Network Modeling

Viviana Amati   Jürgen Lerner   David Schoch

Dept. Computer & Information Science
University of Konstanz

Winter 2013/2014
(last updated: December 2, 2013)
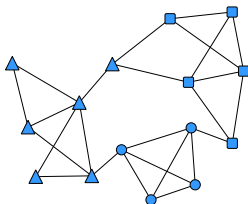
Outline.

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 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**.

What would happen, if we measured the data again?
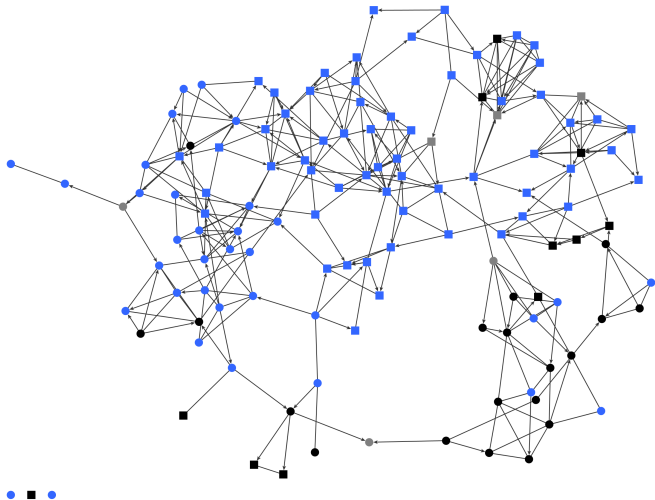
- at a different point in time,
- on a different set of actors,
- with different environmental factors, ...

estimate expected outcome $\pm$ **variability**

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

Example: friendship network among teenagers.

Can you see some pattern? Can you find explanations?

black: smokers; gray: occasional smokers; blue: non-smokers

Social influence vs. social selection.

Network ties and actor behavior evolve over time.

network(t) ⟶ network($t+1$)   social selection

behavior(t) ⟶ behavior($t+1$)   social influence
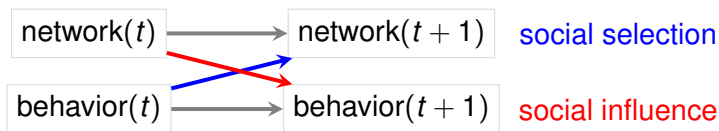
Social influence.

▶ E. g., friends of smokers start smoking.

Social selection.

▶ E. g., smokers choose smokers as friends (*homophily*).

Dependency among network ties.

▶ E. g., friends of friends become friends (*transitivity*).

Correlation of individual attributes.

▶ E. g., smokers typically drink more alcohol.

Topic of this lecture.

Statistical models for social network data.

Specify realistic **probability distributions** for social networks formalizing hypothetical dependencies in the data.

Statistical network models serve several purposes.

Explaining social relations and/or behavior

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

Predicting social relations and/or behavior

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

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 spreading, spread of disease).

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 $P(G)$.

Networks observed at two or more points in time

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

Continuously observed network changes or events

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

Treated in three parts of this lecture.

Outline.

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)
$\Omega = \{X \subset \{1, \ldots, 49\} \; ; \; |X| = 6\}$ (sets 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 be defined by

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

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 be defined by

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

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

Random graph models: remarks and 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:
  $D = \{\{u, v\} \, ; \, u, v \in V, \, u \neq v\}$.

▶ For directed, loopless graphs:
  $D = \{(u, v) \, ; \, u, v \in V, \, u \neq v\}$.

The set of vertices is fixed; all the randomness is in the edges.

Random graph models: remarks and 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:
  $D = \{\{u, v\} ; \ u, v \in V, \ u \neq v\}$.
- For directed, loopless graphs:
  $D = \{(u, v) ; \ u, v \in V, \ u \neq v\}$.

The set of vertices is fixed; all the randomness is in the edges.

Random graph models: remarks and 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:
  $D = \{\{u, v\}\,;\; u, v \in V,\; u \neq v\}$.
- For directed, loopless graphs:
  $D = \{(u, v)\,;\; u, v \in V,\; u \neq v\}$.

The set of vertices is fixed; all the randomness is in the edges.

Random graph models: edge probability.

Let $(\mathcal{G}, P)$ be a random graph model.
$P \colon \mathcal{G} \to [0, 1]$ defines a probability for each graph.

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 $P(\mathcal{G}_e)$;
sometimes written as $P(e)$ or $P(e \in E)$.

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

Does this also hold the other way round?

Random graph models: edge probability.

Let $(\mathcal{G}, P)$ be a random graph model.
$P \colon \mathcal{G} \to [0, 1]$ defines a probability for each graph.

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 $P(\mathcal{G}_e)$;
sometimes written as $P(e)$ or $P(e \in E)$.

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

Does this also hold the other way round?

Random graph models: edge probability.

Let $(\mathcal{G}, P)$ be a random graph model.
$P \colon \mathcal{G} \to [0, 1]$ defines a probability for each graph.

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 $P(\mathcal{G}_e)$;
sometimes written as $P(e)$ or $P(e \in E)$.

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

Does this also hold the other way round?

Example: two random graph models.

Let $\mathcal{G}$ be the set of undirected, loopless graphs $G = (V, E)$ with $V = \{1, 2, 3\}$.

Define P_1 by $P_1(G) = 1/8$ for all $G \in \mathcal{G}$.

Define P_2 by

$$P_2(G) = \left\{ \begin{array}{ll} 1/2 & \text{if } E = \emptyset \text{ or } E = D; \\ 0 & \text{else} \end{array} \right.$$

Both models define the same edge probabilities; but the models are not the same.

Independence and non-independence of edges.
(intuition)

In some cases the existence of an edge (or several edges) changes the probability of other edges.

For instance: does $P(e)$ change when the nodes incident to e are indirectly connected via a third node? How? Why?

Independence and non-independence of edges.

small facebook network

769 nodes, 295 296 dyads, 16 656 edges
$\Rightarrow$ average edge probability is 0.056
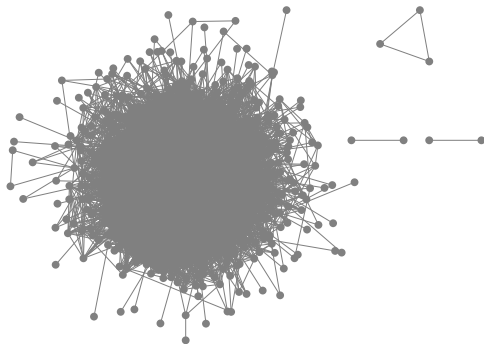
186 722 dyads are indirectly connected via a third node;
16 556 of these are edges $\Rightarrow$ average conditional edge
probability for indirectly connected nodes is 0.089

Independence and non-independence of edges.

small facebook network

769 nodes, 295 296 dyads, 16 656 edges
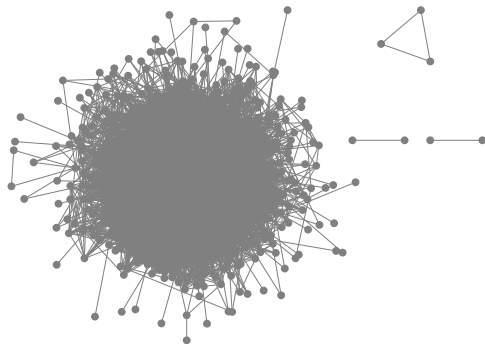$\Rightarrow$ average edge probability is 0.056

186 722 dyads are indirectly connected via a third node;
16 556 of these are edges $\Rightarrow$ average conditional edge
probability for indirectly connected nodes is 0.089

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

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

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

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

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

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

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

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

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

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

Example (probability space: dice)

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

Independence of dyads in random graph models.

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"

Example: two random graph models revisited.

Let $\mathcal{G}$ be the set of undirected, loopless graphs $G = (V, E)$ with $V = \{1, 2, 3\}$.

Let $P_1(G) = 1/8$. It is for two different dyads e_1 and e_2

$$P_1(\mathcal{G}_{e_1} \cap \mathcal{G}_{e_2}) = 1/4 = 1/2 \cdot 1/2 = P_1(\mathcal{G}_{e_1}) \cdot P_1(\mathcal{G}_{e_2})$$

Let

$$P_2(G) = \left\{ \begin{array}{ll} 1/2 & \text{if } E = \emptyset \text{ or } E = D; \\ 0 & \text{else.} \end{array} \right.$$

It is for two different dyads e_1 and e_2

$$P_2(\mathcal{G}_{e_1} \cap \mathcal{G}_{e_2}) = 1/2 \neq 1/2 \cdot 1/2 = P_2(\mathcal{G}_{e_1}) \cdot P_2(\mathcal{G}_{e_2})$$

Structural balance theory (illustrating dependence).

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

balanced　　　　　　　　　　　**not balanced**

SBT claims that actors prefer balanced networks.

In an appropriate random graph model, it holds that

- ▶ all dyads are pairwise independent;
- ▶ every dyad depends on the two others (i. e., there is a higher-order dependence).
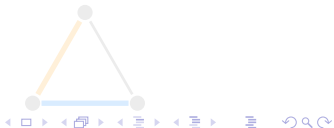
Structural balance theory (illustrating dependence).

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

balanced   **not balanced**
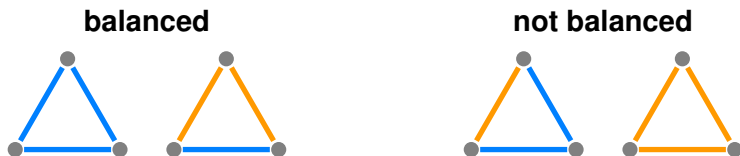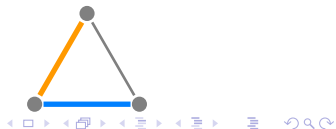

SBT claims that actors prefer balanced networks.

In an appropriate random graph model, it holds that

- all dyads are pairwise independent;
- every dyad depends on the two others (i. e., there is a higher-order dependence).

Fully independent random graph models.

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

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

Recall: 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

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

If every dyad e is independent of every subset $D' \subseteq D \setminus \{e\}$, then we say that the random graph model is *fully independent*.

Fully independent random graph models.

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

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

Recall: 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

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

If every dyad e is independent of every subset $D' \subseteq D \setminus \{e\}$, then we say that the random graph model is *fully independent*.

Fully independent random graph models.

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

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

Recall: 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

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

If every dyad e is independent of every subset $D' \subseteq D \setminus \{e\}$, then we say that the random graph model is *fully independent*.

A fully independent random graph model is determined by the edge probabilities of all dyads.

Let $(\mathcal{G}, P)$ be a fully independent random graph model.
Then the probability of a graph $G = (V, E) \in \mathcal{G}$ is

$$
\begin{aligned}
P(G) &= P(\{G\}) \\
&= P\left(\bigcap_{d \in E} \mathcal{G}_d \cap \bigcap_{d \in D \setminus E} \overline{\mathcal{G}_d}\right) \\
&= \prod_{d \in E} P(\mathcal{G}_d) \cdot \prod_{d \in D \setminus E} P(\overline{\mathcal{G}_d}) \\
&= \prod_{d \in E} P(d \in E) \cdot \prod_{d \in D \setminus E} 1 - P(d \in E) \ .
\end{aligned}
$$

For illustration, we treat in the following

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

of the uniform random graph model.

Uniform graph model: edge probability.

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

Proof.
The two sets

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

- 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.

Claim

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.

Claim
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.

Claim
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.

Claim
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.

Claim
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 $\mathcal{G} \Rightarrow P(\mathcal{G}_e \cup \overline{\mathcal{G}_e}) = 1$.

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

$\square$

Uniform graph model: independence.

Claim
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

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

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). $\qquad\square$

Uniform graph model: independence.

Claim
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

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

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). $\qquad \square$

Uniform graph model: independence.

Claim
*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

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

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). $\qquad\square$

Uniform graph model: independence.

Claim

*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

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

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). $\qquad \square$

Background: random variable and expectation.

Let (Ω, P) be a finite probability space.

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

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

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

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

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 finite probability space.

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

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

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

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

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.

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

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

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

$\square$

Uniform graph model: expected number of edges.

Claim
The expected number of edges in the uniform random graph model with n vertices equals $\frac{n(n-1)}{4}$.

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

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

$\square$

Uniform graph model: expected number of edges.

Claim

The expected number of edges in the uniform random graph model with n vertices equals $\frac{n(n-1)}{4}$.

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

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

$\square$

Uniform graph model: expected number of edges.

Claim

The expected number of edges in the uniform random graph model with n vertices equals $\frac{n(n-1)}{4}$.

Proof.

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

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

$\square$

Uniform graph model: expected number of edges.

Claim

The expected number of edges in the uniform random graph model with n vertices equals $\frac{n(n-1)}{4}$.

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

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

$\square$

Uniform graph model: summary.

Characterizing properties:

- the model is fully independent;
- all edges are equally likely;
- no preference for edges over non-edges or vice versa.

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:

- the model is fully independent;
- all edges are equally likely;
- no preference for edges over non-edges or vice versa.

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.

Outline.

Outline.

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

Let $n \in \mathbb{N}_{\geq 1}$ and p be a real number $0 < p < 1$.

$\mathcal{G}(n, p)$ is the random graph model on the set of undirected, loopless graphs with vertex set $V = \{1, \ldots, n\}$ that defines the probability of a graph G with m edges by

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

Note: P is normalized since (let $M = n(n - 1)/2$)

$$\begin{aligned}
\sum_{G \in \mathcal{G}} P(G) &= \sum_{m=0}^{M} \binom{M}{m} p^m (1 - p)^{M-m} \\
&= (p + (1 - p))^M = 1^M = 1 .
\end{aligned}$$

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

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

Let $n \in \mathbb{N}_{\geq 1}$ and p be a real number $0 < p < 1$.

$\mathcal{G}(n, p)$ is the random graph model on the set of undirected, loopless graphs with vertex set $V = \{1, \ldots, n\}$ that defines the probability of a graph G with m edges by

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

Note: P is normalized since (let $M = n(n-1)/2$)

$$
\begin{aligned}
\sum_{G \in \mathcal{G}} P(G) &= \sum_{m=0}^{M} \binom{M}{m} p^m (1 - p)^{M - m} \\
&= (p + (1 - p))^M = 1^M = 1 .
\end{aligned}
$$

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

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

Let $n \in \mathbb{N}_{\geq 1}$ and p be a real number $0 < p < 1$.

$\mathcal{G}(n, p)$ is the random graph model on the set of undirected, loopless graphs with vertex set $V = \{1, \ldots, n\}$ that defines the probability of a graph G with m edges by

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

Note: P is normalized since (let $M = n(n - 1)/2$)

$$
\begin{aligned}
\sum_{G \in \mathcal{G}} P(G) &= \sum_{m=0}^{M} \binom{M}{m} p^m (1 - p)^{M-m} \\
&= (p + (1 - p))^M = 1^M = 1 .
\end{aligned}
$$

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

Characterizing properties of $\mathcal{G}(n, p)$.

The probability of a graph G with m edges is defined by

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

Claim

1. *The edge probability of every dyad is equal to p.*
2. *The model is fully independent.*
3. *There is just one model satisfying properties (1) and (2).*

Proof.
See next exercise sheet. □

Further properties of $\mathcal{G}(n, p)$.

- Expected number of edges is $p\frac{n(n-1)}{2}$.
- Expected density is p.

Outline.

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

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.

Runtime is in $\Theta(n^2)$ (independent of p)
$\Rightarrow$ inefficient for small p (i. e., sparse graphs).

The expected size of a graph from $\mathcal{G}(n, p)$ is in $\Theta(n + p \cdot n^2)$.

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

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.

Runtime is in $\Theta(n^2)$ (independent of p)
$\Rightarrow$ inefficient for small p (i. e., sparse graphs).

The expected size of a graph from $\mathcal{G}(n, p)$ is in $\Theta(n + p \cdot n^2)$.

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

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.

Runtime is in $\Theta(n^2)$ (independent of p)
$\Rightarrow$ inefficient for small p (i. e., sparse graphs).

The expected size of a graph from $\mathcal{G}(n, p)$ is in $\Theta(n + p \cdot n^2)$.

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 $m \in \mathcal{O}(n)$;

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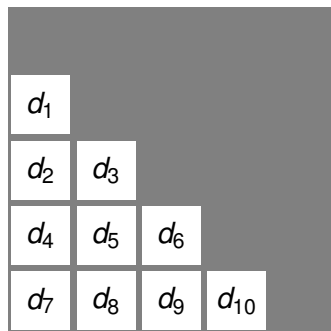$\overline{d} \in \mathcal{O}(1)$.

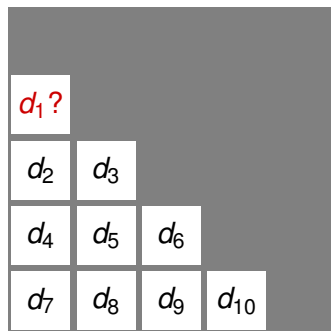Empirical observation: social networks are typically sparse.

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

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

d_1			
d_2	d_3		
d_4	d_5	d_6	
d_7	d_8	d_9	d_{10}

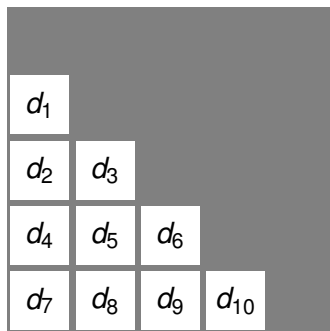inefficient, when p is small (too many NOs)

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

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. . .)

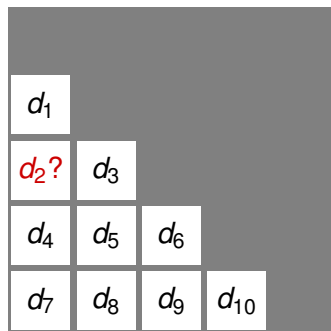| d_1?			
d_2	d_3		
d_4	d_5	d_6	
d_7	d_8	d_9	d_{10}

inefficient, when p is small (too many NOs)

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

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)

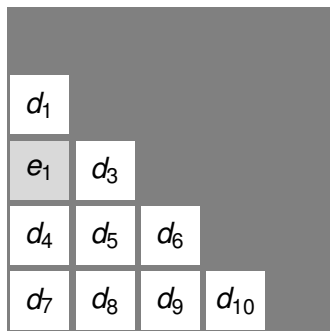# Generating graphs from $\mathcal{G}(n, p)$ (example).

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. . .)

d_1			
d_2?	d_3		
d_4	d_5	d_6	
d_7	d_8	d_9	d_{10}

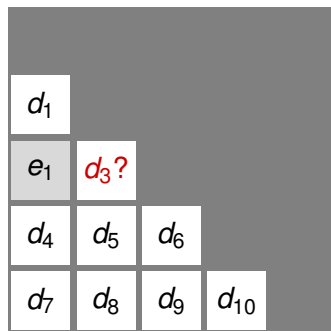inefficient, when p is small (too many NOs)

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

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

d_1			
e_1	d_3		
d_4	d_5	d_6	
d_7	d_8	d_9	d_{10}

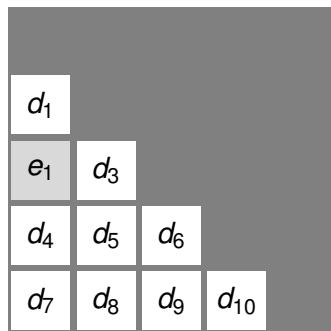inefficient, when p is small (too many NOs)

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

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...)

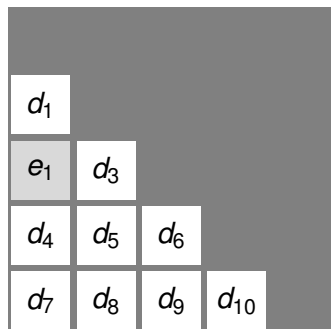|			
d_1			
e_1	d_3?		
d_4	d_5	d_6	
d_7	d_8	d_9	d_{10}

inefficient, when p is small (too many NOs)

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

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)

d_1			
e_1	d_3		
d_4	d_5	d_6	
d_7	d_8	d_9	d_{10}

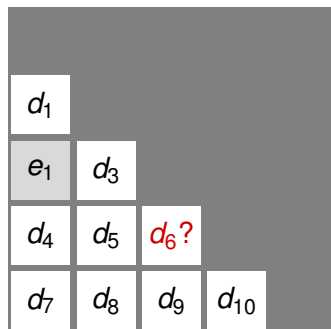inefficient, when p is small (too many NOs)

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

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)

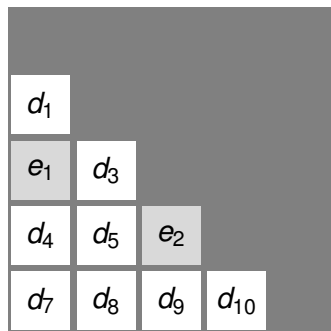# Generating graphs from $\mathcal{G}(n, p)$ (example).

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. . .)

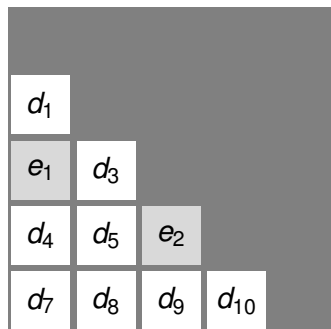| d_1 | | | |
e_1	d_3		
d_4	d_5	d_6?	
d_7	d_8	d_9	d_{10}

inefficient, when p is small (too many NOs)

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

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

d_1			
e_1	d_3		
d_4	d_5	e_2	
d_7	d_8	d_9	d_{10}

inefficient, when p is small (too many NOs)

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

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

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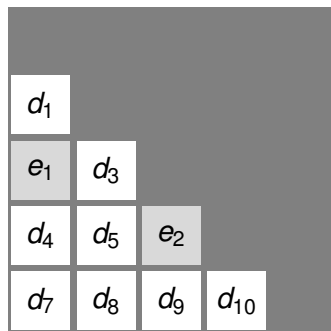- if $r \leq p$ add e to the edge set.

to be continued . . .

d_1			
e_1	d_3		
d_4	d_5	e_2	
d_7	d_8	d_9	d_{10}

inefficient, when p is small (too many NOs)

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

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 . . .

d_1			
e_1	d_3		
d_4	d_5	e_2	
d_7	d_8	d_9	d_{10}

inefficient, when p is small (too many NOs)

Sampling efficiently from $\mathcal{G}(n, p)$.

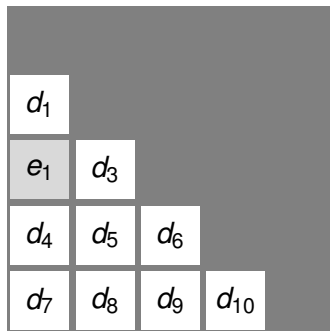**Better** ask the question:

> *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

d_1			
d_2	d_3		
d_4	d_5	d_6	
d_7	d_8	d_9	d_{10}

Sampling efficiently from $\mathcal{G}(n, p)$.

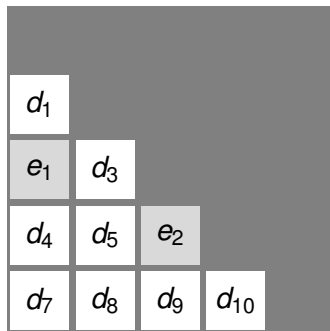**Better** ask the question:

> *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

d_1			
e_1	d_3		
d_4	d_5	d_6	
d_7	d_8	d_9	d_{10}
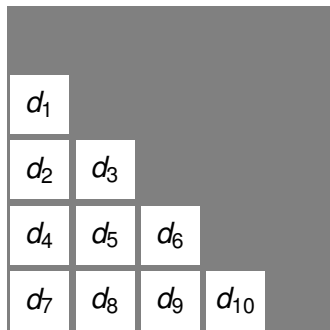
Sampling efficiently from $\mathcal{G}(n, p)$.

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

Sampling efficiently from $\mathcal{G}(n, p)$.

How many dyads shall be left out?

(Notation: $q = 1 - p$ on this slide.)

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 $0, 1, \dots$) with probability $q^k p$ and add the $k + 1$th dyad to the edge set.

draw $k = 1$ (happens with probability qp)

draw $k = 3$ (happens with probability $q^3 p$)

Sampling efficiently from $\mathcal{G}(n, p)$.
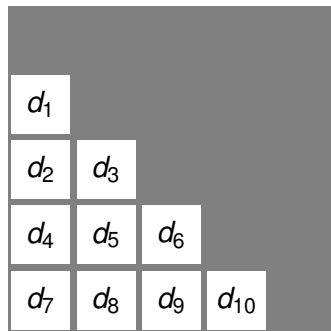
How many dyads shall be left out?

(Notation: $q = 1 - p$ on this slide.)

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 $0, 1, \dots$) with probability $q^k p$ and add the $k + 1$th dyad to the edge set.

draw $k = 1$ (happens with probability qp)

draw $k = 3$ (happens with probability $q^3 p$)

Sampling efficiently from $\mathcal{G}(n, p)$.
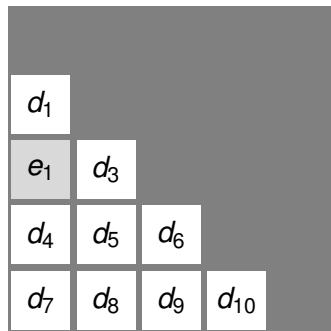
How many dyads shall be left out?

(Notation: $q = 1 - p$ on this slide.)

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 $0, 1, \dots$) with probability $q^k p$ and add the $k + 1$th dyad to the edge set.

draw $k = 1$ (happens with probability qp)

draw $k = 3$ (happens with probability $q^3 p$)

Sampling efficiently from $\mathcal{G}(n, p)$.
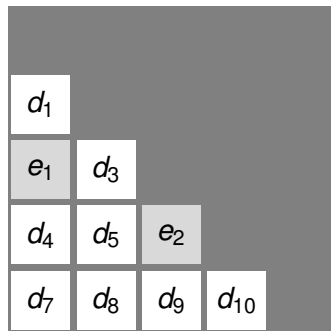
How many dyads shall be left out?

(Notation: $q = 1 - p$ on this slide.)

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 $k = 1$ (happens with probability qp)

draw $k = 3$ (happens with probability $q^3 p$)

Background: geometric distribution.

The distribution that assigns the probability $P(k) = p \cdot (1 - p)^k$ to the non-negative integers $k = 0, 1, 2, \ldots$ is called the **geometric distribution**.

Such a random number generator is implemented in R (function `rgeom`).

Equivalent: draw a uniformly distributed real number r from $(0, 1)$ and return

$$k = \left\lfloor \frac{\log(r)}{\log(1 - p)} \right\rfloor$$

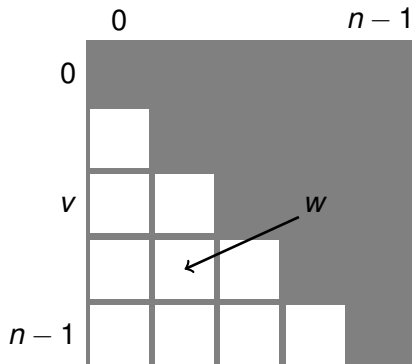# Sampling efficiently from $\mathcal{G}(n, p)$ (algorithm).

```
E ← ∅
v ← 1  w ← −1
while v < n do
    k ← rgeom(p)
    w ← w + k + 1
    while w ≥ v and v < n do
        w ← w − v
        v ← v + 1
    if v < n then
        E ← E ∪ {{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.

Sampling efficiently from $\mathcal{G}(n, p)$ (algorithm).

```
E ← ∅
v ← 1  w ← −1
while v < n do
    k ← rgeom(p)
    w ← w + k + 1
    while w ≥ v and v < n do
        w ← w − v
        v ← v + 1
    if v < n then
        E ← E ∪ {{v, w}}
return G = (V, E)
```
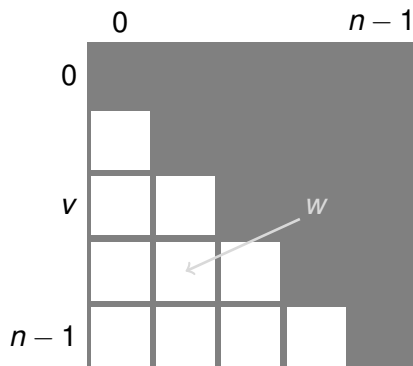

If $w \geq v$ then w is reduced by v and the row index v is incremented by one.

Sampling efficiently from $\mathcal{G}(n, p)$ (runtime).

```
E ← ∅
v ← 1  w ← −1
while v < n do
    k ← rgeom(p)
    w ← w + k + 1
    while w ≥ v and v < n do
        w ← w − v
        v ← v + 1
    if v < n then
        E ← E ∪ {{v, w}}
return G = (V, E)
```

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

Inner while loop is executed (in total) $n − 1$ times.

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

Efficient sampling from $\mathcal{G}(n, p)$ (correctness).

Let G be a graph with m edges; compute probability that G is returned by the sampling algorithm.

For $i = 1, \ldots, m + 1$ let k_i be number of non-edges between $(i - 1)$th and ith edge.

Algorithm returns G if and only if

1. For all $i = 1, \ldots, m$, the random number k in the ith iteration satisfies $k = k_i$.
2. For $i = m + 1$ the random number k in the $m + 1$th iteration satisfies $k \geq k_{m+1}$.

Efficient sampling from $\mathcal{G}(n, p)$ (correctness).

Algorithm returns G if and only if

1. For all $i = 1, \ldots, m$, the random number k in the ith iteration satisfies $k = k_i$.
   Happens with probability pq^{k_i}.

2. For $i = m + 1$ the random number k in the $m + 1$th iteration satisfies $k \geq k_{m+1}$.
   Happens with probability $q^{k_{m+1}}$.

Efficient sampling from $\mathcal{G}(n, p)$ (correctness).

Algorithm returns G if and only if

1. For all $i = 1, \ldots, m$, the random number k in the ith iteration satisfies $k = k_i$.
   Happens with probability pq^{k_i}.

2. For $i = m + 1$ the random number k in the $m + 1$th iteration satisfies $k \geq k_{m+1}$.
   Happens with probability $q^{k_{m+1}}$.

Efficient sampling from $\mathcal{G}(n, p)$ (correctness).

Algorithm returns G if and only if

1. For all $i = 1, \ldots, m$, the random number k in the ith iteration satisfies $k = k_i$.
   Happens with probability pq^{k_i}.

2. For $i = m + 1$ the random number k in the $m + 1$th iteration satisfies $k \geq k_{m+1}$.
   Happens with probability $q^{k_{m+1}}$.

Efficient sampling from $\mathcal{G}(n, p)$ (correctness).

Algorithm returns G if and only if

1. For all $i = 1, \ldots, m$, the random number k in the ith iteration satisfies $k = k_i$.
   Happens with probability pq^{k_i}.

2. For $i = m + 1$ the random number k in the $m + 1$th iteration satisfies $k \geq k_{m+1}$.
   Happens with probability $q^{k_{m+1}}$.

$$\sum_{j=k_{m+1}}^{\infty} pq^j = \sum_{j=0}^{\infty} pq^j - \sum_{j=0}^{k_{m+1}-1} pq^j = 1 - (1 - q^{k_{m+1}}) = q^{k_{m+1}} \ .$$

Efficient sampling from $\mathcal{G}(n, p)$ (correctness).

Algorithm returns G if and only if

1. For all $i = 1, \ldots, m$, the random number k in the ith iteration satisfies $k = k_i$.
   Happens with probability pq^{k_i}.

2. For $i = m + 1$ the random number k in the $m + 1$th iteration satisfies $k \geq k_{m+1}$.
   Happens with probability $q^{k_{m+1}}$.

All conditions for $i = 1, \ldots, m + 1$ are satisfied with probability

$$q^{k_{m+1}} \prod_{i=1}^{m} pq^{k_i} = p^m q^{\sum_{i=1}^{m+1} k_i} = p^m q^{\frac{n(n-1)}{2} - m} \ .$$

Outline.

Can such a network be drawn from a $\mathcal{G}(n, p)$ model?

Graph has 769 vertices and about 16 600 edges.
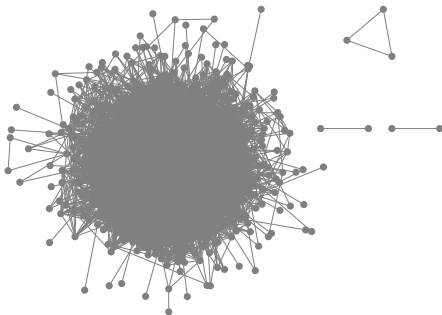
Which $\mathcal{G}(n, p)$?
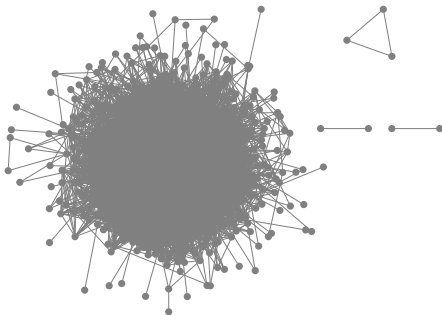*What is the most likely value for the parameter p?*

Can such a network be drawn from a $\mathcal{G}(n, p)$ model?

Graph has 769 vertices and about 16 600 edges.

Which $\mathcal{G}(n, p)$?
What is the most likely value for the parameter p?

Background: maximum likelihood.

Problem: given a graph *G* drawn from some parameterized random graph model (without knowing the parameter value).

What is the most likely parameter value?

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}

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

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.

Background: maximum likelihood.

Problem: given a graph *G* drawn from some parameterized random graph model (without knowing the parameter value).

What is the most likely parameter value?

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}

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

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.

Background: maximum likelihood.

Problem: given a graph *G* drawn from some parameterized random graph model (without knowing the parameter value).

What is the most likely parameter value?

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}

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

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.

Background: maximum likelihood.

Problem: given a graph *G* drawn from some parameterized random graph model (without knowing the parameter value).

What is the most likely parameter value?

Definition (maximum likelihood)

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

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

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.

Background: maximum likelihood.

Problem: given a graph *G* drawn from some parameterized random graph model (without knowing the parameter value).

What is the most likely parameter value?

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}

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

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

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

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 $[\dots]$.

Maximum likelihood estimate of p in $\mathcal{G}(n, p)$.

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 $[\ldots]$.

Maximum likelihood estimate of p in $\mathcal{G}(n, p)$.

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

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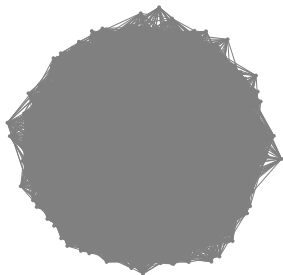m \cdot (1-p) &= p \cdot (M-m) \\
m - pm &= pM - pm \\
\frac{m}{M} &= p
\end{aligned}
$$

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

Both graphs have 769 vertices and about 16 600 edges.

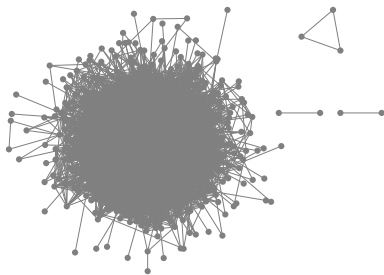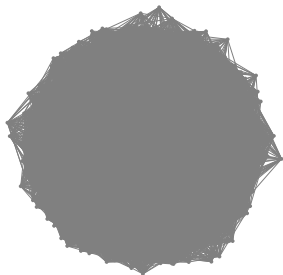Maximum likelihood estimate for p is 0.056
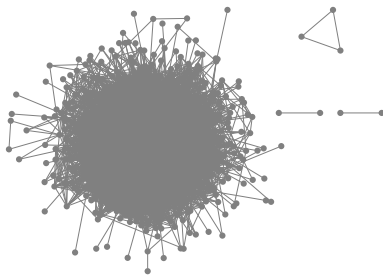
Which graph is more likely to be drawn from a $\mathcal{G}(n, p)$ model?

Both graphs have the same (very small) probability in $\mathcal{G}(n, p)$
$\Rightarrow$ the probability of the graph is not a good criterion.

Both graphs have 769 vertices and about 16 600 edges.

Maximum likelihood estimate for p is 0.056

Which graph is more likely to be drawn from a $\mathcal{G}(n, p)$ model?
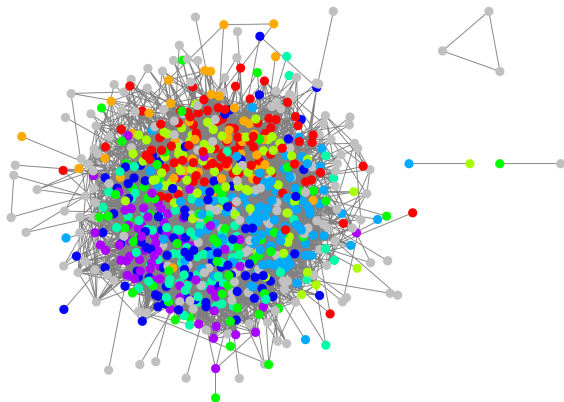
Both graphs have the same (very small) probability in $\mathcal{G}(n, p)$
$\Rightarrow$ the probability of the graph is not a good criterion.

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

Address this question by looking at some network properties:

1. inhomogeneity of the graph density;
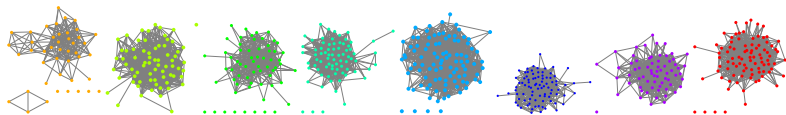2. skewness of the degree distribution;
3. number of triangles.

Inhomogeneity of the graph density.

Colors encode the **dorm** variable (gray for missing value).

Inhomogeneity of the graph density.

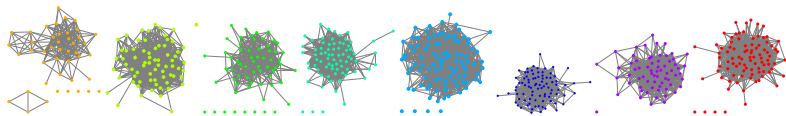Density of the whole network is 0.056

The subnetworks induced by the eight dorms have much higher densities, namely: 0.21, 0.37, 0.20, 0.35, 0.31, 0.24, 0.37, 0.25.

Can this happen in a $\mathcal{G}(n, p)$ model?

Probably not: probability that randomly drawn subnetworks of that size have such high density is very small.

Inhomogeneity of the graph density.

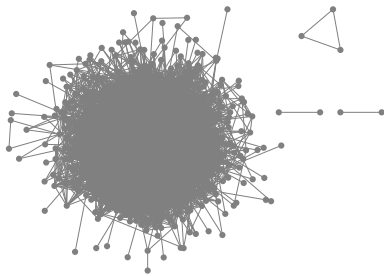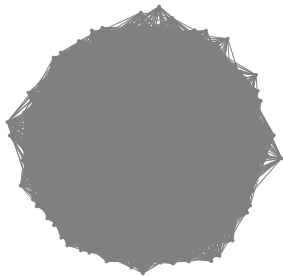Density of the whole network is 0.056

The subnetworks induced by the eight dorms have much higher densities, namely: 0.21, 0.37, 0.20, 0.35, 0.31, 0.24, 0.37, 0.25.

Can this happen in a $\mathcal{G}(n, p)$ model?

Probably not: probability that randomly drawn subnetworks of that size have such high density is very small.

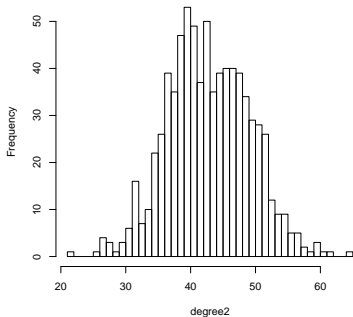Which graph is drawn from a $\mathcal{G}(n, p)$ model?

Comparing degree distributions.

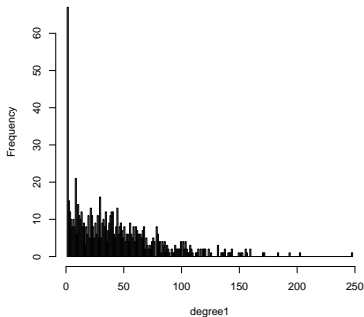Which graph is drawn from a $\mathcal{G}(n, p)$ model?

Plotting number of vertices (*y*-axis) with given degree (*x*-axis).

Histogram of degree2

Histogram of degree1

max degree is 65
min degree is 21

max degree is 248
min degree is 1

Probability of degree *k* in $\mathcal{G}(n, p)$.

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 $p^k q^{n-1-k}$. $\qquad\square$

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

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 $p^k q^{n-1-k}$. $\square$

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

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

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

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

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

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

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

details on the proof: let

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

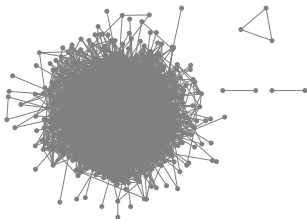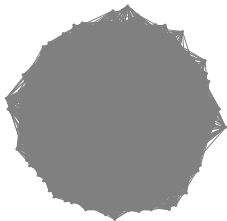be the set of k-element subsets of $V \setminus \{v\}$ (potential neighborhoods of size k of v).
Define for $U \in \mathcal{N}_k(v)$ the subset

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

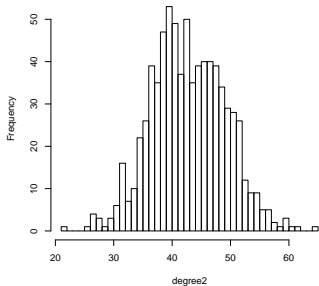(all graphs in which the neighborhood of v equals U).
Important fact: $\mathcal{G}_U$ and $\mathcal{G}_{U'}$ are disjoint for $U \neq U'$. Thus

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

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

Histogram of degree2

Frequency

degree2

Histogram of degree1

Frequency

degree1

Degree distribution in $\mathcal{G}(n, p)$ (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**.)*

Degree distribution in $\mathcal{G}(n, p)$ (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**.)*

Degree distribution in $\mathcal{G}(n, p)$ (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**.)*

Degree distribution in $\mathcal{G}(n, p)$ (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**.)*

Degree distribution in $\mathcal{G}(n, p)$ (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**.)*

Degree distribution in $\mathcal{G}(n, p)$ for large n.

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.
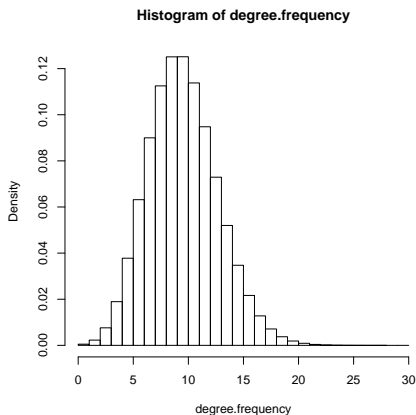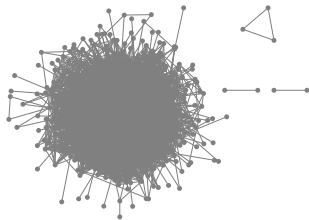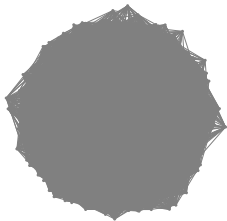
Histogram of degree.frequency

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

$$\lim_{n\to\infty} P_n[d(v)=k] = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \ .$$
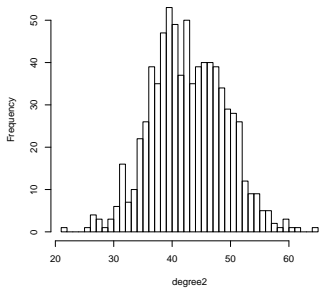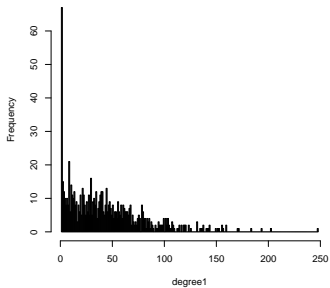
Which graph is drawn from a $\mathcal{G}(n, p)$ model?
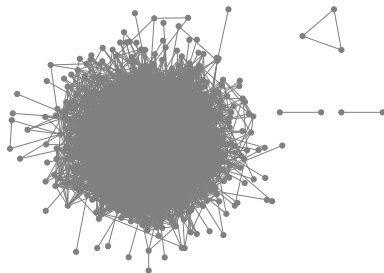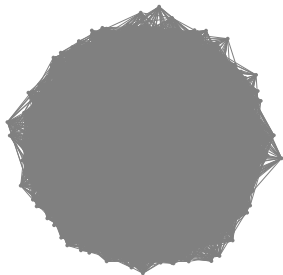
Comparing number of triangles.

Has been done in the last exercise sheet.

A nice quote.

> *"For their part, social scientists have reacted to this practice with considerable amusement. To them, baseline models like simple random graphs seem naïve to the extreme—like comparing the structure of a skyscraper to a random distribution of the same quantities of materials."* [p. 895]

Borgatti et al. **Network analysis in the social sciences.** *Science* 323, 2009.

Outline.

Outline.

Two simple approaches to define more structured models.

1. **Planted partition models:** allow varying probability between different vertices (but keeping independence as in the $\mathcal{G}(n, p)$ model).

2. **Incrementally defined models:** nodes and edges are incrementally added to the network; probability of later edges may depend on earlier ones.
   Example: preferential attachment.

Outline.

Recall: inhomogeneity of the graph density

Density of the whole network is 0.056

The subnetworks induced by the eight dorms have much higher densities, namely: 0.21, 0.37, 0.20, 0.35, 0.31, 0.24, 0.37, 0.25.

Planted partition models.

Definition

A *planted partition model* is defined by

- A partition of the vertex set $V = V_1 \cup \cdots \cup V_k$ into k disjoint classes.

- Probabilities $p_{ij} \in (0, 1)$ associated with each unordered pair of classes V_i and V_j.

- Two vertices $u \in V_i$ and $v \in V_j$ are connected by an edge with probability p_{ij}.

- The model is fully independent.

Planted partition models.

Vertex partition induces a partition of the adjacency matrix into blocks.

$$
\begin{bmatrix}
p_1 & \cdots & p_1 & p_2 & \cdots & p_2 & p_3 & \cdots & p_3 \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
p_1 & \cdots & p_1 & p_2 & \cdots & p_2 & p_3 & \cdots & p_3 \\
p_2 & \cdots & p_2 & p_4 & \cdots & p_4 & p_5 & \cdots & p_5 \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
p_2 & \cdots & p_2 & p_4 & \cdots & p_4 & p_5 & \cdots & p_5 \\
p_3 & \cdots & p_3 & p_5 & \cdots & p_5 & p_6 & \cdots & p_6 \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
p_3 & \cdots & p_3 & p_5 & \cdots & p_5 & p_6 & \cdots & p_6
\end{bmatrix}
$$

Can be used to define random graphs that are well clusterable; are k-colorable; have a large cut; have a small bisection; etc.

Outline.

Recall: degree distributions.
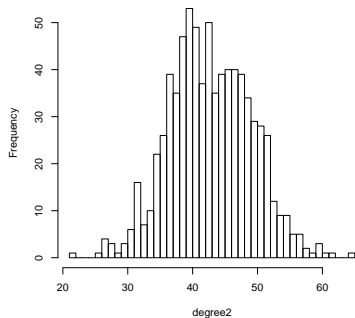
sampled from $\mathcal{G}(n, p)$ empirical network

Histogram of degree2

Histogram of degree1
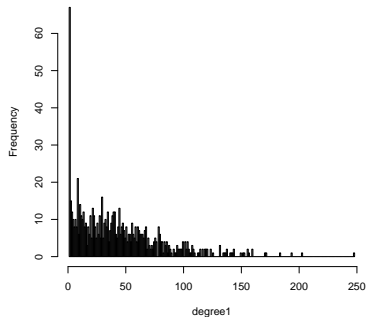
Other empirical dist. (Barabasi and Albert, 1999).

Note: logarithmic scaling of axes.
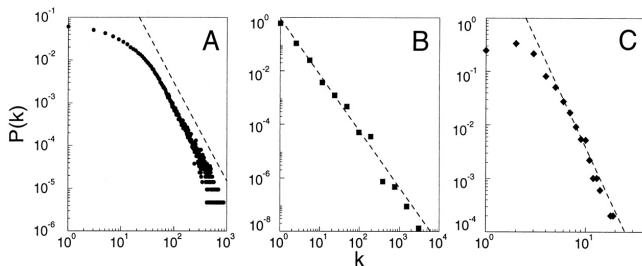
- A Actor collaboration network $n = 212,250$ and $\overline{d} = 28.78$
- B WWW $n = 325,729$ and $\overline{d} = 5.46$
- C Power grid $n = 4,941$ and $\overline{d} = 2.67$

Preferential attachment: motivation and history.

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

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

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*:

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

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*:

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

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*:

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

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
       $\Rightarrow$ target node gets selected with the correct probability.

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
     $\Rightarrow$ target node gets selected with the correct probability.

Preferential attachment (algorithm).

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

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

foreach $v = 0, \dots, n - 1$ **do**
    foreach $j = 0, \dots, b - 1$ **do**
        $A[2m] \leftarrow v$ //*v is source of next edge*
        $w \leftarrow A[\text{random}(\{0, \dots, 2m\})]$ //*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*
    $\Rightarrow$ target node gets selected with the correct probability.

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*
     $\Rightarrow$ target node gets selected with the correct probability.

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*
    $\Rightarrow$ target node gets selected with the correct probability.

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*
    $\Rightarrow$ target node gets selected with the correct probability.

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*
   $\Rightarrow$ target node gets selected with the correct probability.

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*
       $\Rightarrow$ target node gets selected with the correct probability.

Some remarks.

It is relatively easy to define a simple model that reproduces a given property of empirical social networks.

But different properties might be interrelated:

For instance, a planted partition model with dense diagonal blocks yields more triangles than a $\mathcal{G}(n, p)$ model with the same global density.

Difficulty lies in assessing some network property while controlling for others.

Outline.

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 (may) have an influence on the probability of a graph.
2. A set of **parameters** (associated with statistics) that determine how network statistics 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 function P_θ 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

- $g_i \colon \mathcal{G} \to \mathbb{R}$ 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

$$\kappa(\theta) = \sum_{G' \in \mathcal{G}} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G')\right) \ .$$

ERGM (example).

Consider undirected, loopless 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
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

$\Rightarrow \kappa = 1 + 3 \cdot 1/2 + 3 \cdot 1/4 + 2 = 21/4$

Relation between statistics and probability.

Probability function

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right) \ .$$

Isolating the effect of one specific statistic g_{i_0}:

$$P_\theta(G) = \exp[\theta_{i_0} \cdot g_{i_0}(G)] \cdot \frac{1}{\kappa(\theta)} \exp\left(\sum_{i \neq i_0} \theta_i \cdot g_i(G)\right) \ .$$

$\Rightarrow$ if $g_{i_0}(G') = g_{i_0}(G) + c$ and $g_i(G') = g_i(G)$ for all $i \neq i_0$,
then $P(G') = \exp(\theta_{i_0})^c \cdot P(G)$.

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.

Probability function

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right) .$$

Isolating the effect of one specific statistic g_{i_0}:

$$P_\theta(G) = \exp[\theta_{i_0} \cdot g_{i_0}(G)] \cdot \frac{1}{\kappa(\theta)} \exp\left(\sum_{i \neq i_0} \theta_i \cdot g_i(G)\right) .$$

$\Rightarrow$ if $g_{i_0}(G') = g_{i_0}(G) + c$ and $g_i(G') = g_i(G)$ for all $i \neq i_0$, then $P(G') = \exp(\theta_{i_0})^c \cdot P(G)$.

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.

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 $g_{i_0}(G') = g_{i_0}(G) + c$ and $g_i(G') = g_i(G)$ for all $i \neq i_0$, then $P(G') = \exp(\theta_{i_0})^c \cdot P(G)$.

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.

Probability function

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right) \ .$$

Isolating the effect of one specific statistic g_{i_0}:

$$P_\theta(G) = \exp[\theta_{i_0} \cdot g_{i_0}(G)] \cdot \frac{1}{\kappa(\theta)} \exp\left(\sum_{i \neq i_0} \theta_i \cdot g_i(G)\right) \ .$$

$\Rightarrow$ if $g_{i_0}(G') = g_{i_0}(G) + c$ and $g_i(G') = g_i(G)$ for all $i \neq i_0$, then $P(G') = \exp(\theta_{i_0})^c \cdot P(G)$.

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.

$$P_\theta(G) = \exp[\theta_{i_0} \cdot g_{i_0}(G)] \cdot \frac{1}{\kappa(\theta)} \exp\left(\sum_{i \neq i_0} \theta_i \cdot g_i(G)\right) \ .$$
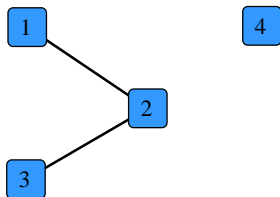
Edge between 1 and 3 is $\exp(\theta_{i_0})$-times as likely as between 1 and 4.

If other statistics change identically!

Positive $\theta_{i_0} \Rightarrow$ more likely; negative $\theta_{i_0} \Rightarrow$ less likely.

Example: $\mathcal{G}(n, p)$ belongs to the ERGM class.

Lemma

$\mathcal{G}(n, p)$ *is identical with the ERGM defined by*

$$P_1(G) = \exp\left[\theta \cdot m(G)\right] \cdot \kappa(\theta)^{-1}$$

where $\theta = \log\left(\frac{p}{1-p}\right)$.

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'. $\qquad \square$

Example: $\mathcal{G}(n, p)$ belongs to the ERGM class.

Lemma

$\mathcal{G}(n, p)$ *is identical with the ERGM defined by*

$$P_1(G) = \exp\left[\theta \cdot m(G)\right] \cdot \kappa(\theta)^{-1}$$

where $\theta = \log\left(\frac{p}{1-p}\right)$.

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'. $\qquad\square$

Example: $\mathcal{G}(n, p)$ belongs to the ERGM class.

Lemma

$\mathcal{G}(n, p)$ *is identical with the ERGM defined by*

$$P_1(G) = \exp\left[\theta \cdot m(G)\right] \cdot \kappa(\theta)^{-1}$$

where $\theta = \log\left(\frac{p}{1-p}\right)$.

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'. $\qquad\square$

Example: $\mathcal{G}(n, p)$ belongs to the ERGM class.

Lemma
$\mathcal{G}(n, p)$ *is identical with the ERGM defined by*

$$P_1(G) = \exp\left[\theta \cdot m(G)\right] \cdot \kappa(\theta)^{-1}$$

where $\theta = \log\left(\frac{p}{1-p}\right)$.

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'. $\qquad\square$

Example: $\mathcal{G}(n, p)$ belongs to the ERGM class.

Lemma

$\mathcal{G}(n, p)$ *is identical with the ERGM defined by*

$$P_1(G) = \exp\left[\theta \cdot m(G)\right] \cdot \kappa(\theta)^{-1}$$

where $\theta = \log\left(\frac{p}{1-p}\right)$.

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'. $\qquad\square$

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
$\mathcal{G}(n, p)$ *is identical with the ERGM defined by*

$$P_1(G) = \exp\left[\theta \cdot m(G)\right] \cdot \kappa(\theta)^{-1}$$

where $\theta = \log\left(\frac{p}{1-p}\right)$.

Relation 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
$\mathcal{G}(n, p)$ *is identical with the ERGM defined by*

$$P_1(G) = \exp\left[\theta \cdot m(G)\right] \cdot \kappa(\theta)^{-1}$$

where $\theta = \log\left(\frac{p}{1-p}\right)$.

Relation 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
$\mathcal{G}(n, p)$ *is identical with the ERGM defined by*

$$P_1(G) = \exp\left[\theta \cdot m(G)\right] \cdot \kappa(\theta)^{-1}$$

where $\theta = \log\left(\frac{p}{1-p}\right)$.

Relation between θ and *p*

- $\theta < 0 \Longleftrightarrow$ expected density $p < 1/2$;
- $\theta = 0 \Longleftrightarrow$ expected density $p = 1/2$;
- $\theta > 0 \Longleftrightarrow$ expected density $p > 1/2$.

Does not hold in general (if the ERGM contains other statistics).

Commonly used network statistics.

Commonly used statistics *g* count the number of specific subgraphs in the network.

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

If a subgraph count is associated with a positive (negative) parameter, then those subgraphs become more (less) frequent.

Commonly used network statistics.

Commonly used statistics *g* count the number of specific subgraphs in the network.

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

If a subgraph count is associated with a positive (negative) parameter, then those subgraphs become more (less) frequent.

Commonly used network statistics (I).

Statistic $m(G)$ counts the number of **edges**.

A positive (negative) parameter associated with $m(G)$ increases (decreases) the expected density.

Commonly used network statistics (II).

Assume that actors have attribute values $a\colon V \to \{1, \ldots, c\}$, such as age, gender, nationality, religion, ...

Let statistic

$$m_a(G) = |\{\{u, v\} \in E \,;\; a(u) = a(v)\}|$$

count the number of edges connecting actors with the same attribute value.

A positive (negative) parameter associated with $m_a(G)$ models tendency for (against) creating edges to similar actors **homophily (heterophily)**.
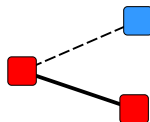
Commonly used network statistics (III).

Statistic $t(G)$ counts the number of **triangles** in G.

A positive (negative) parameter models a preference (reluctance) to close triangles (transitivity).

"*A friend of a friend is a friend.*"
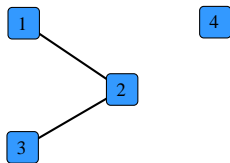
Commonly used network statistics (IV).

For $\ell = 2, \ldots, n-1$ statistic $s_\ell(G)$ counts the number of ℓ-**stars**.

A positive (negative) parameter models the tendency for (against) connecting to high-degree vertices.
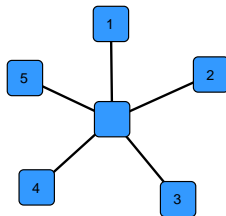
Note: a vertex of degree d contributes $\binom{d}{\ell}$ to the ℓ-star count.

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.

Estimation of ERGM parameters.

Given an observed network G_{obs} and a set of statistics g_i, $i = 1, \ldots, k$.

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G) \right)$$

The *maximum likelihood estimate* of the parameters is the vector $\hat{\theta} \in \mathbb{R}^k$ that maximizes the likelihood function

$$L : \mathbb{R}^k \to \mathbb{R}; \ \theta \mapsto P_\theta(G_{obs}) \ .$$

Estimation can be done with the R function `ergm`.

Outline.

Given an ERGM $(\mathcal{G}, P_\theta)$ with

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right) \quad ,$$

we want to design a probabilitstic algorithm

- returning at each call a graph G from $\mathcal{G}$,
- with probability defined by $P_\theta(G)$.

First try: factorize the probability.

Let $D = \{d_1, \ldots, d_M\}$ be the set of dyads in an arbitrary but fixed order.

For a given graph $G = (V, E)$ let $E_i = E \cap \{d_1, \ldots, d_i\}$ and $\overline{E}_i = \{d_j \in \{d_1, \ldots, d_i\} ; \; d_j \notin E\}$.

For two disjoint subsets $E, \overline{E} \subseteq D$ let

$$\mathcal{G}_{E,\overline{E}} = \{G = (V_G, E_G) \in \mathcal{G} ; \; E \subseteq E_G \text{ and } \overline{E} \cap E_G = \emptyset\} .$$

Then, for a given graph G it is

$$P(G) = \prod_{d_i \in E} P(\mathcal{G}_{d_i} | \mathcal{G}_{E_{i-1}, \overline{E_{i-1}}}) \cdot \prod_{d_i \in D \setminus E} 1 - P(\mathcal{G}_{d_i} | \mathcal{G}_{E_{i-1}, \overline{E_{i-1}}})$$

First try: factorize the probability.

Let $D = \{d_1, \ldots, d_M\}$ be the set of dyads in an arbitrary but fixed order.

For a given graph $G = (V, E)$ let $E_i = E \cap \{d_1, \ldots, d_i\}$ and $\overline{E_i} = \{d_j \in \{d_1, \ldots, d_i\}\,;\; d_j \notin E\}$.

For two disjoint subsets $E, \overline{E} \subseteq D$ let

$$\mathcal{G}_{E,\overline{E}} = \{G = (V_G, E_G) \in \mathcal{G}\,;\; E \subseteq E_G \text{ and } \overline{E} \cap E_G = \emptyset\}\ .$$

Then, for a given graph G it is

$$P(G) = \prod_{d_i \in E} P(\mathcal{G}_{d_i} | \mathcal{G}_{E_{i-1}, \overline{E_{i-1}}}) \cdot \prod_{d_i \in D \setminus E} 1 - P(\mathcal{G}_{d_i} | \mathcal{G}_{E_{i-1}, \overline{E_{i-1}}})$$

Sampling from an ERGM: first try.

For a given graph G it is

$$P(G) = \prod_{d_i \in E} P(\mathcal{G}_{d_i} | \mathcal{G}_{E_{i-1}, \overline{E_{i-1}}}) \cdot \prod_{d_i \in D \setminus E} 1 - P(\mathcal{G}_{d_i} | \mathcal{G}_{E_{i-1}, \overline{E_{i-1}}})$$

Sample from a given ERGM:

$E \leftarrow \emptyset; \ \overline{E} \leftarrow \emptyset$
for $i = 1, \ldots, M$ **do**
    with probability $P(\mathcal{G}_{d_i} | \mathcal{G}_{E, \overline{E}})$
      put d_i into E;
    otherwise
      put d_i into $\overline{E}$.

Problem: probabilities are computationally intractable.

Sampling from an ERGM: first try.

For a given graph G it is

$$P(G) = \prod_{d_i \in E} P(\mathcal{G}_{d_i}|\mathcal{G}_{E_{i-1}, \overline{E_{i-1}}}) \cdot \prod_{d_i \in D \setminus E} 1 - P(\mathcal{G}_{d_i}|\mathcal{G}_{E_{i-1}, \overline{E_{i-1}}})$$

Sample from a given ERGM:

$E \leftarrow \emptyset$; $\overline{E} \leftarrow \emptyset$
for $i = 1, \ldots, M$ **do**
    with probability $P(\mathcal{G}_{d_i}|\mathcal{G}_{E, \overline{E}})$
      put d_i into E;
    otherwise
      put d_i into $\overline{E}$.

Problem: probabilities are computationally intractable.

Conditional probabilities are computationally intractable in general.

Probability of a graph G in an ERGM:

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

Normalizing constant κ cancels out when computing conditional probabilities

$$P(\mathcal{G}_{d_i} | \mathcal{G}_{E_{i-1}, \overline{E_{i-1}}}) \ .$$

But there are $M - i + 1$ unconstrained dyads in $\mathcal{G}_{E_{i-1}, \overline{E_{i-1}}}$.

Computationally intractable, unless $M - i + 1$ is very small; that is, if i is almost as large as M.

Sampling from an ERGM: second try.

For a given graph $G = (V, E)$ and a dyad d define

$$\mathcal{G}[G^{(-d)}] = \{(V, E \setminus \{d\}), (V, E \cup \{d\})\} \ ,$$

(the set of two graphs that are identical with G on all dyads except d).

Sample from a given ERGM:

$E \leftarrow \emptyset$
for $i = 1, \ldots, M$ **do**
    with probability $P(\mathcal{G}_{d_i}|\mathcal{G}[(V, E)^{(-d_i)}])$
       put d_i into E;
    otherwise
       do not put d_i into E (i. e., do nothing).

Problem: graphs are not returned with the correct probabilities.

For a given graph $G = (V, E)$ and a dyad d define

$$\mathcal{G}[G^{(-d)}] = \{(V, E \setminus \{d\}), (V, E \cup \{d\})\} \ ,$$

(the set of two graphs that are identical with G on all dyads except d).

Sample from a given ERGM:

$E \leftarrow \emptyset$
for $i = 1, \ldots, M$ **do**
   with probability $P(\mathcal{G}_{d_i} | \mathcal{G}[(V, E)^{(-d_i)}])$
     put d_i into E;
   otherwise
     do not put d_i into E (i. e., do nothing).

Problem: graphs are not returned with the correct probabilities.

Sampling from an ERGM: second try.

For a given graph $G = (V, E)$ and a dyad d define

$$\mathcal{G}[G^{(-d)}] = \{(V, E \setminus \{d\}), (V, E \cup \{d\})\} \ ,$$

(the set of two graphs that are identical with G on all dyads except d).

Sample from a given ERGM:

$E \leftarrow \emptyset$
for $i = 1, \ldots, M$ **do**
    with probability $P(\mathcal{G}_{d_i} | \mathcal{G}[(V, E)^{(-d_i)}])$
       put d_i into E;
    otherwise
       do not put d_i into E (i.e., do nothing).

Problem: graphs are not returned with the correct probabilities.

Sampling from an ERGM: yet another try.

Sample from a given ERGM:

start with some arbitrary graph (V, E)
for *some number of steps T* **do**
  draw a random dyad $d \in D$
  with probability $P(\mathcal{G}_d | \mathcal{G}[(V, E)^{(-d)}])$
    put d into E;
  otherwise
    remove d from E.

Fact: graphs are still not returned with the correct probabilities.

But probability converges to the correct probability when
$T \to \infty$.

That's what we are going to show in this section.

Sampling from an ERGM: yet another try.

Sample from a given ERGM:

start with some arbitrary graph (V, E)
for *some number of steps T* **do**
    draw a random dyad $d \in D$
    with probability $P(\mathcal{G}_d | \mathcal{G}[(V, E)^{(-d)}])$
      put d into E;
    otherwise
      remove d from E.

Fact: graphs are still not returned with the correct probabilities.

But probability converges to the correct probability when $T \to \infty$.

That's what we are going to show in this section.

Giving it a different name: Markov chain simulation.

Informally, 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 ∞.

⇒ Simulate many steps and return the current graph.

Giving it a different name: Markov chain simulation.

Informally, 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.

Giving it a different name: Markov chain simulation.

Informally, 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.

Giving it a different name: Markov chain simulation.

Informally, 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.
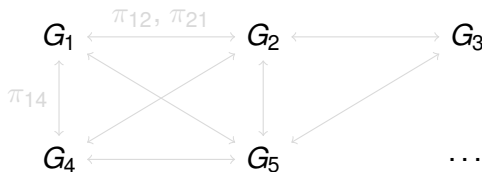
Giving it a different name: Markov chain simulation.

Informally, 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.

How to define the transition probabilities.

Goal: given an ERGM $(\mathcal{G}, P)$

- define transition probabilities π on the set of graphs $\mathcal{G}$ 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 ∞.

An important observation.

π_{ij} interpreted as the probability to jump from state G_i to G_j

$$P(G_1) \xrightarrow{\pi_{12}} P(G_2) \xleftarrow{\pi_{32}} P(G_3)$$

$$\nearrow^{\pi_{42}} \quad \uparrow^{\pi_{52}}$$

$$P(G_4) \qquad P(G_5) \qquad \cdots$$
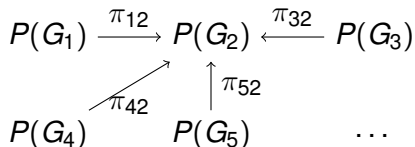
If the probability to be on a graph G after t iteration steps is
denoted by $P^{(t)}(G)$, then
(with $P^{(t)} = [P^{(t)}(G_1), \ldots, P^{(t)}(G_N)] \in \mathbb{R}^N$) it is

$$P^{(t+1)} = P^{(t)} \pi \ , \text{(matrix-vector multiplication)}.$$

This holds since for any $j = 1, \ldots, N$ it is

$$P^{(t+1)}(G_j) = \sum_{i=1}^{N} P^{(t)}(G_i) \pi_{ij}$$

Stationary state space distributions.

A probability vector P can only be a limit of the Markov chain if it is a fix-point or the mapping $P \mapsto P\pi$.

$$P(G_1) \xrightarrow{\pi_{12}} P(G_2) \xleftarrow{\pi_{32}} P(G_3)$$

$$\nearrow_{\pi_{42}} \qquad \uparrow \pi_{52}$$

$$P(G_4) \qquad P(G_5) \qquad \cdots$$

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

$$P = P\pi \ ,$$

i. e., P is an eigenvector of π with eigenvalue one.

Stationary state space distributions.

A probability vector P can only be a limit of the Markov chain if it is a fix-point or the mapping $P \mapsto P\pi$.

$$P(G_1) \xrightarrow{\pi_{12}} P(G_2) \xleftarrow{\pi_{32}} P(G_3)$$

$$\nearrow_{\pi_{42}} \quad \uparrow^{\pi_{52}}$$

$$P(G_4) \qquad P(G_5) \qquad \cdots$$

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

$$P = P\pi \ ,$$

i. e., P is an eigenvector of π with eigenvalue one.

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.

Irreducible and aperiodic Markov chains.

π_{ij} interpreted as the probability to jump from state G_i to G_j.
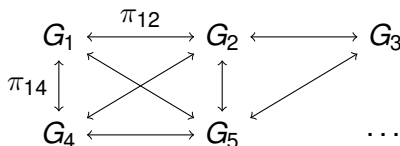
A sequence of states $G_{i_1}, G_{i_2}, \ldots, G_{i_k}$ is called a *(directed) path* if for all $j = 1, \ldots, k-1$ it is $\pi_{i_j i_{j+1}} > 0$.

Definition
The Markov chain $(\mathcal{G}, \pi)$ is called

- *irreducible* if for any two states $G_i, G_j \in \mathcal{G}$ there is a path from G_i to G_j;
- *aperiodic* if the greatest common divisor of the length of all cycles (i. e., paths from a state to itself) equals one.

Irreducible and aperiodic Markov chains.

π_{ij} interpreted as the probability to jump from state G_i to G_j.
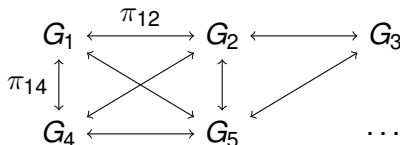
A sequence of states $G_{i_1}, G_{i_2}, \ldots, G_{i_k}$ is called a *(directed) path*
if for all $j = 1, \ldots, k - 1$ it is $\pi_{i_j i_{j+1}} > 0$.

Definition
The Markov chain $(\mathcal{G}, \pi)$ is called

- ▶ *irreducible* if for any two states $G_i, G_j \in \mathcal{G}$ there is a path
  from G_i to G_j;
- ▶ *aperiodic* if the greatest common divisor of the length of all
  cycles (i. e., paths from a state to itself) 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

$$\lim_{K \to \infty} P' \pi^K = P \ .$$

Conditions will be 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

$$\lim_{K \to \infty} P'\pi^K = P .$$

Conditions will be 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, aperiodic matrix A is an eigenvalue of multiplicity one, all entries of an associated eigenvector are non-zero and have the same sign, and the absolute values of all smaller eigenvalues are strictly smaller than ρ.

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.

P is an eigenvector of π with eigenvalue one.

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.

P is an eigenvector of π with eigenvalue one.

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.

The spectral radius ρ of π is one.

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 $\sum_{i=1}^{N} x_i \neq 0$, it must be $\rho = 1$.

The spectral radius ρ of π is one.

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 $\sum_{i=1}^{N} x_i \neq 0$, it must be $\rho = 1$.

The spectral radius ρ of π is one.

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 $\sum_{i=1}^{N} x_i \neq 0$, it must be $\rho = 1$.

The spectral radius ρ of π is one.

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 $\sum_{i=1}^{N} x_i \neq 0$, it must be $\rho = 1$.

The spectral radius ρ of π is one.

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 $\sum_{i=1}^{N} x_i \neq 0$, 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

$$\pi_{ij} \ = \ \frac{P(G_j)}{\binom{n}{2}(P(G_i) + P(G_j))} \ .$$

▶ $\pi_{ii} = \sum \frac{P(G)}{\binom{n}{2}(P(G_i)+P(G))}$

(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

$$\pi_{ij} = \frac{P(G_j)}{\binom{n}{2}(P(G_i) + P(G_j))} .$$

- $\pi_{ii} = \sum \frac{P(G)}{\binom{n}{2}(P(G_i)+P(G))}$

  (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

$$\pi_{ij} = \frac{P(G_j)}{\binom{n}{2}(P(G_i) + P(G_j))} .$$

- $\pi_{ij} = \sum \frac{P(G_i)}{\binom{n}{2}(P(G_i)+P(G))}$

  (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 (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 $g_\ell(G)$.

Outline.

Estimating the parameters of ERGMs

The general form of the class of ERGMs is:

$$P_\theta(G) = \frac{1}{\kappa(\theta)} exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

where:

- $g_i : \mathcal{G} \to \mathbb{R}$ for $i = 1, \ldots, k$ are statistics
- θ_i for $i = 1, \ldots, k$ are parameters
- $\kappa(\theta)$ is a normalizing constant defined by:

$$\kappa(\theta) = \sum_{G' \in \mathcal{G}} exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G')\right)$$

Estimating the parameters of ERGMs

The general form of the class of ERGMs is:

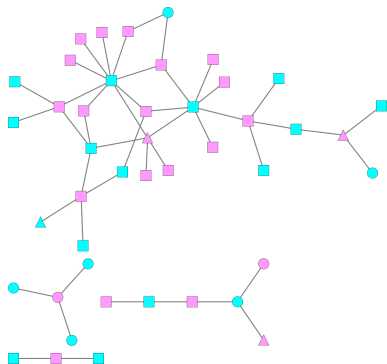$$P_\theta(G) = \frac{1}{\kappa(\theta)} exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

where:

- $g_i \colon \mathcal{G} \to \mathbb{R}$ for $i = 1, \ldots, k$ are statistics
- θ_i for $i = 1, \ldots, k$ are parameters
- $\kappa(\theta)$ is a normalizing constant defined by:

$$\kappa(\theta) = \sum_{G' \in \mathcal{G}} exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G')\right)$$

**How to specify the vector of statistics g
given an observed graph $G_{obs} \in \mathcal{G}$?**

Estimating the parameters of ERGMs

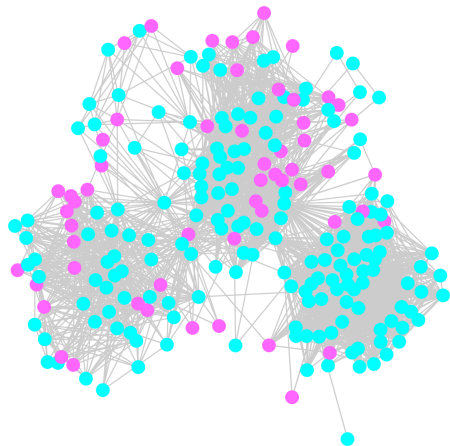**How to specify the vector of statistics g given an observed graph $G_{obs} \in \mathcal{G}$?**

▶ Gender by color:
- pink = female
- turquoise = male

▶ Race by shape:
- square = white
- circle = black
- triangle = other

"Grey's Anatomy Network of Sexual Relations" collected by Gary Weissman

Looking at "patterns" in the graph...

Estimating the parameters of ERGMs

How to specify the vector of statistics g given an observed graph $G_{obs} \in \mathcal{G}$?
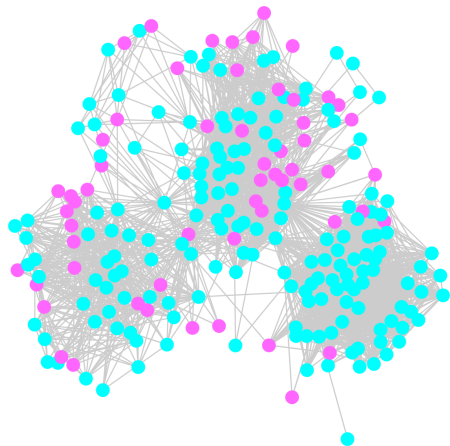

Friendship in FB network data

Gender by color:

- pink = female
- turquoise = male

... does not always help, ...

Estimating the parameters of ERGMs

How to specify the vector of statistics g given an observed graph $G_{obs} \in \mathcal{G}$?

Friendship in FB network data

Gender by color:

- pink = female
- turquoise = male

..., but theory related to the phenomenon under study can help!

Estimating the parameters of ERGMs

How to specify the vector of statistics g given an observed graph $G_{obs} \in \mathcal{G}$?

Conclusions:

- *it is not simple to specify which effects might be relevant for explaining*
  an observed graph G based solely on visual inspection!

- *a **theory-guided approach** is necessary!*

  Statistics are "translations" of hypotheses concerning the local social processes that might have generated an observed graph.

  ERGMs allow researchers to *test* their hypotheses:

  do data (i.e., the observed graph) support or are against researchers' hypotheses?

Estimating the parameters of ERGMs

The general form of the class of ERGMs is:

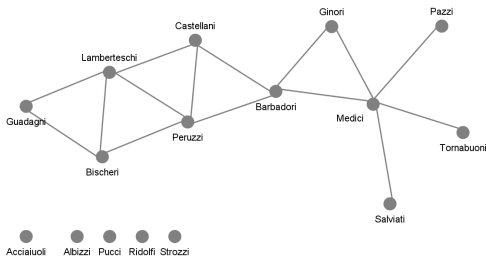$$P_\theta(G) = \frac{1}{\kappa(\theta)} exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

where:

- $g_i : \mathcal{G} \to \mathbb{R}$ for $i = 1, \ldots, k$ are statistics
- θ_i for $i = 1, \ldots, k$ are parameters
- $\kappa(\theta)$ is a normalizing constant defined by:

$$\kappa(\theta) = \sum_{G' \in \mathcal{G}} exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G')\right)$$

**How to estimate and to interpret the vector of parameters θ
given an observed graph $G \in \mathcal{G}$?**
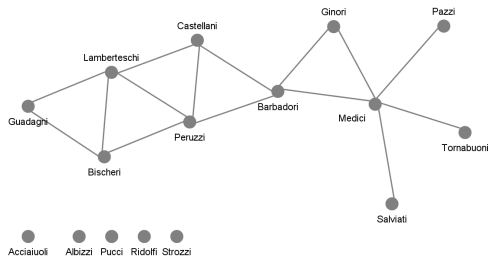
Estimating the parameters of ERGMs
Example

John Padgett Florentine families data set

	Estimate	Std.Error	p-value
edges	-4.21	1.23	0.000837
triangle	1.31	0.37	0.044581
kstar2	1.04	0.65	0.134105
kstar3	-0.63	0.41	0.128747

Estimating the parameters of ERGMs
Example

John Padgett Florentine families data set

	Estimate	Std.Error	p-value
edges	-4.21	1.23	0.000837
triangle	1.31	0.37	0.044581
kstar2	1.04	0.65	0.134105
kstar3	-0.63	0.41	0.128747

Estimating the parameters of ERGMs

Definition

Let $(\mathcal{G}, \theta)$ be a random graph model parametrized by $\theta \in \Theta \subseteq \mathbb{R}^k$ and G_{obs} an observed graph.

The *likelihood function* associated with G_{obs} is defined by:

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

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

$$\widehat{\theta} = arg \max_{\theta \in \theta} L(\theta)$$

is called maximum likelihood estimate (MLE) for θ.

Estimating the parameters of ERGMs

Therefore, estimating the vector of parameters θ of ERGMs requires solving
the following optimization problem:

$$
\begin{aligned}
\widehat{\theta} &= arg \max_{\theta \in \theta} L(\theta) = arg \max_{\theta \in \theta} P_\theta(G_{obs}) = \\
&= arg \max_{\theta \in \theta} \frac{1}{\kappa(\theta)} exp \left(\sum_{i=1}^{k} \theta_i \cdot g_i(G) \right)
\end{aligned}
$$

or equivalently:

$$
\begin{aligned}
\widehat{\theta} &= arg \max_{\theta \in \theta} log L(\theta) \\
&= arg \max_{\theta \in \theta} \left(\sum_{i=1}^{k} \theta_i \cdot g_i(G) \right) - ln \left[\kappa(\theta) \right]
\end{aligned}
$$

which is not simple!!!

Interpreting the parameters of ERGMs

Something to keep in mind:

1. Statistics are "translations" of hypotheses concerning a particular phenomenon into local configurations of a graph

2. A parameter θ_i is a measure of the "presence" of a specific local configuration $s_i(G)$ in the observed graph:

   ▶ $\theta_i > 0$ the number of $s_i(G)$ is higher than that expected by chance
   $\Rightarrow$ evidence towards the hypothesis

   ▶ $\theta_i < 0$ the number of $s_i(G)$ is lower than that expected by chance
   $\Rightarrow$ evidence against the hypothesis

   ▶ $\theta_i = 0$ the number of $s_i(G)$ is equal to that expected by chance
   $\Rightarrow$ evidence that the local process involved in the hypothesis has
   no role in the process of tie formation

When a parameter is statistically different from 0?
We need a statistical test!!!

Estimating the parameters of ERGMs
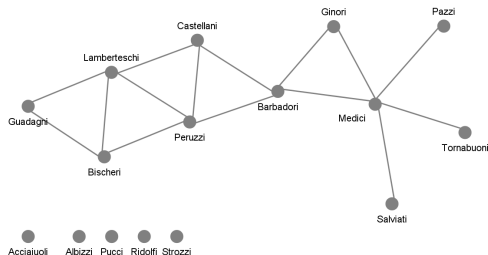
Example

	Estimate	Std.Error	p-value
edges	-4.21	1.23	0.000837
triangle	1.31	0.37	0.044581
kstar2	1.04	0.65	0.134105
kstar3	-0.63	0.41	0.128747

Hypothesis test:

1. State the hypotheses

    - The *null hypothesis* (H_0) states that the observed number of $s_i(G)$ is equal to that expected by chance

    $$H_0 : \theta_i = 0$$

    - The *alternative hypothesis* (H_1) states that number of $s_i(G)$ is influenced by some non-random cause
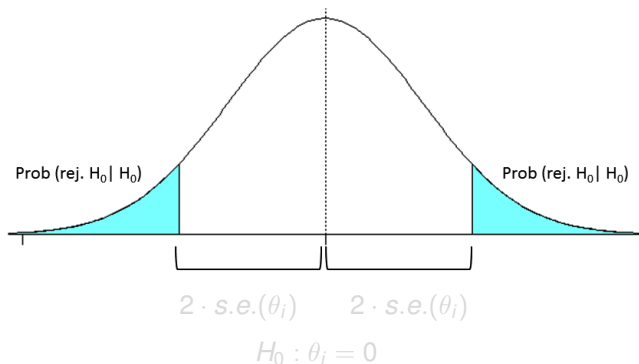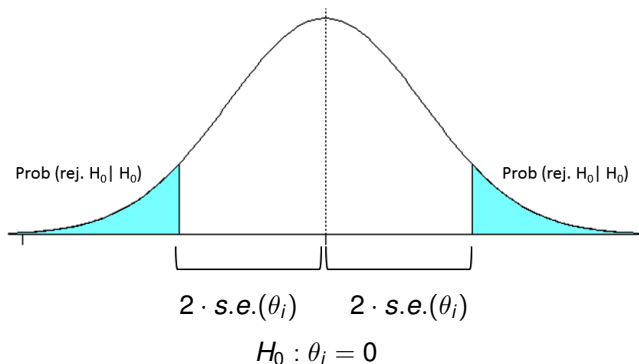
    $$H_1 : \theta_i \neq 0$$

Testing and interpreting the parameters of ERGMs

2. Decision rule (1):

$$|z_{obs}| = \left| \frac{\theta_i}{s.e.(\theta_i)} \right| \geq 2 \qquad \textit{reject } H_0$$

$$|z_{obs}| = \left| \frac{\theta_i}{s.e.(\theta_i)} \right| < 2 \qquad \textit{fail to reject } H_0$$

Prob (rej. $H_0|$ H_0)

Prob (rej. $H_0|$ H_0)

$2 \cdot s.e.(\theta_i)$ $\qquad$ $2 \cdot s.e.(\theta_i)$

$H_0 : \theta_i = 0$

Testing and interpreting the parameters of ERGMs

2. Decision rule (1):

$$|z_{obs}| = \left| \frac{\theta_i}{s.e.(\theta_i)} \right| \geq 2 \qquad \text{reject } H_0$$

$$|z_{obs}| = \left| \frac{\theta_i}{s.e.(\theta_i)} \right| < 2 \qquad \text{fail to reject } H_0$$

Prob (rej. H_0| H_0)   Prob (rej. H_0| H_0)

$$2 \cdot s.e.(\theta_i) \qquad 2 \cdot s.e.(\theta_i)$$

$$H_0 : \theta_i = 0$$

Testing and interpreting the parameters of ERGMs
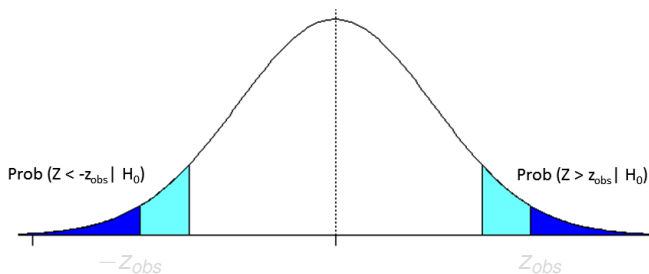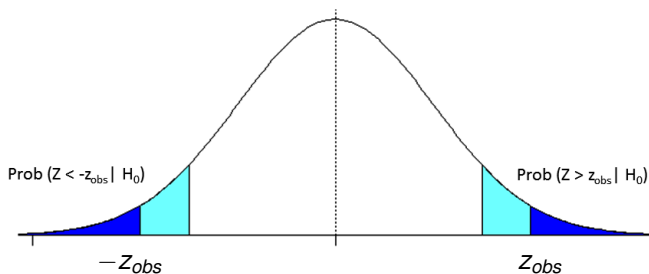
2. Decision rule (2) based on

$$p - value = P(|Z| > z_{obs}|H_0)$$

Then,

$$p - value < 0.05 \quad \text{reject } H_0$$

$$p - value > 0.05 \quad \text{fail to reject } H_0$$

Prob $(Z < -z_{obs}| H_0)$

Prob $(Z > z_{obs}| H_0)$

$-z_{obs}$

z_{obs}

$H_0 : \theta_i = 0$

Testing and interpreting the parameters of ERGMs

2. Decision rule (2) based on

$$p - value = P(|Z| > z_{obs}|H_0)$$

Then,

$$p - value < 0.05 \quad \textit{reject } H_0$$

$$p - value > 0.05 \quad \textit{fail to reject } H_0$$

Prob $(Z < -z_{obs}| H_0)$

Prob $(Z > z_{obs}| H_0)$

$-z_{obs}$

z_{obs}

$$H_0 : \theta_i = 0$$

Outline.

Near-degeneracy and multi-modality of ERGMs

An ERGM $(\mathcal{G}, P_\theta)$ is *near-degenerate* if it places all the probability mass on a small subset of $\mathcal{G}$

Examples

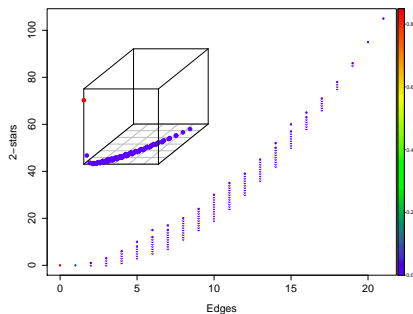$$P_\theta(G) = \frac{1}{\kappa(\theta)} exp\left(\eta m(G) + \sigma_2 s_2(G)\right) \qquad (\eta, \sigma_2) = (-2, -0.2)$$

Near-degeneracy and multi-modality of ERGMs

An ERGM $(\mathcal{G}, P_\theta)$ is *near-degenerate* if it places all the probability mass on a small subset of $\mathcal{G}$

Examples

$$P_\theta(G) = \frac{1}{\kappa(\theta)} exp\left(\eta m(G) + \sigma_2 s_2(G)\right) \qquad (\eta, \sigma_2) = (1, 0.2)$$
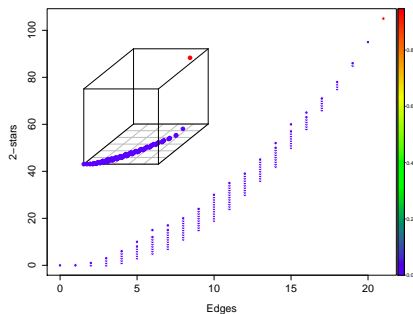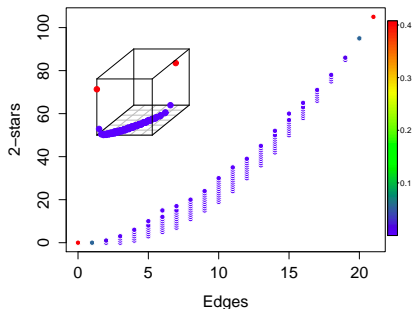
Near-degeneracy and multi-modality of ERGMs

An ERGM $(\mathcal{G}, P_\theta)$ is *near-degenerate* if it places all the probability mass on a small subset of $\mathcal{G}$

Examples

$$P_\theta(G) = \frac{1}{\kappa(\theta)} exp\left(\eta m(G) + \sigma_2 s_2(G)\right) \qquad (\eta, \sigma_2) = (-2, 0.4)$$

Near-degeneracy and multi-modality of ERGMs

Consequences of near-degeneracy

1. The Markov chain for parameter estimations and simulations converges very slow towards the stationary distribution

$$\Downarrow$$

Parameter estimates are not reliable

2. The distribution concentrates on "uninteresting" network configurations, e.g. near-empty or near-complete graphs

How to overcome near-degeneracy???

Near-degeneracy and multi-modality of ERGMs
Overcoming near-degeneracy

1. Improve the convergence of the Markov chain via different updating rules (e.g., Metropolis-Hastings algorithm vs. Gibbs sampling)

   $\Downarrow$ but

   the resulting ERGMs are still inadequate for empirical networks

2. Conditioning on some network characteristics (e.g., $m(G_{obs})$)

   $\Downarrow$ but

   there are still many datasets for which satisfactory parameters value cannot be obtained

3. New specifications for the ERGMs

Near-degeneracy and multi-modality of ERGMs
Overcoming near-degeneracy

1. Improve the convergence of the Markov chain via different updating rules (e.g., Metropolis-Hastings algorithm vs. Gibbs sampling)

$$\Downarrow \text{ but}$$

the resulting ERGMs are still inadequate for empirical networks

2. Conditioning on some network characteristics (e.g., $m(G_{obs})$)

$$\Downarrow \text{ but}$$

there are still many datasets for which satisfactory parameters value cannot be obtained

3. New specifications for the ERGMs

Near-degeneracy and multi-modality of ERGMs
Overcoming near-degeneracy

1. Improve the convergence of the Markov chain via different updating rules (e.g., Metropolis-Hastings algorithm vs. Gibbs sampling)

$$\Downarrow \text{ but}$$

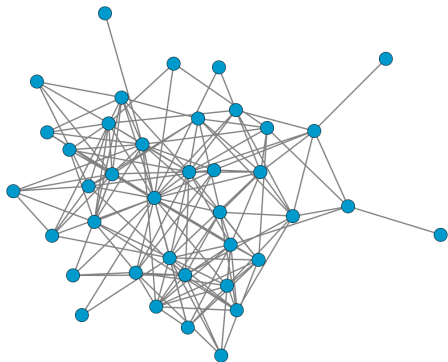the resulting ERGMs are still inadequate for empirical networks

2. Conditioning on some network characteristics (e.g., $m(G_{obs})$)

$$\Downarrow \text{ but}$$

there are still many datasets for which satisfactory parameters value cannot be obtained

3. New specifications for the ERGMs

Near-degeneracy and multi-modality of ERGMs

Interactions among 39 workers in a tailor shop in Zambia

An interaction is: "a continuous uninterrupted social activity involving the participation of at least two persons"

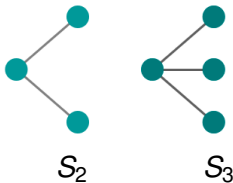"Kapferer's tailor shop data"
collected by Bruce Kapferer in Zambia

$$P(G_{obs}) = \frac{1}{\kappa(\theta)} exp(\eta m(G) + \sigma_2 s_2(G) + \sigma_3 s_3(G) + \tau t(G))$$

is a degenerate model!!!

Near-degeneracy and multi-modality of ERGMs

New specifications for ERGMS

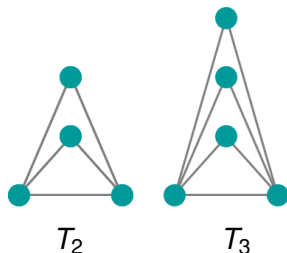1. *Alternating k-stars* (GWDEGREE)

$$s_{altS}(G) = \sum_{k=2}^{n-1} (-1)^k \frac{S_k(G)}{\lambda^{k-2}}$$

- represents the distribution of the degree
- balance between positive and negative stars parameters to prevent empty or complete graphs
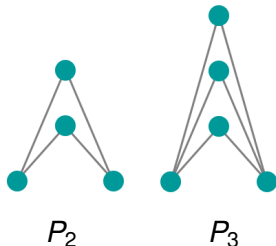
2. Alternating k-triangles (GWESP/ESP)

$$s_{altT}(G) = \sum_{k=2}^{n-1} (-1)^k \frac{T_k(G)}{\lambda^{k-2}}$$

T_2 T_3

- number of shared partners of adjacent actors
- controls the tendency to have many cohesive subsets of 3 or more nodes
- N.b.
  requires a partial dependence assumption
  $\Rightarrow$ the resulting graph is no longer a Markov graph!!!

Near-degeneracy and multi-modality of ERGMs

New specifications for ERGMS

3. Alternating k-two-paths (GWDSP/DSP)

$$s_{altP}(G) = P_1(G) - \frac{2P_2(G)}{\lambda} +$$
$$\sum_{k=3}^{n-1} \left(-\frac{1}{\lambda}\right)^{k-1} P_k(G)$$

P_2 P_3

- number of shared partners of two actors
- models the pre-conditions for forming triangles
  $\Rightarrow$ Used in conjunction with the *alternating k-triangles*
- N.b.
  requires a partial dependence assumption
  $\Rightarrow$ the resulting graph is no longer a Markov graph!!!

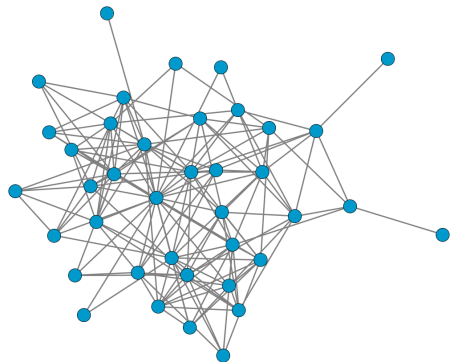# Near-degeneracy and multi-modality of ERGMs

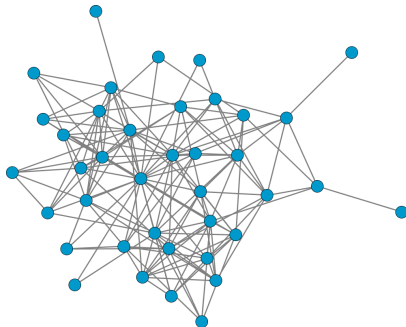Interactions among 39 workers in a tailor shop in Zambia

An interaction is: "a continuous uninterrupted social activity involving the participation of at least two persons"

"Kapferer's tailor shop data" collected by Bruce Kapferer in Zambia

$$P(G_{obs}) = \frac{1}{\kappa(\theta)} exp(\eta m(G) + \theta_{altS} s_{altS}(G) + \theta_{altT} s_{altT}(G) + \theta_{altP} s_{altP}(G))$$

is not a degenerate model!!!

Estimating the parameter of ERGMs

	Estimate	Std.Error
edges	-3.779	1.495
alt-k-stars	-0.081	0.409
alt-k-triangles	1.129	0.206
alt-k-2path	0.029	0.054

Outline.

Recall: some statistics make edge probabilities dependent – others not.

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(G)\right)$$

$g_i(G)$		edge prob.
number of edges		independent
edges connecting same attribute		independent
number of triangles		**dependent**
number of ℓ-stars		**dependent**

Conditional independence of edges (informally).

Two dyads d_1 and d_2 are said to be *conditionally independent* (given the rest of the graph) if—under the condition that all other dyads are fixed—the state of the dyad d_2 does not provide any additional information about the probability $P(d_1 \in E)$.
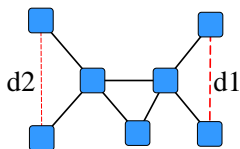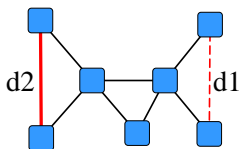
Conditional independence of edges.

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 $\mathcal{G}_{D^+, D^-}$ be defined by

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

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

$$P(\mathcal{G}_{d_1} | \mathcal{G}_{D^+, D^-}) = P(\mathcal{G}_{d_1} | \mathcal{G}_{D^+, D^-} \cap \mathcal{G}_{d_2}) \;.$$

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 $P(d_1 \in E)$.

Hammersley-Clifford Theorem; special case.

Theorem (first part)

Let $\mathcal{G}$ be a set of graphs that is closed under taking subsets of the edge set. Let $(\mathcal{G}, P)$ be a random graph satisfying $P(G) > 0$ for all $G \in \mathcal{G}$ and let D be the set of dyads.

There are constants $\{\alpha_A \in \mathbb{R} \,;\; A \subseteq D\}$, satisfying $\alpha_A = 0$ if A contains two conditionally independent dyads, such that

$$P(G) = \frac{1}{\kappa} \exp \left(\sum_{A \subseteq E(G)} \alpha_A \right) , \qquad where \qquad (1)$$

$$\kappa = \sum_{G' \in \mathcal{G}} \exp \left(\sum_{A \subseteq E(G')} \alpha_A \right) . \qquad (2)$$

Hammersley-Clifford Theorem; special case.

Theorem (first part)

Let $\mathcal{G}$ be a set of graphs that is closed under taking subsets of the edge set. Let $(\mathcal{G}, P)$ be a random graph satisfying $P(G) > 0$ for all $G \in \mathcal{G}$ and let D be the set of dyads.

There are constants $\{\alpha_A \in \mathbb{R};\ A \subseteq D\}$, satisfying $\alpha_A = 0$ if A contains two conditionally independent dyads, such that

$$P(G) = \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A \right) , \qquad where \qquad (1)$$

$$\kappa = \sum_{G' \in \mathcal{G}} \exp\left(\sum_{A \subseteq E(G')} \alpha_A \right) . \qquad (2)$$

Hammersley-Clifford Theorem; special case.

Theorem (second part)

Conversely, if the probability P on $\mathcal{G}$ is defined by

$$P(G) \;=\; \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A \right) \;, \qquad \text{where}$$

$$\kappa \;=\; \sum_{G' \in \mathcal{G}} \exp\left(\sum_{A \subseteq E(G')} \alpha_A \right) \;,$$

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

$$P(G) = \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A\right) , \qquad \text{where}$$

$$\kappa = \sum_{G' \in \mathcal{G}} \exp\left(\sum_{A \subseteq E(G')} \alpha_A\right) ,$$

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

$$P(G) \;=\; \frac{1}{\kappa} \exp \left(\sum_{A \subseteq E(G)} \alpha_A \right) \;.$$

$\Rightarrow$ every random graph $(\mathcal{G}, P)$ 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

$$P(G) \;=\; \frac{1}{\kappa} \exp \left(\sum_{A \subseteq D} \alpha_A \cdot g_A(G) \right) \;.$$

Conclusion from the Hammersley-Clifford Theorem.

There are constants $\{\alpha_A \in \mathbb{R} \, ; \ A \subseteq D\}$, such that

$$P(G) = \frac{1}{\kappa} \exp \left(\sum_{A \subseteq E(G)} \alpha_A \right) \ .$$

$\Rightarrow$ every random graph $(\mathcal{G}, P)$ 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

$$P(G) = \frac{1}{\kappa} \exp \left(\sum_{A \subseteq D} \alpha_A \cdot g_A(G) \right) \ .$$

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.

Example
$\{i, j\}$ and $\{u, v\}$ conditionally
independent;

$\{i, j\}$ and $\{j, u\}$ might be
conditionally dependent;
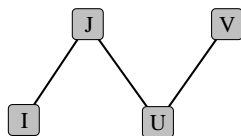
Markov graphs are a specific subclass of the ERGM class.

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 not conditionally independent, given
the rest of the graph.

Example
the dependence graph of a
Markov graph on vertices
$V = \{1, 2, 3, 4\}$ is
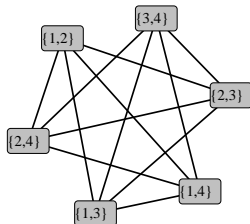
A subset $A \subseteq D$ is a clique in the dependence graph if A does
not contain two conditionally independent dyads.

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*

$$
\begin{aligned}
\eta_{uv} & \quad \text{for all} \quad \text{dyads } \{u, v\} \\
\tau_{uvw} & \quad \text{for all} \quad \text{triangles } \{u, v, w\} \\
\sigma_{uv_1 \ldots v_\ell} & \quad \text{for all} \quad 2 \leq \ell \leq n - 1, \text{ and all} \\
& \quad\quad\quad\quad \ell\text{-stars } (u, \{v_1, \ldots, v_\ell\})
\end{aligned}
$$

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.

Start from the ERGM of a general Markov graph.

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.

Start from the ERGM of a general Markov graph.

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. . .

$\square$

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,

$$P(G) = \frac{1}{\kappa} \exp\left(\eta \cdot m(G)\right) \ .$$

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

$$
P(G) = \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A \right) , \qquad where
$$

$$
\kappa = \sum_{G' \in \mathcal{G}} \exp\left(\sum_{A \subseteq E(G')} \alpha_A \right) .
$$

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.

$$\text{Want to show } P(G) \;=\; \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A \right) \;.$$

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.

For $A \subseteq D$ define

$$\alpha_A := \sum_{B \subseteq A} (-1)^{|A \setminus B|} \log P(G_B) \;.$$

Motivation (Möbius Inversion Theorem)

$$f(A) = \sum_{B \subseteq A} g(B) \Longleftrightarrow g(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} f(B) \;.$$

Proof of the Hammersley-Clifford Theorem.

Want to show $P(G) = \dfrac{1}{\kappa} \exp\left(\displaystyle\sum_{A \subseteq E(G)} \alpha_A \right)$.

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.

For $A \subseteq D$ define

$$\alpha_A := \sum_{B \subseteq A} (-1)^{|A \setminus B|} \log P(G_B) .$$

Motivation (Möbius Inversion Theorem)

$$f(A) = \sum_{B \subseteq A} g(B) \iff g(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} f(B) .$$

Proof of the Hammersley-Clifford Theorem.

Want to show $P(G) = \dfrac{1}{\kappa} \exp\left(\displaystyle\sum_{A \subseteq E(G)} \alpha_A\right)$.

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.

For $A \subseteq D$ define

$$\alpha_A := \sum_{B \subseteq A} (-1)^{|A \setminus B|} \log P(G_B) \ .$$

Motivation (Möbius Inversion Theorem)

$$f(A) = \sum_{B \subseteq A} g(B) \Longleftrightarrow g(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} f(B) \ .$$

Proof of the Hammersley-Clifford Theorem.

Want to show $P(G) = \dfrac{1}{\kappa} \exp\left(\displaystyle\sum_{A \subseteq E(G)} \alpha_A \right)$.

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.

For $A \subseteq D$ define

$$\alpha_A := \sum_{B \subseteq A} (-1)^{|A \setminus B|} \log P(G_B) .$$

Motivation (Möbius Inversion Theorem)

$$f(A) = \sum_{B \subseteq A} g(B) \Longleftrightarrow g(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} f(B) .$$

Proof of the Hammersley-Clifford Theorem.

By definition we have

$$\alpha_A = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \log P(G_B) \ .$$

Möbius Inversion Theorem:

$$f(A) = \sum_{B \subseteq A} g(B) \Longleftrightarrow g(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} f(B) \ .$$

Thus, for $A \subseteq D$ it is

$$\log P(G_A) = \sum_{B \subseteq A} \alpha_B \ .$$

In particular, for $A = E(G)$ we get

$$P(G) = \exp \left(\sum_{B \subseteq E(G)} \alpha_B \right) \ .$$

Proof of the Hammersley-Clifford Theorem.

We have

$$P(G) = \exp\left(\sum_{A \subseteq E(G)} \alpha_A\right) \ .$$

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

$$P(G) = \exp\left(\sum_{A \subseteq E(G)} \alpha_A\right) \ .$$

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.

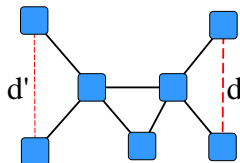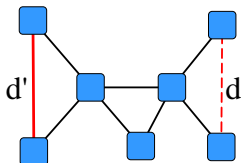# Proof of the Hammersley-Clifford Theorem.

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

$$\frac{P(G_{B \cup \{d,d'\}})}{P(G_{B \cup d'}) + P(G_{B \cup \{d,d'\}})} = \frac{P(G_{B \cup \{d\}})}{P(G_B) + P(G_{B \cup \{d\}})} .$$

which is true if and only if

$$\frac{P(G_{B \cup \{d,d'\}})}{P(G_{B \cup \{d'\}})} = \frac{P(G_{B \cup \{d\}})}{P(G_B)} . \tag{3}$$
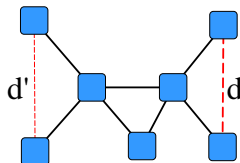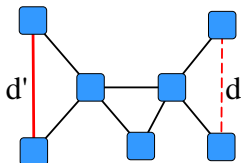
Proof of the Hammersley-Clifford Theorem.

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

$$\frac{P(G_{B \cup \{d, d'\}})}{P(G_{B \cup d'}) + P(G_{B \cup \{d, d'\}})} = \frac{P(G_{B \cup \{d\}})}{P(G_B) + P(G_{B \cup \{d\}})} \ .$$

which is true if and only if

$$\frac{P(G_{B \cup \{d, d'\}})}{P(G_{B \cup \{d'\}})} = \frac{P(G_{B \cup \{d\}})}{P(G_B)} \ . \tag{3}$$

Proof of the Hammersley-Clifford Theorem.

$$\alpha_A = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \log P(G_B) \ .$$

Let $A \subseteq D$, contain two conditionally independent dyads d, d'.

$$
\begin{aligned}
\alpha_A &= \sum_{\substack{B \subseteq A \\ d, d' \in B}} (-1)^{|A \setminus B|} \log P(G_B) + \sum_{\substack{B \subseteq A \\ d \in B, \, d' \notin B}} (-1)^{|A \setminus B|} \log P(G_B) \\
&\quad + \sum_{\substack{B \subseteq A \\ d \notin B, \, d' \in B}} (-1)^{|A \setminus B|} \log P(G_B) + \sum_{\substack{B \subseteq A \\ d, d' \notin B}} (-1)^{|A \setminus B|} \log P(G_B) \\
&= \sum_{B \subseteq A \setminus \{d, d'\}} (-1)^{|A \setminus B|} \log \left(\frac{P(G_{B \cup \{d, d'\}})}{P(G_{B \cup \{d'\}})} \Big/ \frac{P(G_{B \cup \{d\}})}{P(G_B)} \right) \\
&= 0 \ , \text{ follows from (3)}
\end{aligned}
$$

Thus, $\alpha_A = 0$ 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 and d' are conditionally independent in $(\mathcal{G}, P)$, unless there is a subset $A \subseteq D$ with $d, d' \in A$ and $\alpha_A \neq 0$.
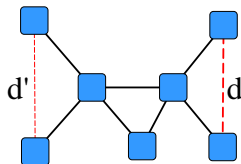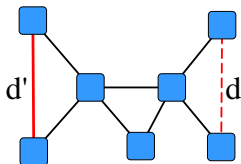
Proof of the Hammersley-Clifford Theorem.

Suppose that d, d' are two dyads such that there is no subset $A \subseteq D$ with $\alpha_A \neq 0$ and $d, d' \in A$.

We show that d and d' are conditionally independent.

Equivalently, for any $B \subseteq D$ with $d, d' \notin B$ it is

$$\frac{P(G_{B \cup \{d, d'\}})}{P(G_{B \cup \{d'\}})} = \frac{P(G_{B \cup \{d\}})}{P(G_B)} \ .$$

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 $d, d' \in A$; $d, d' \notin B$)

$$
\begin{aligned}
\log \left(\frac{P(G_{B \cup \{d,d'\}})}{P(G_{B \cup \{d'\}})} \right) &= \sum_{A \subseteq B \cup \{d,d'\}} \alpha_A - \sum_{A \subseteq B \cup \{d'\}} \alpha_A \\
&= \sum_{\substack{A \subseteq B \cup \{d,d'\} \\ d \in A}} \alpha_A \\
&= \sum_{\substack{A \subseteq B \cup \{d\} \\ d \in A}} \alpha_A \\
&= \sum_{A \subseteq B \cup \{d\}} \alpha_A - \sum_{A \subseteq B} \alpha_A \\
&= \log \left(\frac{P(G_{B \cup \{d\}})}{P(G_B)} \right) \ .
\end{aligned}
$$

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

$$P(G) = \frac{1}{\kappa} \exp \left(\sum_{A \subseteq E(G)} \alpha_A \right) . \tag{4}$$

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 $\alpha_A \neq 0$.