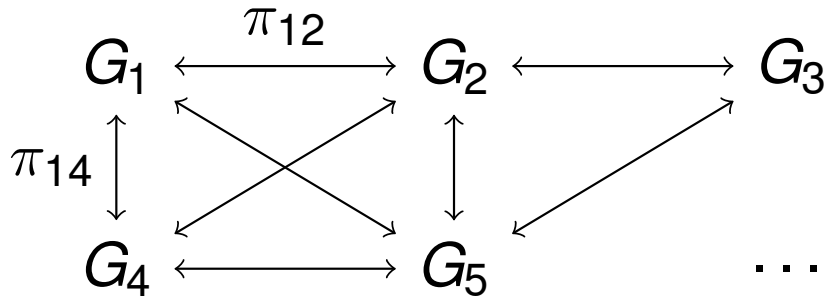
Satisfied if and only if (with $P = [P(G_1) \ldots P(G_N)] \in \mathbb{R}^N$ written as a row vector) it is

$$P = P\pi \quad ,$$

i. e., P is an eigenvector of π with eigenvalue one.

Graph associated with a Markov chain.

π_{ij} interpreted as the probability to jump from state G_i to G_j.



A Markov chain $(\mathcal{G}, \pi)$ is associated with a directed graph $(\mathcal{G}, E)$ whose vertices are the states and where

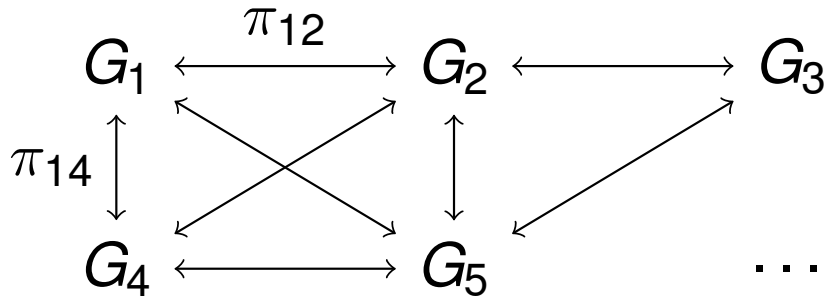
$$(G_i, G_j) \in E \iff \pi_{ij} > 0 \ .$$

The Markov chain $(\mathcal{G}, \pi)$ is called

- *irreducible* if $(\mathcal{G}, E)$ is strongly connected;
- *aperiodic* if the greatest common divisor of all cycles in $(\mathcal{G}, E)$ equals one.

Graph associated with a Markov chain.

π_{ij} interpreted as the probability to jump from state G_i to G_j.



A Markov chain $(\mathcal{G}, \pi)$ is associated with a directed graph $(\mathcal{G}, E)$ whose vertices are the states and where

$$(G_i, G_j) \in E \iff \pi_{ij} > 0 \ .$$

The Markov chain $(\mathcal{G}, \pi)$ is called

- *irreducible* if $(\mathcal{G}, E)$ is strongly connected;
- *aperiodic* if the greatest common divisor of all cycles in $(\mathcal{G}, E)$ equals one.

Stationary distribution of reversible Markov chains.

Theorem

If a probability distribution P on $\mathcal{G}$ satisfies for all graphs G_i, G_j

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji}$$

(Markov chain is then called reversible)
and the Markov chain is irreducible and aperiodic
then P is the unique stationary distribution of the Markov chain
and for any initial distribution P' it is

$$\lim_{K\to\infty} P'\pi^K = P \ .$$

Reversibility condition used to find an appropriate π if P is given.

Stationary distribution of reversible Markov chains.

Theorem

If a probability distribution P on $\mathcal{G}$ satisfies for all graphs G_i, G_j

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji}$$

(Markov chain is then called reversible)
and the Markov chain is irreducible and aperiodic
then P is the unique stationary distribution of the Markov chain
and for any initial distribution P' it is

$$\lim_{K\to\infty} P'\pi^K = P \ .$$

Reversibility condition used to find an appropriate π if P is given.

Stationary distribution of reversible Markov chains.

Theorem

If a probability distribution P on $\mathcal{G}$ satisfies for all graphs G_i, G_j

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji}$$

(Markov chain is then called reversible)
and the Markov chain is irreducible and aperiodic
then P is the unique stationary distribution of the Markov chain
and for any initial distribution P' it is

$$\lim_{K\to\infty} P'\pi^K = P \ .$$

Reversibility condition used to find an appropriate π if P is given.

Stationary distribution of reversible Markov chains.

Theorem

If a probability distribution P on $\mathcal{G}$ satisfies for all graphs G_i, G_j

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji}$$

(Markov chain is then called reversible)
and the Markov chain is irreducible and aperiodic
then P is the unique stationary distribution of the Markov chain
and for any initial distribution P' it is

$$\lim_{K\to\infty} P'\pi^K = P \ .$$

Reversibility condition used to find an appropriate π if P is given.

Stationary distribution of reversible Markov chains.

Theorem

If a probability distribution P on $\mathcal{G}$ satisfies for all graphs G_i, G_j

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji}$$

(Markov chain is then called reversible)
and the Markov chain is irreducible and aperiodic
then P is the unique stationary distribution of the Markov chain
and for any initial distribution P' it is

$$\lim_{K\to\infty} P'\pi^K = P \ .$$

Reversibility condition used to find an appropriate π if P is given.

Background: Perron-Frobenius Theorem.

spectral radius $\rho(A) = \max\{|\lambda| \,;\ \lambda \text{ is eigenvalue of } A\}$

Theorem (Perron-Frobenius)

The spectral-radius $\rho(A)$ *of a non-negative, irreducible matrix A is an eigenvalue of multiplicity one and all entries of an associated eigenvector are non-zero and have the same sign.*

Background: power iteration.

Theorem (power iteration)

Let A be a non-negative, irreducible, aperiodic matrix and x a normalized eigenvector with associated eigenvalue $\rho(A)$. For a vector $y^{(0)}$ whose projection onto x is not zero define a sequence of vectors by

$$y^{(i+1)} = \frac{A \cdot y^{(i)}}{\|A \cdot y^{(i)}\|} \ .$$

Then $\lim_{i\to\infty} y^{(i)} = x$.

Stationary distribution of reversible Markov chains.

Theorem

If a probability distribution P on $\mathcal{G}$ satisfies for all graphs G_i, G_j

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji}$$

and the Markov chain $(\mathcal{G}, \pi)$ is irreducible and aperiodic then P is the unique stationary distribution of $(\mathcal{G}, \pi)$ and for any initial distribution P' it is

$$\lim_{K\to\infty} P'\pi^K = P \ .$$

Proof.

Matrix π satisfies the conditions of the theorems on the previous slides. We show that

- P is an eigenvector of π with eigenvalue one;
- the spectral radius ρ of π is one.

Stationary distribution of reversible Markov chains.

Theorem

If a probability distribution P on $\mathcal{G}$ satisfies for all graphs G_i, G_j

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji}$$

and the Markov chain $(\mathcal{G}, \pi)$ is irreducible and aperiodic then P is the unique stationary distribution of $(\mathcal{G}, \pi)$ and for any initial distribution P' it is

$$\lim_{K\to\infty} P'\pi^K = P \ .$$

Proof.

Matrix π satisfies the conditions of the theorems on the previous slides. We show that

- P is an eigenvector of π with eigenvalue one;
- the spectral radius ρ of π is one.

Stationary distribution of reversible Markov chains.

Theorem

If a probability distribution P on $\mathcal{G}$ satisfies for all graphs G_i, G_j

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji}$$

and the Markov chain $(\mathcal{G}, \pi)$ is irreducible and aperiodic then P is the unique stationary distribution of $(\mathcal{G}, \pi)$ and for any initial distribution P' it is

$$\lim_{K\to\infty} P'\pi^K = P \ .$$

Proof.

Matrix π satisfies the conditions of the theorems on the previous slides. We show that

- P is an eigenvector of π with eigenvalue one;
- the spectral radius ρ of π is one.

Stationary distribution of reversible Markov chains.

From

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji}$$

it follows that for all $G_i \in \mathcal{G}$ it is

$$\sum_{G_j \in \mathcal{G}} P(G_j)\pi_{ji} = \sum_{G_j \in \mathcal{G}} P(G_i)\pi_{ij} = P(G_i)$$

(since the rows of π sum up to one).

Thus, P and π satisfy the matrix equation $P\pi = P$, i. e., P is an eigenvector of π with eigenvalue one.

Stationary distribution of reversible Markov chains.

From

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji}$$

it follows that for all $G_i \in \mathcal{G}$ it is

$$\sum_{G_j \in \mathcal{G}} P(G_j)\pi_{ji} = \sum_{G_j \in \mathcal{G}} P(G_i)\pi_{ij} = P(G_i)$$

(since the rows of π sum up to one).

Thus, P and π satisfy the matrix equation $P\pi = P$, i. e., P is an eigenvector of π with eigenvalue one.

Stationary distribution of reversible Markov chains.

Show that the spectral radius $\rho = \rho(\pi)$ of π is equal to one.

Proof.

We have that for all i it is $\sum_{j=1}^{N} \pi_{ij} = 1$.

Let x be an eigenvector of π with eigenvalue ρ.

For all j it is $\rho x_j = \sum_{i=1}^{N} x_i \pi_{ij}$. Thus

$$\rho \sum_{j=1}^{N} x_j = \sum_{j=1}^{N} \rho x_j = \sum_{j=1}^{N} \sum_{i=1}^{N} x_i \pi_{ij}$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} x_i \pi_{ij} = \sum_{i=1}^{N} x_i \sum_{j=1}^{N} \pi_{ij} = \sum_{i=1}^{N} x_i$$

Since $\sum_{i=1}^{N} x_i \neq 0$ (why?) it must be $\rho = 1$. □

Stationary distribution of reversible Markov chains.

Show that the spectral radius $\rho = \rho(\pi)$ of π is equal to one.

Proof.

We have that for all i it is $\sum_{j=1}^{N} \pi_{ij} = 1$.

Let x be an eigenvector of π with eigenvalue ρ.

For all j it is $\rho x_j = \sum_{i=1}^{N} x_i \pi_{ij}$. Thus

$$\begin{aligned} \rho \sum_{j=1}^{N} x_j &= \sum_{j=1}^{N} \rho x_j = \sum_{j=1}^{N} \sum_{i=1}^{N} x_i \pi_{ij} \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} x_i \pi_{ij} = \sum_{i=1}^{N} x_i \sum_{j=1}^{N} \pi_{ij} = \sum_{i=1}^{N} x_i \end{aligned}$$

Since $\sum_{i=1}^{N} x_i \neq 0$ (why?) it must be $\rho = 1$. □

Stationary distribution of reversible Markov chains.

Show that the spectral radius $\rho = \rho(\pi)$ of π is equal to one.

Proof.

We have that for all i it is $\sum_{j=1}^{N} \pi_{ij} = 1$.

Let x be an eigenvector of π with eigenvalue ρ.

For all j it is $\rho x_j = \sum_{i=1}^{N} x_i \pi_{ij}$. Thus

$$\rho \sum_{j=1}^{N} x_j = \sum_{j=1}^{N} \rho x_j = \sum_{j=1}^{N} \sum_{i=1}^{N} x_i \pi_{ij}$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} x_i \pi_{ij} = \sum_{i=1}^{N} x_i \sum_{j=1}^{N} \pi_{ij} = \sum_{i=1}^{N} x_i$$

Since $\sum_{i=1}^{N} x_i \neq 0$ (why?) it must be $\rho = 1$. □

Stationary distribution of reversible Markov chains.

Show that the spectral radius $\rho = \rho(\pi)$ of π is equal to one.

Proof.

We have that for all i it is $\sum_{j=1}^{N} \pi_{ij} = 1$.

Let x be an eigenvector of π with eigenvalue ρ.

For all j it is $\rho x_j = \sum_{i=1}^{N} x_i \pi_{ij}$. Thus

$$\begin{aligned} \rho \sum_{j=1}^{N} x_j &= \sum_{j=1}^{N} \rho x_j = \sum_{j=1}^{N} \sum_{i=1}^{N} x_i \pi_{ij} \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} x_i \pi_{ij} = \sum_{i=1}^{N} x_i \sum_{j=1}^{N} \pi_{ij} = \sum_{i=1}^{N} x_i \end{aligned}$$

Since $\sum_{i=1}^{N} x_i \neq 0$ (why?) it must be $\rho = 1$. □

Stationary distribution of reversible Markov chains.

Show that the spectral radius $\rho = \rho(\pi)$ of π is equal to one.

Proof.

We have that for all i it is $\sum_{j=1}^{N} \pi_{ij} = 1$.

Let x be an eigenvector of π with eigenvalue ρ.

For all j it is $\rho x_j = \sum_{i=1}^{N} x_i \pi_{ij}$. Thus

$$\begin{aligned} \rho \sum_{j=1}^{N} x_j &= \sum_{j=1}^{N} \rho x_j = \sum_{j=1}^{N} \sum_{i=1}^{N} x_i \pi_{ij} \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} x_i \pi_{ij} = \sum_{i=1}^{N} x_i \sum_{j=1}^{N} \pi_{ij} = \sum_{i=1}^{N} x_i \end{aligned}$$

Since $\sum_{i=1}^{N} x_i \neq 0$ (why?) it must be $\rho = 1$. □

Stationary distribution of reversible Markov chains.

Theorem

If a probability distribution P on $\mathcal{G}$ satisfies for all graphs G_i, G_j

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji}$$

(Markov chain is called reversible)
and the Markov chain is irreducible and aperiodic
then P is the unique stationary distribution of the Markov chain
and for any initial distribution P' it is

$$\lim_{K\to\infty} P'\pi^K = P \ .$$

Gibbs sampling.

Given P, define π such that

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji} \ .$$

Gibbs sampling: define π as follows

- $\pi_{ij} = 0$ if G_i and G_j differ in more than one dyad;
- if G_i and G_j differ in exactly one dyad, then

$$\pi_{ij} = \frac{P(G_j)}{\binom{n}{2}(P(G_i) + P(G_j))} \ .$$

- $\pi_{ii} = \sum \frac{P(G_i)}{\binom{n}{2}(P(G_i)+P(G))}$

  (sum over all G that differ from G_i in exactly one dyad)

Show: π is normalized, irreducible, aperiodic, reversible.

Gibbs sampling.

Given P, define π such that

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji} \ .$$

Gibbs sampling: define π as follows

- $\pi_{ij} = 0$ if G_i and G_j differ in more than one dyad;
- if G_i and G_j differ in exactly one dyad, then

$$\pi_{ij} = \frac{P(G_j)}{\binom{n}{2}(P(G_i) + P(G_j))} \ .$$

- $\pi_{ii} = \sum \frac{P(G_i)}{\binom{n}{2}(P(G_i)+P(G))}$

  (sum over all G that differ from G_i in exactly one dyad)

Show: π is normalized, irreducible, aperiodic, reversible.

Gibbs sampling.

Given P, define π such that

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji} \ .$$

Gibbs sampling: define π as follows

- $\pi_{ij} = 0$ if G_i and G_j differ in more than one dyad;
- if G_i and G_j differ in exactly one dyad, then

$$\pi_{ij} = \frac{P(G_j)}{\binom{n}{2}(P(G_i) + P(G_j))} \ .$$

- $\pi_{ii} = \sum \frac{P(G_i)}{\binom{n}{2}(P(G_i)+P(G))}$

  (sum over all G that differ from G_i in exactly one dyad)

Show: π is normalized, irreducible, aperiodic, reversible.

Gibbs sampling.

Given P, define π such that

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji} \ .$$

Gibbs sampling: define π as follows

- $\pi_{ij} = 0$ if G_i and G_j differ in more than one dyad;
- if G_i and G_j differ in exactly one dyad, then

$$\pi_{ij} \quad = \quad \frac{P(G_j)}{\binom{n}{2}(P(G_i) + P(G_j))} \ .$$

- $\pi_{ii} = \sum \frac{P(G_i)}{\binom{n}{2}(P(G_i)+P(G))}$

  (sum over all G that differ from G_i in exactly one dyad)

Show: π is normalized, irreducible, aperiodic, reversible.

Gibbs sampling.

Given P, define π such that

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji} \ .$$

Gibbs sampling: define π as follows

- $\pi_{ij} = 0$ if G_i and G_j differ in more than one dyad;
- if G_i and G_j differ in exactly one dyad, then

$$\pi_{ij} \quad = \quad \frac{P(G_j)}{\binom{n}{2}(P(G_i) + P(G_j))} \ .$$

- $\pi_{ii} = \sum \frac{P(G_i)}{\binom{n}{2}(P(G_i)+P(G))}$

  (sum over all G that differ from G_i in exactly one dyad)

Show: π is normalized, irreducible, aperiodic, reversible.

Gibbs sampling.

Given P, define π such that

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji} \ .$$

Gibbs sampling: define π as follows

- $\pi_{ij} = 0$ if G_i and G_j differ in more than one dyad;
- if G_i and G_j differ in exactly one dyad, then

$$\pi_{ij} \quad = \quad \frac{P(G_j)}{\binom{n}{2}(P(G_i) + P(G_j))} \ .$$

- $\pi_{ii} = \sum \frac{P(G_i)}{\binom{n}{2}(P(G_i)+P(G))}$

  (sum over all G that differ from G_i in exactly one dyad)

Show: π is normalized, irreducible, aperiodic, reversible.

Gibbs sampling.

Transition probabilities defined by

$$\pi_{ij} = \frac{P(G_j)}{\binom{n}{2}(P(G_i) + P(G_j))} .$$

Didn't we claim that it is intractable to compute $P(G)$?

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

$$\kappa(\theta) = \sum_{G' \in \mathcal{G}} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G')\right)$$

Gibbs sampling.

Transition probabilities defined by

$$\pi_{ij} = \frac{P(G_j)}{\binom{n}{2}(P(G_i) + P(G_j))} .$$

Didn't we claim that it is intractable to compute $P(G)$?

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

$$\kappa(\theta) = \sum_{G' \in \mathcal{G}} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G')\right)$$

Gibbs sampling (algorithm).

initialize G by any graph from $\mathcal{G}$;
repeat *many* times

- select a dyad $\{i, j\}$ uniformly at random;
- with probability $\frac{P(G^{+ij})}{P(G^{+ij})+P(G^{-ij})}$
  - replace $G = (V, E)$ by $G^{+ij} = (V, E \cup \{i, j\})$
  - otherwise replace $G = (V, E)$ by $G^{-ij} = (V, E \setminus \{i, j\})$;

return G;

Note: (in practice) the statistics $g_\ell(G^{+ij})$ and $g_\ell(G^{-ij})$ can be efficiently derived by computing changes to the statistics $g_\ell(G)$.

Gibbs sampling (algorithm).

initialize G by any graph from $\mathcal{G}$;
repeat *many* times

- select a dyad $\{i, j\}$ uniformly at random;
- with probability $\frac{P(G^{+ij})}{P(G^{+ij})+P(G^{-ij})}$
  - replace $G = (V, E)$ by $G^{+ij} = (V, E \cup \{i, j\})$
  - otherwise replace $G = (V, E)$ by $G^{-ij} = (V, E \setminus \{i, j\})$;

return G;

Note: (in practice) the statistics $g_\ell(G^{+ij})$ and $g_\ell(G^{-ij})$ can be efficiently derived by computing changes to the statistics $g_\ell(G)$.

The ability to sample from an ERGM enables us to efficiently estimate quantities that are computationally intractable. For instance,

- the expected number of edges, triangles, ℓ-stars,. . . ;
- the normalizing constant κ;
- the probability $P(G)$ of a specific graph G.

Estimation of the expected number of edges.

Let $(\mathcal{G}, P)$ be an ERGM. By definition it is

$$\mathbb{E}(m) = \sum_{G \in \mathcal{G}} P(G) \cdot m(G) \ .$$

To approximate $\mathbb{E}(m)$ draw K random samples $G_1, \ldots, G_K$ from $(\mathcal{G}, P)$ and compute

$$\widehat{\mathbb{E}(m)} = \sum_{i=1}^{K} \frac{1}{K} \cdot m(G_i) \ .$$

$\widehat{\mathbb{E}(m)}$ converges to $\mathbb{E}(m)$ in probability when $K \to \infty$.

Outline.

Conditional independence of edges.

Let $(\mathcal{G}, P)$ be a random graph model where D is the set of dyads of graphs in $\mathcal{G}$ and assume that $P(G) > 0$ for all $G \in \mathcal{G}$. Let $d_1, d_2 \in D$ be two different dyads.

For a partition $D^+ \uplus D^- = D \setminus \{d_1, d_2\}$ of the set of dyads different from d_1 and d_2 let the subset $\mathcal{G}_{D^+ \uplus D^-}$ be defined by

$$\mathcal{G}_{D^+ \uplus D^-} = \{G \in \mathcal{G}\,;\ D^+ \subseteq E_G \text{ and } D^- \cap E_G = \emptyset\}\ .$$

We say that d_1 and d_2 are *conditionally independent* (given the rest of the graph) if for all partitions $D^+ \uplus D^- = D \setminus \{d_1, d_2\}$ it is

$$P(\mathcal{G}_{d_1} | \mathcal{G}_{D^+ \uplus D^-}) = P(\mathcal{G}_{d_1} | \mathcal{G}_{D^+ \uplus D^-} \cap \mathcal{G}_{d_2})\ .$$

Informally: if we know the state of all dyads in $D \setminus \{d_1, d_2\}$, the state of the dyad d_2 does not provide any additional information about the probability $P(d_1 \in E)$.

Conditional independence of edges.

Let $(\mathcal{G}, P)$ be a random graph model where D is the set of dyads of graphs in $\mathcal{G}$ and assume that $P(G) > 0$ for all $G \in \mathcal{G}$. Let $d_1, d_2 \in D$ be two different dyads.

For a partition $D^+ \uplus D^- = D \setminus \{d_1, d_2\}$ of the set of dyads different from d_1 and d_2 let the subset $\mathcal{G}_{D^+ \uplus D^-}$ be defined by

$$\mathcal{G}_{D^+ \uplus D^-} = \{G \in \mathcal{G}\,;\ D^+ \subseteq E_G \text{ and } D^- \cap E_G = \emptyset\}\ .$$

We say that d_1 and d_2 are *conditionally independent* (given the rest of the graph) if for all partitions $D^+ \uplus D^- = D \setminus \{d_1, d_2\}$ it is

$$P(\mathcal{G}_{d_1} | \mathcal{G}_{D^+ \uplus D^-}) = P(\mathcal{G}_{d_1} | \mathcal{G}_{D^+ \uplus D^-} \cap \mathcal{G}_{d_2})\ .$$

Informally: if we know the state of all dyads in $D \setminus \{d_1, d_2\}$, the state of the dyad d_2 does not provide any additional information about the probability $P(d_1 \in E)$.

Conditional independence of edges.

Let $(\mathcal{G}, P)$ be a random graph model where D is the set of dyads of graphs in $\mathcal{G}$ and assume that $P(G) > 0$ for all $G \in \mathcal{G}$. Let $d_1, d_2 \in D$ be two different dyads.

For a partition $D^+ \uplus D^- = D \setminus \{d_1, d_2\}$ of the set of dyads different from d_1 and d_2 let the subset $\mathcal{G}_{D^+ \uplus D^-}$ be defined by

$$\mathcal{G}_{D^+ \uplus D^-} = \{G \in \mathcal{G}\,;\ D^+ \subseteq E_G \text{ and } D^- \cap E_G = \emptyset\}\ .$$

We say that d_1 and d_2 are *conditionally independent* (given the rest of the graph) if for all partitions $D^+ \uplus D^- = D \setminus \{d_1, d_2\}$ it is

$$P(\mathcal{G}_{d_1} | \mathcal{G}_{D^+ \uplus D^-}) = P(\mathcal{G}_{d_1} | \mathcal{G}_{D^+ \uplus D^-} \cap \mathcal{G}_{d_2})\ .$$

Informally: if we know the state of all dyads in $D \setminus \{d_1, d_2\}$, the state of the dyad d_2 does not provide any additional information about the probability $P(d_1 \in E)$.

Conditional independence of edges.

Let $(\mathcal{G}, P)$ be a random graph model where D is the set of dyads of graphs in $\mathcal{G}$ and assume that $P(G) > 0$ for all $G \in \mathcal{G}$. Let $d_1, d_2 \in D$ be two different dyads.

For a partition $D^+ \uplus D^- = D \setminus \{d_1, d_2\}$ of the set of dyads different from d_1 and d_2 let the subset $\mathcal{G}_{D^+ \uplus D^-}$ be defined by

$$\mathcal{G}_{D^+ \uplus D^-} = \{G \in \mathcal{G}\,;\ D^+ \subseteq E_G \text{ and } D^- \cap E_G = \emptyset\}\ .$$

We say that d_1 and d_2 are *conditionally independent* (given the rest of the graph) if for all partitions $D^+ \uplus D^- = D \setminus \{d_1, d_2\}$ it is

$$P(\mathcal{G}_{d_1} | \mathcal{G}_{D^+ \uplus D^-}) = P(\mathcal{G}_{d_1} | \mathcal{G}_{D^+ \uplus D^-} \cap \mathcal{G}_{d_2})\ .$$

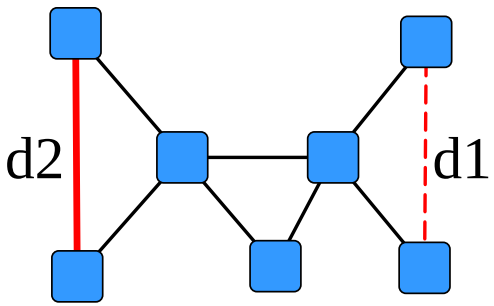
Informally: if we know the state of all dyads in $D \setminus \{d_1, d_2\}$, the state of the dyad d_2 does not provide any additional information about the probability $P(d_1 \in E)$.

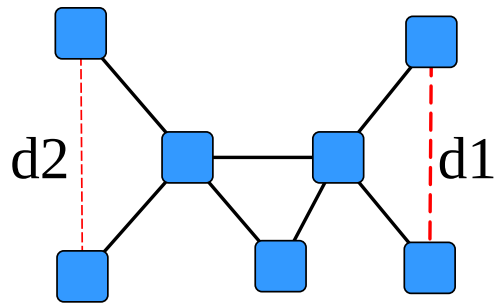
Conditional independence of edges.

We say that d_1 and d_2 are *conditionally independent* (given the rest of the graph) if for all partitions $D^+ \uplus D^- = D \setminus \{d_1, d_2\}$ it is

$$P(\mathcal{G}_{d_1} | \mathcal{G}_{D^+ \uplus D^-}) = P(\mathcal{G}_{d_1} | \mathcal{G}_{D^+ \uplus D^-} \cap \mathcal{G}_{d_2}) \ .$$

Informally: if we know the state of all dyads in $D \setminus \{d_1, d_2\}$, the state of the dyad d_2 does not provide any additional information about the probability $P(d_1 \in E)$.





Markov random graphs.

Definition

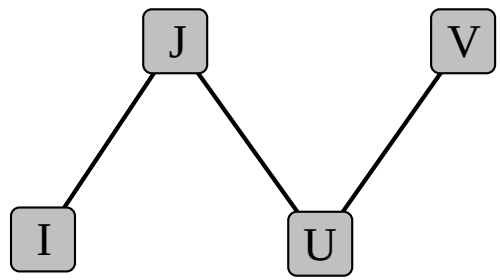
Markov random graphs are a class of random graph models satisfying (1) the probability of every graph is positive and (2) for every set of four pairwise different vertices $\{i, j, u, v\}$ the dyads $\{i, j\}$ and $\{u, v\}$ are conditionally independent, given the rest of the graph.

From another point of view, non-incident dyads are required to be conditionally independent.

Example

$\{i, j\}$ and $\{u, v\}$ conditionally independent;

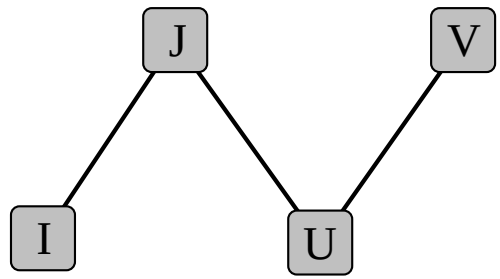
$\{i, j\}$ and $\{j, u\}$ might be conditionally dependent;



Markov random graphs (comments).

$\{i, j\}$ and $\{u, v\}$ conditionally independent;

$\{i, j\}$ and $\{j, u\}$ might be conditionally dependent;



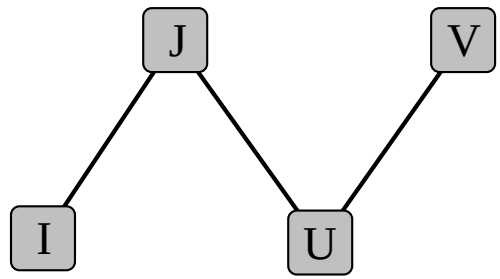
Note: $\{i, j\}$ and $\{u, v\}$ might be (unconditionally) dependent.

We'll see later: Markov graphs are a sub-class of ERGMs.

Markov random graphs (comments).

$\{i, j\}$ and $\{u, v\}$ conditionally independent;

$\{i, j\}$ and $\{j, u\}$ might be conditionally dependent;



Note: $\{i, j\}$ and $\{u, v\}$ might be (unconditionally) dependent.

We'll see later: Markov graphs are a sub-class of ERGMs.

Dependence graph (of a random graph model).

Definition

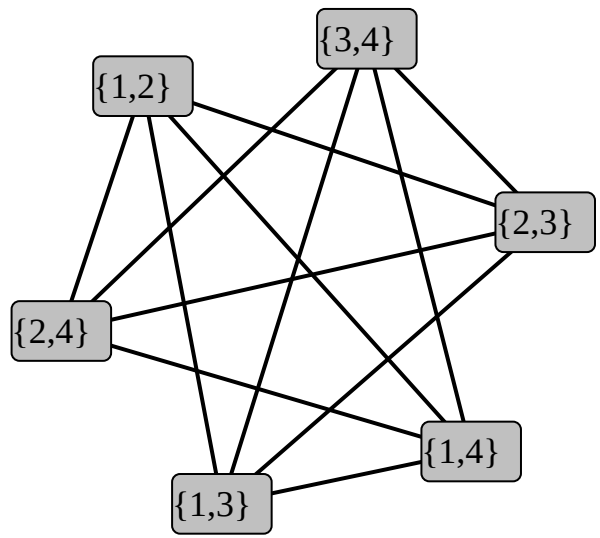
Let $(\mathcal{G}, P)$ be a random graph model and let D be the set of dyads of graphs in $\mathcal{G}$.

The *dependence graph* $\mathcal{D} = (D, E)$ of $(\mathcal{G}, P)$ has vertex set D,

$\{d_i, d_j\} \in E$ if d_i and d_j are conditionally dependent, given the rest of the graph.

Example

the dependence graph of a Markov graph on vertices $V = \{1, 2, 3, 4\}$ is



Hammersley-Clifford Theorem; special case.

Theorem (first part)

Let $(\mathcal{G}, P)$ be a random graph satisfying $P(G) > 0$ for all $G \in \mathcal{G}$, let D be the set of dyads and $\mathcal{D}$ the dependence graph.

There are constants $\{\alpha_A \in \mathbb{R};\ A \subseteq D\}$, satisfying $\alpha_A = 0$ if A is not a clique in $\mathcal{D}$, such that

$$P(G) = \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A\right), \quad \text{where} \tag{1}$$

$$\kappa = \sum_{G' \in \mathcal{G}} \exp\left(\sum_{A \subseteq E(G')} \alpha_A\right). \tag{2}$$

Note: A is a clique in $\mathcal{D}$ if dyads in A are pairwise dependent. $A \subseteq E(G)$ means that all dyads in A are edges in G.

Hammersley-Clifford Theorem; special case.

Theorem (first part)

Let $(\mathcal{G}, P)$ be a random graph satisfying $P(G) > 0$ for all $G \in \mathcal{G}$, let D be the set of dyads and $\mathcal{D}$ the dependence graph.

There are constants $\{\alpha_A \in \mathbb{R};\ A \subseteq D\}$, satisfying $\alpha_A = 0$ if A is not a clique in $\mathcal{D}$, such that

$$P(G) = \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A\right), \qquad \textit{where} \tag{1}$$

$$\kappa = \sum_{G' \in \mathcal{G}} \exp\left(\sum_{A \subseteq E(G')} \alpha_A\right). \tag{2}$$

Note: A is a clique in $\mathcal{D}$ if dyads in A are pairwise dependent. $A \subseteq E(G)$ means that all dyads in A are edges in G.

Hammersley-Clifford Theorem; special case.

Theorem (second part)

Conversely, if the probability P on $\mathcal{G}$ is defined by

$$P(G) = \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A\right), \qquad \textit{where}$$

$$\kappa = \sum_{G' \in \mathcal{G}} \exp\left(\sum_{A \subseteq E(G')} \alpha_A\right),$$

then two dyads d_1 and d_2 are conditionally independent in $(\mathcal{G}, P)$, unless there is a subset $A \subseteq D$ with $d_1, d_2 \in A$ and $\alpha_A \neq 0$.

Hammersley-Clifford Theorem; special case.

Theorem (second part)

Conversely, if the probability P on $\mathcal{G}$ is defined by

$$P(G) = \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A\right), \qquad \textit{where}$$

$$\kappa = \sum_{G' \in \mathcal{G}} \exp\left(\sum_{A \subseteq E(G')} \alpha_A\right),$$

then two dyads d_1 and d_2 are conditionally independent in $(\mathcal{G}, P)$, unless there is a subset $A \subseteq D$ with $d_1, d_2 \in A$ and $\alpha_A \neq 0$.

Conclusion from the Hammersley-Clifford Theorem.

There are constants $\{\alpha_A \in \mathbb{R}\,;\ A \subseteq D\}$, such that

$$P(G) \quad = \quad \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A\right) \quad .$$

Every random graph model $(\mathcal{G}, P)$ with $P > 0$ is an ERGM

- statistics: for $A \subseteq D$ define $g_A(G) = \begin{cases} 1 & A \subseteq E(G) \\ 0 & \text{else} \end{cases}$
- parameters: α_A

$$P(G) \quad = \quad \frac{1}{\kappa} \exp\left(\sum_{A \subseteq D} \alpha_A \cdot g_A(G)\right) \quad .$$

Conclusion from the Hammersley-Clifford Theorem.

There are constants $\{\alpha_A \in \mathbb{R}\,;\ A \subseteq D\}$, such that

$$P(G) \quad = \quad \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A\right) \quad .$$

Every random graph model $(\mathcal{G}, P)$ with $P > 0$ is an ERGM

- statistics: for $A \subseteq D$ define $g_A(G) = \begin{cases} 1 & A \subseteq E(G) \\ 0 & \text{else} \end{cases}$
- parameters: α_A

$$P(G) \quad = \quad \frac{1}{\kappa} \exp\left(\sum_{A \subseteq D} \alpha_A \cdot g_A(G)\right) \quad .$$

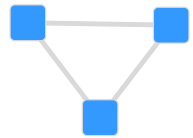
Cliques in the dependence graph of a Markov graph.

Markov random graphs: edges $\{i, j\}$ and $\{u, v\}$ are conditionally independent, unless they have a vertex in common.

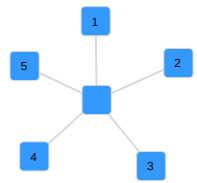
Cliques in the dependence graph of a Markov graph are

edges

triangles



ℓ-stars, for $\ell = 2, \ldots, n-1$

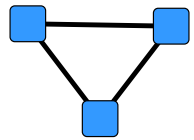
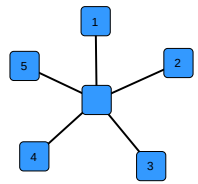


No other subgraphs are cliques in the dependence graph.

Cliques in the dependence graph of a Markov graph.

Markov random graphs: edges $\{i, j\}$ and $\{u, v\}$ are conditionally independent, unless they have a vertex in common.

Cliques in the dependence graph of a Markov graph are

edges	
triangles	
ℓ-stars, for $\ell = 2, \ldots, n-1$	

No other subgraphs are cliques in the dependence graph.

ERGM of general Markov graphs.

Corollary

Let $(\mathcal{G}, P)$ be a Markov random graph on vertices $V = \{1, \ldots, n\}$. Then there are real constants

$$\begin{aligned} \eta_{uv} \quad & \text{for all} \quad \text{dyads } \{u, v\} \\ \tau_{uvw} \quad & \text{for all} \quad \text{triangles } \{u, v, w\} \\ \sigma_{uv_1 \ldots v_\ell} \quad & \text{for all} \quad 2 \leq \ell \leq n-1, \text{ and all} \\ & \qquad \ell\text{-stars } (u, \{v_1, \ldots, v_\ell\}) \end{aligned}$$

such that the probability of a graph $G \in \mathcal{G}$ can be written as

$$P(G) = \frac{1}{\kappa} \exp\left(\sum_{uv \in E(G)} \eta_{uv} + \sum_{uvw \in T(G)} \tau_{uvw} + \sum_{\ell=2}^{n-1} \sum_{uv_1 \ldots v_\ell \in S_\ell(G)} \sigma_{uv_1 \ldots v_\ell} \right)$$

ERGM of general Markov graphs (remarks).

$$P(G) = \frac{1}{\kappa} \exp\left(\sum_{uv \in E(G)} \eta_{uv} + \sum_{uvw \in T(G)} \tau_{uvw} + \sum_{\ell=2}^{n-1} \sum_{uv_1 \ldots v_\ell \in S_\ell(G)} \sigma_{uv_1 \ldots v_\ell} \right)$$

Each dyad, triangle, ℓ-star can contribute differently to the probability of a graph.

$\Rightarrow$ unreasonably high number of parameters.

Homogeneous random graph model.

Two graphs $G = (V, E)$ and $H = (W, F)$ are called *isomorphic* if there is a bijection $\varphi\colon V \to W$ such that

$$\forall u, v \in V\colon \{u, v\} \in E \Leftrightarrow \{\varphi(u), \varphi(v)\} \in F \ .$$

Definition

A random graph model $(\mathcal{G}, P)$ is called *homogeneous* if for any pair of isomorphic graphs G and H it is $P(G) = P(H)$.

ERGM of homogeneous Markov graphs.

Corollary

Let $(\mathcal{G}, P)$ be a homogeneous Markov random graph. Then there are real constants η, τ, and σ_ℓ for $\ell = 2, \ldots, n-1$ such that the probability of a graph $G \in \mathcal{G}$ can be written as

$$P(G) = \frac{1}{\kappa} \exp\left(\eta \cdot m(G) + \tau \cdot t(G) + \sum_{\ell=2}^{n-1} \sigma_\ell \cdot s_\ell(G)\right)$$

Proof.

Show that any two edge parameters are equal...

For $\ell = 2, \ldots, n-1$, show that any two ℓ-star parameters are equal...

Show that any two triangle parameters are equal...

□

ERGM of homogeneous Markov graphs.

Corollary

Let $(\mathcal{G}, P)$ be a homogeneous Markov random graph. Then there are real constants η, τ, and σ_ℓ for $\ell = 2, \ldots, n-1$ such that the probability of a graph $G \in \mathcal{G}$ can be written as

$$P(G) = \frac{1}{\kappa} \exp\left(\eta \cdot m(G) + \tau \cdot t(G) + \sum_{\ell=2}^{n-1} \sigma_\ell \cdot s_\ell(G)\right)$$

Proof.

Show that any two edge parameters are equal...
For $\ell = 2, \ldots, n-1$, show that any two ℓ-star parameters are equal...
Show that any two triangle parameters are equal...

Example: dependence graph of $\mathcal{G}(n, p)$.

$\mathcal{G}(n, p)$ is a homogeneous random graph model whose dependence graph has no edges. Thus,

$$P(G) = \frac{1}{\kappa} \exp\left(\eta \cdot m(G)\right) \quad .$$

Hammersley-Clifford Theorem; special case.

proof

Theorem (first part)

Let $(\mathcal{G}, P)$ be a random graph satisfying $P(G) > 0$ for all $G \in \mathcal{G}$, let D be the set of dyads and $\mathcal{D}$ the dependence graph.

There are constants $\{\alpha_A \in \mathbb{R};\ A \subseteq D\}$, satisfying $\alpha_A = 0$ if A is not a clique in $\mathcal{D}$, such that

$$P(G) = \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A\right), \qquad \textit{where}$$

$$\kappa = \sum_{G' \in \mathcal{G}} \exp\left(\sum_{A \subseteq E(G')} \alpha_A\right).$$

Möbius Inversion Theorem.

Needed for the proof of the Hammersley-Clifford Theorem.

Let S be a finite set and

$$f\colon \mathcal{P}(S) \to \mathbb{R}; \qquad\qquad g\colon \mathcal{P}(S) \to \mathbb{R};$$

two functions defined on the set of subsets of S.

Then, for all subsets $A \subseteq S$ it is

$$f(A) = \sum_{B \subseteq A} g(B)$$

if and only if for all subsets $A \subseteq S$ it is

$$g(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} f(B) \ .$$

Möbius Inversion Theorem.

Needed for the proof of the Hammersley-Clifford Theorem.

Let S be a finite set and

$$f\colon \mathcal{P}(S) \to \mathbb{R}; \qquad\qquad g\colon \mathcal{P}(S) \to \mathbb{R};$$

two functions defined on the set of subsets of S.

Then, for all subsets $A \subseteq S$ it is

$$f(A) = \sum_{B \subseteq A} g(B)$$

if and only if for all subsets $A \subseteq S$ it is

$$g(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} f(B) \ .$$

Proof of the Hammersley-Clifford Theorem.

$$\text{Want to show } P(G) \quad = \quad \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A \right) \quad .$$

For a set $B \subseteq D$ define $G_B = (V, B) \in \mathcal{G}$ to be the graph whose edge set is equal to B.

Define $Q \colon \mathcal{P}(D) \to \mathbb{R}$ by setting for $A \subseteq D$

$$Q(G_A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \log P(G_B) \quad .$$

Motivation (Möbius Inversion Theorem)

$$f(A) = \sum_{B \subseteq A} g(B) \iff g(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} f(B) \quad .$$

Proof of the Hammersley-Clifford Theorem.

$$\text{Want to show } P(G) \quad = \quad \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A \right) \quad .$$

For a set $B \subseteq D$ define $G_B = (V, B) \in \mathcal{G}$ to be the graph whose edge set is equal to B.

Define $Q \colon \mathcal{P}(D) \to \mathbb{R}$ by setting for $A \subseteq D$

$$Q(G_A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \log P(G_B) \quad .$$

Motivation (Möbius Inversion Theorem)

$$f(A) = \sum_{B \subseteq A} g(B) \iff g(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} f(B) \quad .$$

Proof of the Hammersley-Clifford Theorem.

$$\text{Want to show } P(G) \quad = \quad \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A \right) \quad .$$

For a set $B \subseteq D$ define $G_B = (V, B) \in \mathcal{G}$ to be the graph whose edge set is equal to B.

Define $Q \colon \mathcal{P}(D) \to \mathbb{R}$ by setting for $A \subseteq D$

$$Q(G_A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \log P(G_B) \quad .$$

Motivation (Möbius Inversion Theorem)

$$f(A) = \sum_{B \subseteq A} g(B) \iff g(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} f(B) \quad .$$

Proof of the Hammersley-Clifford Theorem.

$$\text{Want to show } P(G) \quad = \quad \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A \right) \quad .$$

For a set $B \subseteq D$ define $G_B = (V, B) \in \mathcal{G}$ to be the graph whose edge set is equal to B.

Define $Q \colon \mathcal{P}(D) \to \mathbb{R}$ by setting for $A \subseteq D$

$$Q(G_A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \log P(G_B) \quad .$$

Motivation (Möbius Inversion Theorem)

$$f(A) = \sum_{B \subseteq A} g(B) \iff g(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} f(B) \quad .$$

Proof of the Hammersley-Clifford Theorem.

We have

$$Q(G_A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \log P(G_B) \ .$$

If follows from the Möbius inversion theorem that for $A \subseteq D$ it is

$$\log P(G_A) = \sum_{B \subseteq A} Q(G_B) \ .$$

In particular, for $A = E(G)$ we get

$$P(G) = \exp \left(\sum_{B \subseteq E(G)} Q(G_B) \right) \ .$$

Proof of the Hammersley-Clifford Theorem.

We have

$$P(G) = \exp\left(\sum_{A \subseteq E(G)} Q(G_A)\right) \quad .$$

Thus, $\alpha_A := Q(G_A)$ are the desired constants.

It remains to show that $\alpha_A = 0$ if A is not a clique in $\mathcal{D}$; and we are done with the first part of the theorem.

Proof of the Hammersley-Clifford Theorem.

We have

$$P(G) = \exp\left(\sum_{A \subseteq E(G)} Q(G_A)\right) \quad .$$

Thus, $\alpha_A := Q(G_A)$ are the desired constants.

It remains to show that $\alpha_A = 0$ if A is not a clique in $\mathcal{D}$; and we are done with the first part of the theorem.

Proof of the Hammersley-Clifford Theorem.

We have

$$P(G) = \exp\left(\sum_{A \subseteq E(G)} Q(G_A)\right) \quad .$$

Thus, $\alpha_A := Q(G_A)$ are the desired constants.

It remains to show that $\alpha_A = 0$ if A is not a clique in $\mathcal{D}$; and we are done with the first part of the theorem.

Proof of the Hammersley-Clifford Theorem.

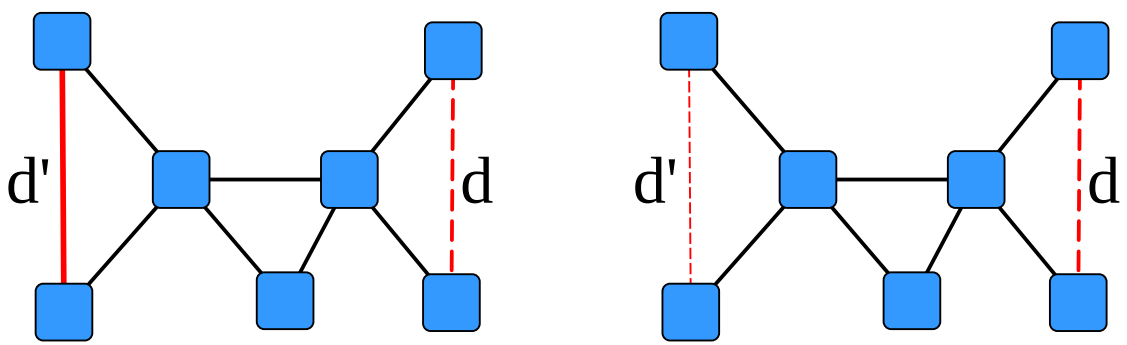
Want to show that $\alpha_A = 0$ if A is not a clique in $\mathcal{D}$.

Let $d, d' \in D$ be two conditionally independent dyads and $B \subseteq D$ with $d, d' \notin B$. It is

$$\frac{P(G_{B\cup\{d,d'\}})}{P(G_{B\cup d'}) + P(G_{B\cup\{d,d'\}})} = \frac{P(G_{B\cup\{d\}})}{P(G_B) + P(G_{B\cup\{d\}})} \quad .$$

and, hence

$$\frac{P(G_{B\cup\{d,d'\}})}{P(G_{B\cup\{d'\}})} = \frac{P(G_{B\cup\{d\}})}{P(G_B)} \quad . \tag{3}$$



Proof of the Hammersley-Clifford Theorem.

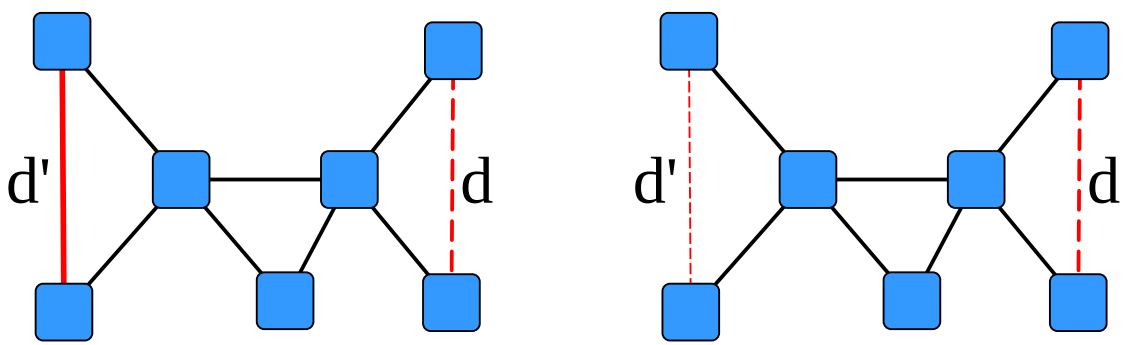
Want to show that $\alpha_A = 0$ if A is not a clique in $\mathcal{D}$.

Let $d, d' \in D$ be two conditionally independent dyads and $B \subseteq D$ with $d, d' \notin B$. It is

$$\frac{P(G_{B\cup\{d,d'\}})}{P(G_{B\cup d'}) + P(G_{B\cup\{d,d'\}})} = \frac{P(G_{B\cup\{d\}})}{P(G_B) + P(G_{B\cup\{d\}})} \quad .$$

and, hence

$$\frac{P(G_{B\cup\{d,d'\}})}{P(G_{B\cup\{d'\}})} = \frac{P(G_{B\cup\{d\}})}{P(G_B)} \quad . \tag{3}$$



Proof of the Hammersley-Clifford Theorem.

$$Q(G_A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \log P(G_B) \ .$$

Let $A \subseteq D$, contain two conditionally independent dyads d, d'.

$$\begin{aligned}
Q(G_A) &= \sum_{\substack{B \subseteq A \\ d,d' \in B}} (-1)^{|A \setminus B|} \log P(G_B) + \sum_{\substack{B \subseteq A \\ d \in B,\, d' \notin B}} (-1)^{|A \setminus B|} \log P(G_B) \\
&\quad + \sum_{\substack{B \subseteq A \\ d \notin B,\, d' \in B}} (-1)^{|A \setminus B|} \log P(G_B) + \sum_{\substack{B \subseteq A \\ d,d' \notin B}} (-1)^{|A \setminus B|} \log P(G_B) \\
&= \sum_{B \subseteq A \setminus \{d,d'\}} (-1)^{|A \setminus B|} \log \left(\frac{P(G_{B \cup \{d,d'\}})}{P(G_{B \cup \{d'\}})} \Big/ \frac{P(G_{B \cup \{d\}})}{P(G_B)} \right) \\
&= 0 \ , \text{ follows from (3)}
\end{aligned}$$

Thus, $Q(G_A) = 0$ if A is not a clique in $\mathcal{D}$.

Hammersley-Clifford Theorem.

Theorem (second part)

Conversely, if the probability P on $\mathcal{G}$ is defined by

$$P(G) = \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A\right), \qquad \textit{where}$$

$$\kappa = \sum_{G' \in \mathcal{G}} \exp\left(\sum_{A \subseteq E(G')} \alpha_A\right),$$

then two dyads d_1 and d_2 are conditionally independent in $(\mathcal{G}, P)$, unless there is a subset $A \subseteq D$ with $d_1, d_2 \in A$ and $\alpha_A \neq 0$.

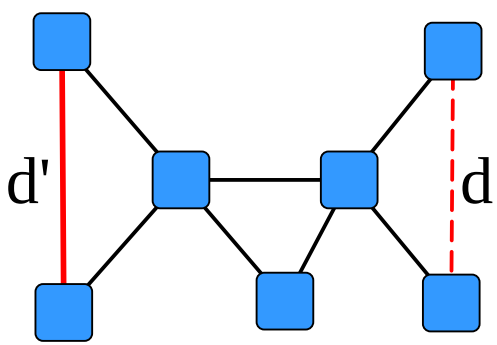
Proof of the Hammersley-Clifford Theorem.

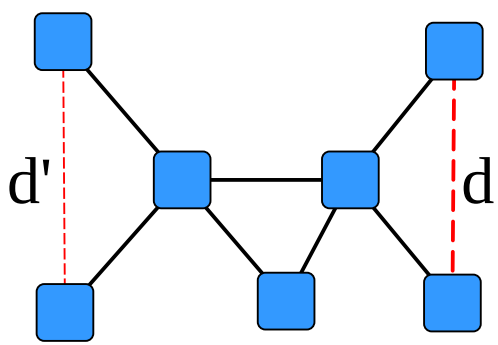
Suppose that d, d' are two dyads such that there is no subset $A \subseteq D$ with $\alpha_A \neq 0$ and $d, d' \in A$.
Let $B \subseteq D$ be any subset of dyads with $d, d' \notin B$. We show that

$$\frac{P(G_{B\cup\{d,d'\}})}{P(G_{B\cup d'}) + P(G_{B\cup\{d,d'\}})} = \frac{P(G_{B\cup\{d\}})}{P(G_B) + P(G_{B\cup\{d\}})} \ .$$

i. e., d and d' are conditionally independent. Equivalently,

$$\frac{P(G_{B\cup\{d,d'\}})}{P(G_{B\cup\{d'\}})} = \frac{P(G_{B\cup\{d\}})}{P(G_B)} \ .$$





Proof of the Hammersley-Clifford Theorem.

(d, d' are two dyads for which there is no subset $A \subseteq D$ with $\alpha_A \neq 0$ and $d, d' \in A$; $d_1, d_2 \notin B$)

$$\begin{aligned}
\log\left(\frac{P(G_{B\cup\{d,d'\}})}{P(G_{B\cup\{d'\}})}\right) &= \sum_{A\subseteq B\cup\{d,d'\}} \alpha_A - \sum_{A\subseteq B\cup\{d'\}} \alpha_A \\
&= \sum_{\substack{A\subseteq B\cup\{d,d'\}\\ d\in A}} \alpha_A \\
&= \sum_{\substack{A\subseteq B\cup\{d\}\\ d\in A}} \alpha_A \\
&= \sum_{A\subseteq B\cup\{d\}} \alpha_A - \sum_{A\subseteq B} \alpha_A \\
&= \log\left(\frac{P(G_{B\cup\{d\}})}{P(G_B)}\right) .
\end{aligned}$$

Hammersley-Clifford Theorem; special case.

Theorem

Let $(\mathcal{G}, P)$ be a random graph satisfying $P(G) > 0$ for all $G \in \mathcal{G}$.

There are constants $\{\alpha_A \in \mathbb{R};\ A \subseteq D\}$, satisfying $\alpha_A = 0$ if A is not a clique in $\mathcal{D}$, such that

$$P(G) \quad = \quad \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A \right) \quad . \tag{4}$$

Conversely, if P is defined by (4), then two dyads $d, d' \in D$ are conditionally independent, unless there is a subset $A \subseteq D$ with $d, d' \in A$ and $\alpha_A \neq 0$.

Outline.

Near-degeneracy and multi-modality of ERGMs.

Many Markov random graphs give rise to multi-modal probability distributions:

- probability mass centered on a small set of graphs
- other graphs are very unlikely.

For instance, only near-empty or near-complete graphs have a non-vanishing probability.

Near-degeneracy and multi-modality of ERGMs.

Consider the following ERGM

$$P(G) \quad = \quad \frac{1}{\kappa} \exp\left(\eta \cdot m(G) + \tau \cdot t(G)\right) \text{ with } \eta < 0,\ \tau > 0 \ .$$

Then, in very **sparse** networks

- there is no possibility to close triangles;
- creation of edges is unlikely;

$\Rightarrow$ very unlikely to leave the set of near-empty graphs.

In contrast, in very **dense** networks

- an edge can close many triangles (up to $n-2$);
- deletion of edges is unlikely;

$\Rightarrow$ very unlikely to leave the set of near-complete graphs.

Near-degeneracy and multi-modality of ERGMs.

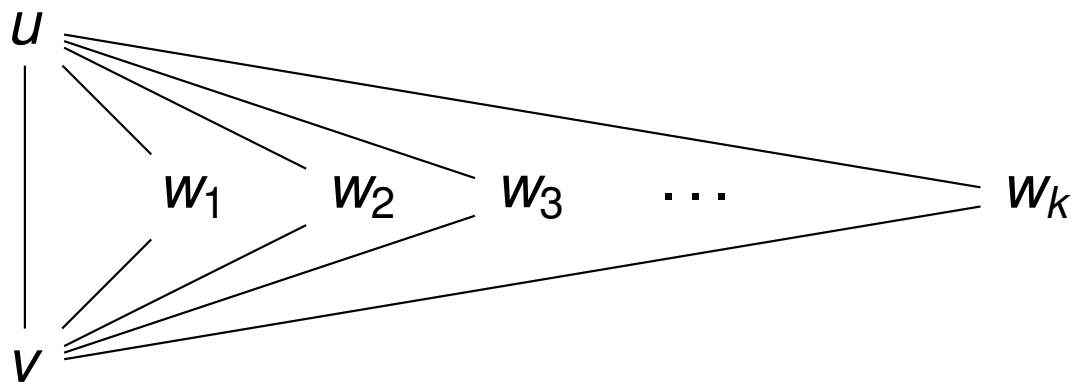
Degeneracy is undesirable for two reasons.

1. Convergence of the Markov chain towards the stationary distribution is very slow.
2. Degenerate models seem to be unreasonable models for empirical networks.

Avoiding near-degeneracy of ERGMs.

Assumed linear marginal effect of closed triangles:

- closing one triangle contributes τ to the log-probability;
- closing two triangles contributes 2τ ...



Geometrically-weighted edgewise shared partner (GWESP) statistic:

- a k-triangle counts more than a single triangle,
- but less than k-times as much.

$\Rightarrow$ no longer a Markov random graph.

Typically leads to less degenerate models.

Outline.

Testing hypotheses with random graph models.

Suppose that we are confronted with a social science hypothesis, such as

> *The friend of a friend is a friend.*

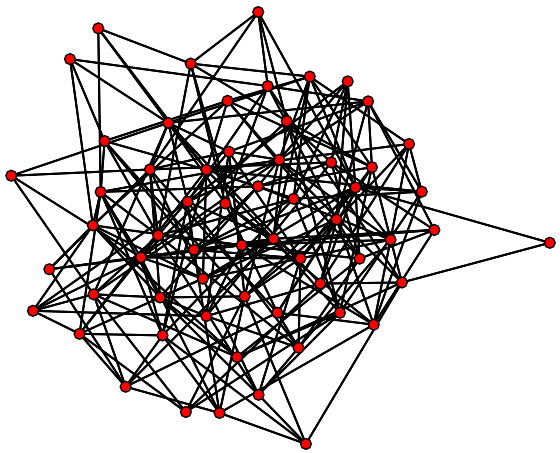
And we are given an observed friendship-network.

Approach to test this hypothesis.

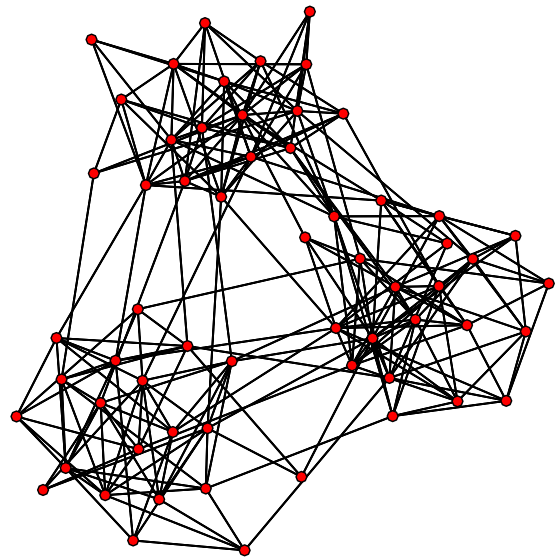
- Fix a numerical indicator for the hypothesis, e. g., number of triangles;
- fix a baseline random graph model (**null model**) violating the hypothesis, e. g., $\mathcal{G}(n, p)$, or an ERGM without the triangle statistic;
- compute probability of observed indicator value (or higher) in the null model;
- decide about rejection of null model.

Testing hypotheses: example.

$\mathcal{G}(60, 0.13)$



observed



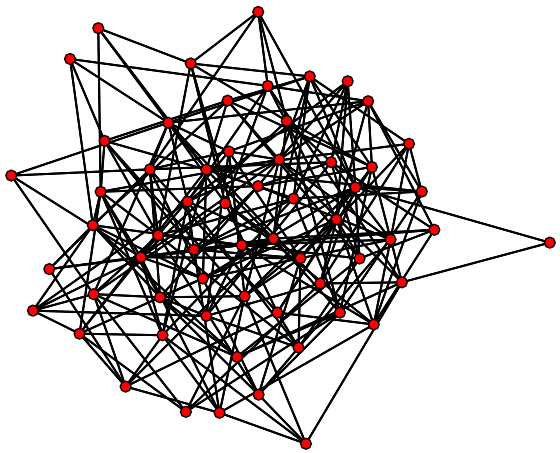
expected **75.2** triangles

148

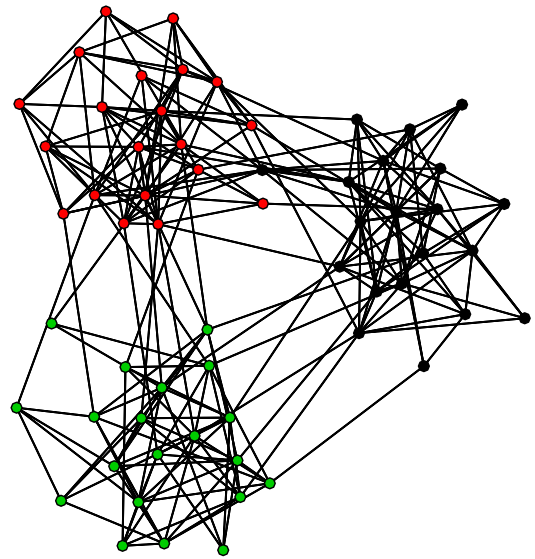
Many more triangles than expected $\Rightarrow$ reject null model.

Testing hypotheses: example.

$\mathcal{G}(60, 0.13)$



observed



expected **75.2** triangles

148

But: observed number of triangles can be expected in a more complex null model (without transitivity).

Testing hypotheses with ERGMs.

Given a hypothesis and an observed network G_{obs}.

- Decide on a reasonable set of statistics g_i, $i = 1, \ldots, k$

$$P(G) = \frac{1}{\kappa} \exp\left(\sum_{i=1}^{k} \theta_i g_i(G)\right)$$

- including a statistic related to the hypothesis, e. g., $g_k =$ number of triangles.
- Compute maximum likelihood estimates $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$.
- Compute probability of observing a network in the **null model** defined by $(\hat{\theta}_1, \ldots, \hat{\theta}_{k-1}, 0)$ that gives rise to θ_k as large as $\hat{\theta}_k$.

Estimation of ERGM Parameters.

Definition (maximum likelihood)

$(\mathcal{G}, P_\theta)$ random graph model parameterized by $\theta \in \Theta \subseteq \mathbb{R}^k$;
$G_{\text{obs}} \in \mathcal{G}$ a graph (observation).
Likelihood function associated with G_{obs}

$$L\colon \Theta \to \mathbb{R};\ \theta \mapsto P_\theta(G_{\text{obs}})$$

A parameter vector $\hat{\theta}$ maximizing L, i. e.,

$$\hat{\theta} = \arg\max_\theta L(\theta)$$

is called a *maximum likelihood estimate (MLE)* for θ.