## Network Modeling

# Viviana Amati Jürgen Lerner 

Dept. Computer \& Information Science
University of Konstanz
Winter 2015/2016
(last updated: January 11, 2016)

## Outline.

## Introduction.

Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## Outline.

Introduction.
Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## Topic of this lecture.

Statistical models for social networks.

## Topic of this lecture.

Statistical models for social networks.

Social networks consist of actors and relations among them.

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


Actors and ties may have associated data (attributes).

## Topic of this lecture.

Statistical models for social networks.

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, ...

Want to estimate expected outcome $\pm$ variability
$\Rightarrow$ to explain and predict social relations and behavior.

## Topic of this lecture.

Statistical models for social networks specify probability distributions for social networks.

- How likely is it that two actors are connected by a tie?
- dependent on their attributes
- dependent on their ties to other actors
- How likely is it that an actor has certain attributes?
- dependent on its ties
- dependent on the attributes of its neighbors

Provide a framework for tackling many research questions in the social sciences.

## Example: friendship network among pupils.

Nodes colored by level of delinquent behavior.


Can you see some pattern? Can you find explanations?

## Social influence vs. social selection.

Network ties and actor behavior evolve over time.


Social influence.

- E. g., friends of delinquent pupils become delinquent.

Social selection.

- E. g., delinquent pupils choose delinquent friends.

Dependency among network ties.

- E. g., friends of friends become friends (transitivity).

Correlation of individual attributes.

- E. g., boys are more delinquent.


## 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

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


## Outline.

## Introduction.

Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## Outline.

## Introduction.

Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## Example data: directed friendship network.

A. Knecht (2008): "Friendship Selection and Friends' Influence".

Four time points in the pupils' first year at secondary school.


Constant actor covariates:

- demographics: sex, age, ethnicity, and religion
- assessment of pupil's capacity at the end of primary school

Changing actor covariates (behavior):

- delinquency (stealing, vandalism, graffiti, and fighting)
- alcohol consumption (only at time steps 2-4)

Constant dyadic variable: same primary school

## Two types of dependent variables: ties and behavior.

Friendship tie from $A$ to $B$
can be modeled as a probabilistic function of

- demographics of $A$ and $B$ (social selection);
- behavior of $A$ and $B$ (social selection);
- (non-)existence of tie from $B$ to $A$ (reciprocity);
- other friends of $A$ or $B$ or both (e. g., transitivity)

Delinquent behavior of student $A$
can be modeled as a probabilistic function of

- A's demographics (sex, age, ethnicity, religion);
- behavior of $A$ 's friends (social influence);
- A's friendship ties; ...


## Running hypotheses.

$H_{1}$ Pupils chose friends with the same gender.
$\mathrm{H}_{2}$ Pupils reciprocate friendship.
$\mathrm{H}_{3}$ The friend of a friend is a friend.
$H_{4}$ Pupils chose friends with similar delinquency behavior.
$H_{5}$ Pupils adopt delinquent behavior from their friends.

## First test: social selection by gender.

Hypothesis: Pupils chose friends with the same gender.
More precisely: the probability of friendship between pupils with the same gender is higher.

Method: divide pairs of pupils (dyads) into two classes

$$
\begin{aligned}
& D_{1}=\{(A, B) ; \text { gender }(A)=\text { gender }(B)\} \\
& D_{2}=\{(A, B) ; \text { gender }(A) \neq \text { gender }(B)\}
\end{aligned}
$$

Compare ratio of friendship ties in the two groups.


Result:


VS.


## First test: social selection by gender.

Hypothesis: Pupils chose friends with the same gender.
More precisely: the probability of friendship between pupils with the same gender is higher.

Method: divide pairs of pupils (dyads) into two classes

$$
\begin{aligned}
& D_{1}=\{(A, B) ; \operatorname{gender}(A)=\operatorname{gender}(B)\} \\
& D_{2}=\{(A, B) ; \operatorname{gender}(A) \neq \operatorname{gender}(B)\}
\end{aligned}
$$

Compare ratio of friendship ties in the two groups.

$$
\frac{\# \text { ties in } D_{1}}{\# \text { dyads in } D_{1}} \quad \text { vs. } \quad \frac{\# \text { ties in } D_{2}}{\# \text { dyads in } D_{2}}
$$

## First test: social selection by gender.

Hypothesis: Pupils chose friends with the same gender.
More precisely: the probability of friendship between pupils with the same gender is higher.

Method: divide pairs of pupils (dyads) into two classes

$$
\begin{aligned}
& D_{1}=\{(A, B) ; \operatorname{gender}(A)=\operatorname{gender}(B)\} \\
& D_{2}=\{(A, B) ; \operatorname{gender}(A) \neq \operatorname{gender}(B)\}
\end{aligned}
$$

Compare ratio of friendship ties in the two groups.

$$
\frac{\# \text { ties in } D_{1}}{\# \text { dyads in } D_{1}} \quad \text { vs. } \quad \frac{\# \text { ties in } D_{2}}{\# \text { dyads in } D_{2}}
$$

Result:

$$
\frac{105}{312}=0.3365 \quad \text { vs. } \quad \frac{31}{288}=0.1076
$$

## Significance of observed difference.

0.11 probability for friendship between different gender
0.34 probability for friendship between same gender

Could this difference be just accidental?
If we divided pupils into two meaningless groups, the tie probability would also not be equal.

Repeat the analysis 1000 times with random gender assignment:
$\Rightarrow$ average difference is 0.035 ; maximum is 0.142

## Significance of observed difference.

0.11 probability for friendship between different gender
0.34 probability for friendship between same gender

Could this difference be just accidental?
If we divided pupils into two meaningless groups, the tie probability would also not be equal.

Repeat the analysis 1000 times with random gender assignment:
$\Rightarrow$ average difference is 0.035 ; maximum is 0.142

## Have to control for alternative explanations.

Maybe friendship is only seemingly influenced by gender equality; the "true" explanatory variable might be

- primary school
and boys/girls happen to go more often to the same one
- behavior
and boys/girls have similar behavior to other boys/girls
- other ties in the networks. . .

We need a model that can control for the influence of other variables.

## Modeling the occurence of ties by logistic regression.

Random variable $Y_{u v}$ for tie from $u$ to $v$

$$
\begin{aligned}
& Y_{u v}= \begin{cases}1 & \text { with probability } p_{u v} ; \\
0 & \text { with probability } 1-p_{u v} .\end{cases} \\
& p_{u v}=\text { someFunctionOf(statistics, parameters) }
\end{aligned}
$$

The statistics (explanatory variables) quantify characteristics of the dyad $(u, v)$ in the observed network.
The parameters quantify the influence of those variables on the tie-probability:

- a positive (negative) parameter means: the higher the statistic the higher (lower)the probability;
- a zero parameter means: the statistic has no influence on the tie-probability.
Parameters are estimated from the observed network.


## Modeling the occurence of ties by logistic regression.

Probability $p_{u v}$ of a tie from $u$ to $v$ specified as

$$
\begin{aligned}
p_{u v} & =\operatorname{logit}^{-1}(\theta \cdot s)=\frac{\exp (\theta \cdot s)}{\exp (\theta \cdot s)+1}, \text { where } \\
s & =\left(s_{1}, \ldots, s_{k}\right) \in \mathbb{R}^{k} \quad \text { statistics } \\
\theta & =\left(\theta_{1}, \ldots, \theta_{k}\right) \in \mathbb{R}^{k} \quad \text { parameters } \\
\theta \cdot s & =\sum_{i=1}^{k} \theta_{i} \cdot s_{i}
\end{aligned}
$$

The statistics $s_{i}=s_{i}(u, v ; y)$ are functions of the observed data.
The parameters are estimated to maximize the probability of the observed network $y$ :

$$
P(Y=y)=\prod_{u \neq v} p_{u v}^{y_{u v}} \cdot\left(1-p_{u v}\right)^{1-y_{u v}}
$$

## Results from logistic regression.

Gender model: friendship ties explained by gender-equality

$$
p_{u v}=\operatorname{logit}^{-1}\left(\theta_{1}+\theta_{2} \cdot \operatorname{sameGender}(u, v)\right)
$$

Results:

| statistic | parameter | Std. Error | $\operatorname{Pr}(>\|z\|)$ |
| :--- | ---: | ---: | ---: |
| (Intercept) | -2.1151 | 0.1901 | $<2 \mathrm{e}-16^{* * *}$ |
| sameGender | 1.4363 | 0.2247 | $1.64 \mathrm{e}-10^{* * *}$ |

Implied probability for ties by gender-equality:
$p=0.1076$ for friendship between pupils with different gender
$p=0.3365$ for friendship between pupils with same gender

## Results from logistic regression.

Delinquency model: friendship ties explained by similar behavior

$$
p_{u v}=\operatorname{logit}^{-1}\left(\theta_{1}+\theta_{2} \cdot \text { similarDelinquency }(u, v)\right)
$$

## Results:

| statistic | parameter | Std. Error | $\operatorname{Pr}(>\|\mathrm{z}\|)$ |
| :--- | ---: | ---: | ---: |
| (Intercept) | -1.5880 | 0.1796 | $<2 \mathrm{e}-16^{* * *}$ |
| similarDelinquency | 0.6568 | 0.2619 | $0.0121^{*}$ |

$\operatorname{similarDelinquency}(u, v)=\frac{\Delta-|\operatorname{delin}(u)-\operatorname{delin}(v)|}{\Delta}$,
where $\Delta=$ maximal difference in delinquency

## Results from logistic regression.

More complex model: control for alternative explanations:

$$
p_{u v}=\operatorname{logit}^{-1}\left(\sum_{i=1}^{k} \theta_{i} \cdot s_{i}(u, v ; y)\right), \text { with }
$$

| $s_{i}(u, v ; y)$ | interpretation |
| ---: | :--- |
| 1 | constant (intercept) |
| sameGender $(u, v)$ | gender homophily |
| similarDelinquency $(u, v)$ | behavior homophily |
| $y_{v u}$ | reciprocity |
| $\sum_{w} y_{u w} \cdot y_{w v}$ | transitivity (friend of friend is friend) |

## Results from logistic regression.

More complex model:

$$
p_{u v}=\operatorname{logit}^{-1}\left(\sum_{i=1}^{k} \theta_{i} \cdot s_{i}(u, v ; y)\right)
$$

Results:

| statistic | parameter | Std. Error | $\operatorname{Pr}(>\|z\|)$ |
| :--- | ---: | ---: | :--- |
| (Intercept) | -4.3664379 | 0.3915032 | $<2 \mathrm{e}-16^{* * *}$ |
| sameGender | 1.2644640 | 0.3036323 | $3.12 \mathrm{e}-05^{* * *}$ |
| similarDelinquency | -0.0009412 | 0.3594857 | 0.998 |
| reciprocity | 2.0621869 | 0.2838916 | $3.76 \mathrm{e}-13^{* * *}$ |
| transitivity | 0.9420077 | 0.0918453 | $<2 \mathrm{e}-16^{* * *}$ |

## The analysis so far is invalid.

Logistic regression is only valid for independent observations.


## The analysis so far is invalid.

Logistic regression is only valid for independent observations.

| $y_{u v}$ | sameGender | simDelinquency | $y_{v u}$ | $\sum_{w} y_{u w} \cdot y_{w v}$ |
| :---: | :---: | :---: | :---: | :---: |
| 1 | 1 | 0.5 | 1 | 3 |
| 1 | 1 | 0.5 | 1 | 0 |
| 0 | 0 | 0.1 | 1 | 1 |
| 1 | 0 | 0.1 | 0 | 2 |
| $\ldots$ | $\ldots$ |  |  |  |

In our case, the different observations (rows) are not independent.
This is even implied by the model itself.

## Discrepancy between observation and model.

Randomly drawing ties from the logistic regression model.

simulated network

## Outline.

## Introduction.

Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## Outline.

Introduction.
Running example: data, questions, and simple answers.

## Random graph models.

$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## Probability space (intuition).

A probability space models the likeliness of outcomes of a random experiment (observation, measurement, ...).

Two components:
What could happen? $\Rightarrow$ defines set of possible outcomes.
How likely is an outcome? $\Rightarrow$ defines probability distribution.

## Background: finite probability space.

## Definition

A finite probability space is a pair $(\Omega, P)$, where

- $\Omega$ is a finite set (possible outcomes)
- $P: \Omega \rightarrow[0,1]$ 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 $\Omega^{\prime} \subseteq \Omega$ is defined by $P\left(\Omega^{\prime}\right)=\sum_{\omega \in \Omega^{\prime}} P(\omega)$.


## Background: finite probability space.

## Definition

A finite probability space is a pair $(\Omega, P)$, where

- $\Omega$ is a finite set (possible outcomes)
- $P: \Omega \rightarrow[0,1]$ 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 $\Omega^{\prime} \subseteq \Omega$ is defined by

$$
P\left(\Omega^{\prime}\right)=\sum_{\omega \in \Omega^{\prime}} P(\omega) .
$$

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

## Background: finite probability space.

## Definition

A finite probability space is a pair $(\Omega, P)$, where

- $\Omega$ is a finite set (possible outcomes)
- $P: \Omega \rightarrow[0,1]$ 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 $\Omega^{\prime} \subseteq \Omega$ is defined by $P\left(\Omega^{\prime}\right)=\sum_{\omega \in \Omega^{\prime}} P(\omega)$.

Example (lottery)

$$
\begin{array}{ll}
\Omega=\{X \subset\{1, \ldots, 49\} ;|X|=6\} & \text { (sets of } 6 \text { different numbers) } \\
P(\omega)=\binom{49}{6}^{-1}=\frac{6 \cdot 133!}{49!} \text { for all } \omega \in \Omega & \text { (uniform probability) }
\end{array}
$$

## 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

The elements that can be edges are called dyads.

Interpretation:

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



## Background: adjacency matrices.

## Definition

Let $G=(V, E)$ be a graph with an ordered set of vertices $V=\left\{v_{1}, \ldots, v_{n}\right\}$. The adjacency matrix associated with $G$ is a binary $n \times n$ matrix $y=\left(y_{i j}\right)_{i, j \in 1, \ldots . n}$ with

$$
y_{i j}= \begin{cases}1 & \text { if }\left(v_{i}, v_{j}\right) \in E \\ 0 & \text { if }\left(v_{i}, v_{j}\right) \notin E\end{cases}
$$



|  | $A$ | $B$ | $C$ | $D$ |
| :---: | :---: | :---: | :---: | :---: |
| $A$ | 0 | 0 | 1 | 0 |
| $B$ | 1 | 0 | 0 | 0 |
| $C$ | 1 | 1 | 0 | 0 |
| $D$ | 1 | 0 | 0 | 0 |

What are the possible outcomes of a network experiment (observation, measurement, ...)?

## Assign probabilities to individual ties.

What are the possible outcomes of a network experiment?

We have a set of actors $V=\left\{v_{1}, \ldots, v_{n}\right\}$

- defining a set of dyads (pairs of actors)

$$
D=\left\{\left(v_{i}, v_{j}\right) \in V \times V\right\}
$$

Each of these dyads $\left(v_{i}, v_{j}\right)$ is associated with a observation $y_{i j}$ that has two possible outcomes:

- $y_{i j}=1$ if there is a tie from $v_{i}$ to $v_{j}$;
- $y_{i j}=0$ if there is no tie from $v_{i}$ to $v_{j}$.

Each of the two outcomes happens with a certain probability where

$$
P\left(Y_{i j}=1\right)+P\left(Y_{i j}=0\right)=1
$$

## Assign probabilities to individual ties.

What are the possible outcomes of a network experiment?
Suppose we are given tie probabilities $p_{i j}=P\left(Y_{i j}=1\right)$ for all dyads $\left(v_{i}, v_{j}\right) \in V \times V$.

Does this imply probabilities for higher-order structures?


Four different possible outcomes with the probabilities:

$$
\begin{aligned}
& P\left(Y_{i j}=1 \wedge Y_{j i}=1\right)=p_{i j} \cdot p_{j i} \\
& P\left(Y_{i j}=1 \wedge Y_{j i}=0\right)=p_{i j} \cdot\left(1-p_{j i}\right) \\
& P\left(Y_{i j}=0 \wedge Y_{j i}=1\right)=\left(1-p_{i j}\right) \cdot p_{j i} \\
& P\left(Y_{i j}=0 \wedge Y_{j i}=0\right)=\left(1-p_{i j}\right) \cdot\left(1-p_{j i}\right)
\end{aligned}
$$

Is this supported by the empirical data?

## Assign probabilities to individual ties.

What are the possible outcomes of a network experiment?

Empirical probabilities in the Knecht Classroom Data.
Tie-probability $p=p_{i j}=P\left(Y_{i j}=1\right)=136 /(25 \cdot 24)=0.23$

|  | predicted | empirical |
| :---: | :---: | :---: |
| $P\left(Y_{i j}=1 \wedge Y_{j i}=1\right)$ | 0.05 | 0.13 |
| $P\left(Y_{i j}=1 \wedge Y_{j i}=0\right)$ | 0.18 | 0.10 |
| $P\left(Y_{i j}=0 \wedge Y_{j i}=1\right)$ | 0.18 | 0.10 |
| $P\left(Y_{i j}=0 \wedge Y_{j i}=0\right)$ | 0.59 | 0.67 |

Apparently it is not enough to specify the individual tie probabilities.

## Assign probabilities to dyads and their reverse.

What are the possible outcomes of a network experiment?

We consider each dyad $\left(v_{i}, v_{j}\right)$ together with its reverse $\left(v_{j}, v_{i}\right)$.


Observing such a pair of dyads can can lead to four different outcomes:

$$
\begin{array}{lc}
Y_{i j}=1 \wedge Y_{j i}=1 & \text { both ties are there } \\
Y_{i j}=1 \wedge Y_{j i}=0 & \left(v_{i}, v_{j}\right) \in E \text { but }\left(v_{j}, v_{i}\right) \notin E \\
Y_{i j}=0 \wedge Y_{j i}=1 & \left(v_{i}, v_{j}\right) \notin E \text { but }\left(v_{j}, v_{i}\right) \in E \\
Y_{i j}=0 \wedge Y_{j i}=0 & \text { none of the two ties is there }
\end{array}
$$

The probabilities of these four outcomes have to add up to one.

## Assign probabilities to dyads and their reverse.

What are the possible outcomes of a network experiment?

Problem: the probability of a tie $\left(v_{i}, v_{j}\right)$ can also depend on ties with other actors.


Specifying different probabilities to all possible outcomes on a triplet $\left(v_{i}, v_{j}, v_{k}\right)$ doesn't help either.

The possible outcomes of a network experiment (observation, measurement, ...) are networks.

## Random graph models (definition).

## Definition

A random graph model is a probability space $(\mathcal{G}, P)$, where the set of possible outcomes $\mathcal{G}$ 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

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

## Random graph models (definition).

## Definition

A random graph model is a probability space ( $\mathcal{G}, P$ ), where the set of possible outcomes $\mathcal{G}$ 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

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

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

## Random graph models: remark.

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 vertices is fixed; all the randomness is in the edges.

## Random graph models: edge probability.

Let $(\mathcal{G}, P)$ be a random graph model.
$P: \mathcal{G} \rightarrow[0,1]$ defines a probability for each graph.
A dyad $e \in D$ is associated with a subset of graphs

$$
\mathcal{G}_{e}=\left\{G \in \mathcal{G} ; e \in E_{G}\right\}
$$

When we say "probability of an edge e", we mean $P\left(\mathcal{G}_{e}\right)$; sometimes written as $P(e)$ or $P(e \in E)$.

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

Note: this does not hold the other way round.

## Random graph models: edge probability.

Let $(\mathcal{G}, P)$ be a random graph model.
$P: \mathcal{G} \rightarrow[0,1]$ defines a probability for each graph.
A dyad $e \in D$ is associated with a subset of graphs

$$
\mathcal{G}_{e}=\left\{G \in \mathcal{G} ; e \in E_{G}\right\} .
$$

When we say "probability of an edge $e$ ", we mean $P\left(\mathcal{G}_{e}\right)$; sometimes written as $P(e)$ or $P(e \in E)$.

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

Note: this does not hold the other way round.

## Random graph models: edge probability.

Let $(\mathcal{G}, P)$ be a random graph model.
$P: \mathcal{G} \rightarrow[0,1]$ defines a probability for each graph.
A dyad $e \in D$ is associated with a subset of graphs

$$
\mathcal{G}_{e}=\left\{G \in \mathcal{G} ; e \in E_{G}\right\}
$$

When we say "probability of an edge e", we mean $P\left(\mathcal{G}_{e}\right)$; sometimes written as $P(e)$ or $P(e \in E)$.

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

Note: this does not hold the other way round.

## Example: two random graph models.

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

Define $P_{1}$ by $P_{1}(G)=1 / 8$ for all $G \in \mathcal{G}$.
Define $P_{2}$ by

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

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: $P(e)$ is expected to increase when the nodes incident to $e$ are indirectly connected via a third node.

Independence and non-independence of edges.
small facebook network
769 nodes, 295296 dyads, 16656 edges
$\Rightarrow$ average edge probability is 0.056


186722 dyads are indirectly connected via a third node;
16556 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, 295296 dyads, 16656 edges
$\Rightarrow$ average edge probability is 0.056


186722 dyads are indirectly connected via a third node; 16556 of these are edges $\Rightarrow$ average conditional edge probability for indirectly connected nodes is 0.089

## Independence and non-independence of edges.

## Knecht Classroom Data

$$
p_{u v}=\operatorname{logit}^{-1}\left(\sum_{i=1}^{k} \theta_{i} \cdot s_{i}(u, v ; y)\right)
$$

| statistic | parameter | Std. Error | $\operatorname{Pr}(>\|z\|)$ |
| :--- | ---: | ---: | :--- |
| Intercept) | -4.3664379 | 0.3915032 | $<2 \mathrm{e}-16^{* * *}$ |
| sameGender | 1.2644640 | 0.3036323 | $3.12 \mathrm{e}-05^{* * *}$ |
| similarDelinquency | -0.0009412 | 0.3594857 | 0.998 |
| reciprocity | $\mathbf{2 . 0 6 2 1 8 6 9}$ | 0.2838916 | $3.76 \mathrm{e}-13^{* * *}$ |
| transitivity | $\mathbf{0 . 9 4 2 0 0 7 7}$ | 0.0918453 | $<2 \mathrm{e}-16^{* * *}$ |

Non-zero parameters for reciprocity and transitivity indicate non-independence of edges.

## 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

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

## 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

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

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

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

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

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

## 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

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

Example (probability space: dice)
$P\left(A_{\text {odd }} \mid A_{\leq 4}\right)=1 / 2$, but $P\left(A_{\text {odd }} \mid A_{\leq 3}\right)=2 / 3$

## Independence of dyads in random graph models.

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

$$
\mathcal{G}_{e}=\left\{G \in \mathcal{G} ; e \in E_{G}\right\} .
$$

- 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 $e_{2}$

$$
P_{1}\left(\mathcal{G}_{e_{1}} \cap \mathcal{G}_{e_{2}}\right)=1 / 4=1 / 2 \cdot 1 / 2=P_{1}\left(\mathcal{G}_{e_{1}}\right) \cdot P_{1}\left(\mathcal{G}_{e_{2}}\right)
$$

Let

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

It is for two different dyads $e_{1}$ and $e_{2}$

$$
P_{2}\left(\mathcal{G}_{e_{1}} \cap \mathcal{G}_{e_{2}}\right)=1 / 2 \neq 1 / 2 \cdot 1 / 2=P_{2}\left(\mathcal{G}_{e_{1}}\right) \cdot P_{2}\left(\mathcal{G}_{e_{2}}\right)
$$

## Structural balance theory (illustrating dependence).

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



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

$$
\mathcal{G}_{e}=\left\{G \in \mathcal{G} ; e \in E_{G}\right\} .
$$

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"


## Fully independent random graph models.

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

$$
\mathcal{G}_{e}=\left\{G \in \mathcal{G} ; e \in E_{G}\right\} .
$$

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 $D^{\prime} \subset D$. A dyad $e \in D \backslash D^{\prime}$ is said to be independent of $D^{\prime}$ if for all partitions $D^{\prime}=D^{+} \cup D^{-}$, the subset $\mathcal{G}_{e}$ is independent of

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

If every dyad $e$ is independent of every subset $D^{\prime} \subseteq D \backslash\{e\}$,
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

$$
\mathcal{G}_{e}=\left\{G \in \mathcal{G} ; e \in E_{G}\right\} .
$$

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 $D^{\prime} \subset D$. A dyad $e \in D \backslash D^{\prime}$ is said to be independent of $D^{\prime}$ if for all partitions $D^{\prime}=D^{+} \cup D^{-}$, the subset $\mathcal{G}_{e}$ is independent of

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

If every dyad $e$ is independent of every subset $D^{\prime} \subseteq D \backslash\{e\}$, 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

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

## Expected values in a random graph model.

Random graph models can be characterized by the expected values of certain random variables.

For instance, expected number of edges, triangles, nodes of certain degrees, ...

## Background: random variable and expectation.

Let $(\Omega, P)$ be a finite probability space.
Definition
A random variable is a function $X: \Omega \rightarrow \mathbb{R}$.
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 $(\Omega, P)$ be a finite probability space.
Definition
A random variable is a function $X: \Omega \rightarrow \mathbb{R}$.
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 $X, Y: \Omega \rightarrow \mathbb{R}$ are two random variables and $\alpha$ a real number, then it is

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

## Example: expected number of edges in a random graph model.

Claim
The expected number of edges equals the sum of the edge-probabilities over all dyads.

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


## Example: expected number of edges in a random graph model.

Claim
The expected number of edges equals the sum of the edge-probabilities over all dyads.

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

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

where $\chi_{e}: \mathcal{G} \rightarrow\{0,1\}$ is defined by

$$
\chi_{e}(G)= \begin{cases}1 & \text { if } e \in E_{G} \\ 0 & \text { else }\end{cases}
$$

## Example: expected number of edges in a random graph model.

Claim
The expected number of edges equals the sum of the edge-probabilities over all dyads.

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

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

## Example: expected number of edges in a random graph model.

Claim
The expected number of edges equals the sum of the edge-probabilities over all dyads.

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

$$
\begin{aligned}
\mathbb{E}[m] & =\sum_{e \in D} \mathbb{E}\left[\chi_{e}\right]=\sum_{e \in D} P(e \in E) \cdot 1+P(e \notin E) \cdot 0 \\
& =\sum_{e \in D} P(e \in E)
\end{aligned}
$$

## Uniform graph model: edge probability.

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

Proof.
The two sets


- have the same cardinality $\Rightarrow P\left(\mathcal{G}_{e}\right)=P\left(\overline{\mathcal{G}_{e}}\right)$,
- are disjoint $\Rightarrow P\left(\mathcal{G}_{e}\right)+P\left(\overline{\mathcal{G}_{e}}\right)=P\left(\mathcal{G}_{e} \cup \overline{\mathcal{G}_{e}}\right)$,
- and their union equals $\mathcal{G} \Rightarrow P\left(\mathcal{G}_{e} \cup \overline{\mathcal{G}_{e}}\right)=1$.
$\Rightarrow P\left(\mathcal{G}_{e}\right)=1 / 2$.


## Uniform graph model: edge probability.

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

## Proof.

The two sets

$$
\begin{aligned}
\frac{\mathcal{G}_{e}}{} & =\left\{G \in \mathcal{G} ; e \in E_{G}\right\}, \\
\overline{\mathcal{G}_{e}} & =\left\{G \in \mathcal{G} ; e \notin E_{G}\right\}
\end{aligned}
$$

- have the same cardinality $\Rightarrow P\left(\mathcal{G}_{e}\right)=P\left(\overline{\mathcal{G}_{e}}\right)$,
- are disjoint $\Rightarrow P\left(\mathcal{G}_{e}\right)+P\left(\overline{\mathcal{G}_{e}}\right)=P\left(\mathcal{G}_{e} \cup \overline{\mathcal{G}_{e}}\right)$,
- and their union equals $\mathcal{G} \Rightarrow P\left(\mathcal{G}_{e} \cup \overline{\mathcal{G}_{e}}\right)=1$.
$\Rightarrow P\left(\mathcal{G}_{e}\right)=1 / 2$.


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


Then, with $\mathcal{G}_{e}^{\prime}=\left\{G \in \mathcal{G}^{\prime} ; e \in E_{G}\right\}$ it follows $P\left(\mathcal{G}_{e}^{\prime} \mid \mathcal{G}^{\prime}\right)=1 / 2$ (as on the previous slide).

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

$$
\mathcal{G}^{\prime}=\left\{G \in \mathcal{G} ; D^{+} \subseteq E_{G}, \text { and } D^{-} \cap E_{G}=\emptyset\right\} .
$$

Then, with $\mathcal{G}_{e}^{\prime}=\left\{G \in \mathcal{G}^{\prime} ; e \in E_{G}\right\}$ it follows $P\left(\mathcal{G}_{e}^{\prime} \mid \mathcal{G}^{\prime}\right)=1 / 2$ (as on the previous slide).

## Random graph models: summary.

A random graph model

- assigns probabilities to entire graphs (rather than to individual edges);
- implies edge probabilities (but is not determined by them).

Dependency among dyads (or higher-order structures)

- is what makes network modeling difficult;
- is what makes network modeling interesting;
- is often the essence of social network theories.


## Outline.

## Introduction.

Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## Outline.

Introduction.
Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## Outline.

## Introduction.

Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## 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 $M=n(n-1) / 2$ )


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

## 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 $G\left(n, \frac{1}{2}\right)$.

## 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 $\mathcal{G}\left(n, \frac{1}{2}\right)$.

## 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.

## Introduction.

Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

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

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

Want to do so efficiently $\Rightarrow$ ability to sample for large $n$.

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

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

Simple 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 $\Theta\left(n^{2}\right)$ (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 $\Theta\left(n+p \cdot n^{2}\right)$.

Observation: density p of social networks typically decreases with growing $n$, for instance, $p \in \mathcal{O}(1 / n)$.

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

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

Simple 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 $\Theta\left(n^{2}\right)$ (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 $\Theta\left(n+p \cdot n^{2}\right)$.

Observation: density $p$ of social networks typically decreases with growing $n$, for instance, $p \in \mathcal{O}(1 / n)$.

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

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

Simple 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 $\Theta\left(n^{2}\right)$ (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 $\Theta\left(n+p \cdot n^{2}\right)$.
Observation: density $p$ of social networks typically decreases with growing $n$, for instance, $p \in \mathcal{O}(1 / n)$.

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

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

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

inefficient, when $p$ is small (too many NOs)


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

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

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

Is $d_{1}$ an edge?
(draw a random number...)

inefficient, when p is small (too many NOs)

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

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

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

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

Is $d_{2}$ an edge?
(draw a random number...)

inefficient, when $p$ is small (too many NOs)

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

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

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

Is $d_{2}$ an edge?
$\rightarrow$ YES (for instance)
$\Rightarrow$ turn $d_{2}$ into the first edge

inefficient, when p is small (too many NOs)

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

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

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

Is $d_{3}$ an edge?
(draw a random number...)

inefficient, when $p$ is small (too many NOs)

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

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

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

Is $d_{3}$ an edge?
$\rightarrow$ NO (for instance)

inefficient, when p is small (too many NOs)

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

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

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

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

Is $d_{6}$ an edge?
(draw a random number...)

inefficient, when p is small (too many NOs)

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

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

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

Is $d_{6}$ an edge?
$\rightarrow$ YES (for instance)
$\Rightarrow$ turn $d_{6}$ into the second edge

inefficient, when $p$ is small (too many NOs)

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

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

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

inefficient, when $p$ is small (too many NOs)


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

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

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

inefficient, when $p$ is small (too many NOs)


## 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 ...


## 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 ...


## 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 $\left(d_{3}, d_{4}, d_{5}\right)$; turn $d_{6}$ into the second edge


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

How many dyads shall be left out?
(Notation: $q=1-p$ in the following.)
Observation: the next dyad that becomes an edge is preceded by exactly $k$ non-edges with probability $q^{k} p$.


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

## How many dyads shall be left out?

(Notation: $q=1-p$ in the following.)
Observation: the next dyad that becomes an edge is preceded 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.


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

## How many dyads shall be left out?

(Notation: $q=1-p$ in the following.)
Observation: the next dyad that becomes an edge is preceded 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 $q p$ )


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

## How many dyads shall be left out?

(Notation: $q=1-p$ in the following.)
Observation: the next dyad that becomes an edge is preceded 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 $q p$ ) 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 $\mathcal{G}(n, p)$ (pseudocode).


return $G=(\{0, \ldots, n-1\}, E)$
If $w \geq v$ then $w$ is reduced by $v$ and the row index $v$ is
incremented by one.

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

$$
\begin{aligned}
& E \leftarrow \emptyset \\
& v \leftarrow 1 w \leftarrow-1 \\
& \text { while } v<n \text { do } \\
& \qquad \begin{array}{l}
k \leftarrow \operatorname{rgeom}(p) \\
w \leftarrow w+k+1 \\
\text { while } w \geq v \text { and } v<n \text { do } \\
\quad \begin{array}{l}
w \leftarrow \bar{w}-v \\
v \leftarrow v+1
\end{array} \\
\text { if } v<n \text { then } \\
L E \leftarrow E \cup\{\{v, w\}\}
\end{array}
\end{aligned}
$$

return $G=(\{0, \ldots, n-1\}, E)$


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).

$E \leftarrow \emptyset$
$v \leftarrow 1 w \leftarrow-1$
while $v<n$ do
$k \leftarrow \operatorname{rgeom}(p)$
$w \leftarrow w+k+1$ while $w \geq v$ and $v<n$ do
$\left\lfloor\begin{array}{l}w \leftarrow w-v \\ v \leftarrow v+1\end{array}\right.$
if $v<n$ then
$L E \leftarrow E \cup\{\{v, w\}\}$
return $G=(V, E)$
$\Rightarrow$ runtime in $\Theta(m+n)$.
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.

## 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 $i$ th edge.

Algorithm returns $G$ if and only if

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

$$
\sum_{j=k_{m+1}}^{\infty} p q^{j}=\sum_{j=0}^{\infty} p q^{j}-\sum_{j=0}^{k_{m+1}-1} p q^{j}=1-\left(1-q^{k_{m+1}}\right)=q^{k_{m+1}}
$$

## 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 $i$ th iteration satisfies $k=k_{i}$. Happens with probability $p q^{k_{i}}$.
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

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

## Outline.

## Introduction.

Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

Can such a network be drawn from a $\mathcal{G}(n, p)$ model?
Graph has 769 vertices and about 16600 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 16600 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)
$\left(\mathcal{G}, P_{\theta}\right)$ 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 $G_{o b s}$

$$
L: \Theta \rightarrow \mathbb{R} ; \theta \mapsto P_{\theta}\left(G_{o b s}\right)
$$

A parameter vector $\hat{\theta}$ maximizing $L$, i. e., $\hat{\theta}=\operatorname{argmax} L(\theta)$
is called a maximum likelihood estimate (MLE) for $\theta$.

## 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 $G_{o b s}$

$$
L: \Theta \rightarrow \mathbb{R} ; \theta \mapsto P_{\theta}\left(G_{o b s}\right)
$$

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

$$
\hat{\theta}=\arg \max _{\theta} L(\theta)
$$

is called a maximum likelihood estimate (MLE) for $\theta$.

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

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

$$
\begin{aligned}
L(p) & =P_{p}\left(G_{o \mathrm{obs}}\right)=p^{m}(1-p)^{M-m} . \\
L^{\prime}(p) & =m \cdot p^{m-1} \cdot(1-p)^{M-m}-p^{m} \cdot(M-m) \cdot(1-p)^{M-m-1}
\end{aligned}
$$

Setting $L^{\prime}(p)=0$ for $0<p<1$ yields

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

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

Both graphs have 769 vertices and about 16600 edges.
Maximum likelihood estimate for $p$ is 0.056


Which graph is more likely 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 16600 edges.
Maximum likelihood estimate for $p$ is 0.056


Which graph is more likely 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

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

## 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 $p^{k} q^{n-1-k}$.

$$
q=1-p
$$



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

Histogram of degree2


Histogram of degree1


## Degree distribution in $\mathcal{G}(n, p)$ (limit $n \rightarrow \infty)$.

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 $\mathcal{G}\left(n, p_{n}\right)$ for fixed $v$.

Then it is

$$
\lim _{n \rightarrow \infty} P_{n}[d(v)=k]=e^{-\lambda} \cdot \frac{\lambda^{k}}{k!}
$$

(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


$$
\lim _{n \rightarrow \infty} P_{n}[d(v)=k]=e^{-\lambda} \cdot \frac{\lambda^{k}}{k!}
$$



$$
\lim _{n \rightarrow \infty} P_{n}[d(v)=k]=e^{-\lambda} \cdot \frac{\lambda^{k}}{k!} .
$$

Histogram of degree2


Histogram of degree1


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


Comparing number of triangles.
expected: 13,000
observed: 119,000

## 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.

All three properties are very different for the small facebook network than for the $\mathcal{G}(n, p)$ model.

## Outline.

## Introduction.

Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## Outline.

Introduction.
Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

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.

Introduction.
Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## 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 $p_{i j} \in(0,1)$ 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 $p_{i j}$.
- The model is fully independent.


## Planted partition models.

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

$$
\left[\begin{array}{ccc|ccc|ccc}
p_{1} & \cdots & p_{1} & p_{2} & \cdots & p_{2} & p_{3} & \cdots & p_{3} \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
p_{1} & \cdots & p_{1} & p_{2} & \cdots & p_{2} & p_{3} & \cdots & p_{3} \\
\hline p_{2} & \cdots & p_{2} & p_{4} & \cdots & p_{4} & p_{5} & \cdots & p_{5} \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
p_{2} & \cdots & p_{2} & p_{4} & \cdots & p_{4} & p_{5} & \cdots & p_{5} \\
\hline p_{3} & \cdots & p_{3} & p_{5} & \cdots & p_{5} & p_{6} & \cdots & p_{6} \\
\vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
p_{3} & \cdots & p_{3} & p_{5} & \cdots & p_{5} & p_{6} & \cdots & p_{6}
\end{array}\right]
$$

## Outline.

Introduction.
Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## 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 $\bar{d}=28.78$
B WWW $n=325,729$ and $\bar{d}=5.46$
C Power grid $n=4,941$ and $\bar{d}=2.67$

## Preferential attachment: motivation and history.

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

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

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:

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

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:

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

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:

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

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, ...,b-1 do
    attach an outgoing edge e=(v,\cdot) to v;
    randomly select target w of e with probability
```



```
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

$$
\frac{d_{G}(w)}{\sum_{w^{\prime} \in V} d_{G}\left(w^{\prime}\right)} ;
$$

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 $A[0 \ldots 2 n b-1] \quad / /$ collects endpoints of edges output multi-graph $G=(\{0, \ldots, n-1\}, E)$

$$
E \leftarrow \emptyset ; m \leftarrow 0 \quad \text { //edge set and edge counter }
$$

foreach $v=0, \ldots, n-1$ do

$$
\text { foreach } j=0, \ldots, b-1 \text { do }
$$

$$
A[2 m] \leftarrow v \quad / / v \text { is source of next edge }
$$

$$
w \leftarrow A[\operatorname{random}(\{0, \ldots, 2 m\})] \quad / / \text { randomly select target }
$$

$$
A[2 m+1] \leftarrow w ; \quad \text { //put target in } A
$$

$$
E \leftarrow E \cup\{(v, w)\} ; m \leftarrow m+1 \quad \text { /lupdate edges }
$$

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.

In incrementally defined models we made an arbitrary choice when ordering nodes.

## Outline.

## Introduction.

Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## Outline.

Introduction.
Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## Outline.

## Introduction.

Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## Exponential random graph models (informal).

Exponential random graph models (ERGMs) are a class of random graph models.
Graph probability is a function of 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 estimated from an observed network to test hypotheses.

## Exponential random graph models (ERGM).

## Definition

The ERGM class consists of random graph models $\left(\mathcal{G}, P_{\theta}\right)$ whose probability function $P_{\theta}$ can be written as

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

with

- $s_{i}: \mathcal{G} \rightarrow \mathbb{R}$ for $i=1, \ldots, k$ (statistics);
- $\theta_{i} \in \mathbb{R}$ for $i=1, \ldots, k$ (parameters); $\theta=\left(\theta_{1}, \ldots, \theta_{k}\right)$;
- normalizing constant $\kappa$ defined by

$$
\kappa(\theta)=\sum_{G^{\prime} \in \mathcal{G}} \exp \left(\sum_{i=1}^{k} \theta_{i} \cdot s_{i}\left(G^{\prime}\right)\right)
$$

## ERGM (example).

Consider undirected, loopless graphs with 3 vertices.
$\left.\begin{array}{l}P(G)=\frac{1}{\kappa} \exp [-\log (2) \cdot m(G)+\log (16) \cdot \text { triangles }(G)] \\ \hline m(G) \\ \text { triangles }(G) \\ P(G) \cdot \kappa \\ \hline\end{array}\right)$

## Relation between statistics and probability.

Probability of a graph $G$

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

What happens if one single statistic $s_{i_{0}}$ changes?

$$
P_{\theta}(G)=\exp \left[\theta_{i_{0}} \cdot s_{i_{0}}(G)\right] \cdot \frac{1}{\kappa(\theta)} \exp \left(\sum_{i \neq i_{0}} \theta_{i} \cdot s_{i}(G)\right)
$$

If $s_{i_{0}}\left(G^{\prime}\right)=s_{i_{0}}(G)+c$ and $s_{i}\left(G^{\prime}\right)=s_{i}(G)$ for all $i \neq i_{0}$,
then $P\left(G^{\prime}\right)=\exp \left(\theta_{i_{0}}\right)^{C} \cdot P(G)$.
It is $\operatorname{exn}\left(\theta_{i_{0}}\right)>1 \Leftrightarrow \theta_{i_{0}}>0$ and $\exp \left(\theta_{i_{0}}\right)<1 \Leftrightarrow \theta_{i_{0}}<0$.

## Relation between statistics and probability.

Probability of a graph $G$

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

What happens if one single statistic $s_{i_{0}}$ changes?

$$
P_{\theta}(G)=\exp \left[\theta_{i_{0}} \cdot s_{i_{0}}(G)\right] \cdot \frac{1}{\kappa(\theta)} \exp \left(\sum_{i \neq i_{0}} \theta_{i} \cdot s_{i}(G)\right)
$$

If $s_{i_{0}}\left(G^{\prime}\right)=s_{i_{0}}(G)+c$ and $s_{i}\left(G^{\prime}\right)=s_{i}(G)$ for all $i \neq i_{0}$, then $P\left(G^{\prime}\right)=\exp \left(\theta_{i_{0}}\right)^{c} \cdot P(G)$.

It is $\exp \left(\theta_{i_{0}}\right)>1 \Leftrightarrow \theta_{i_{0}}>0$ and $\exp \left(\theta_{i_{0}}\right)<1 \Leftrightarrow \theta_{i_{0}}<0$.

## Relation between statistics and probability.

Probability of a graph $G$

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

What happens if one single statistic $s_{i_{0}}$ changes?

$$
P_{\theta}(G)=\exp \left[\theta_{i_{0}} \cdot s_{i_{0}}(G)\right] \cdot \frac{1}{\kappa(\theta)} \exp \left(\sum_{i \neq i_{0}} \theta_{i} \cdot s_{i}(G)\right)
$$

If $s_{i_{0}}\left(G^{\prime}\right)=s_{i_{0}}(G)+c$ and $s_{i}\left(G^{\prime}\right)=s_{i}(G)$ for all $i \neq i_{0}$, then $P\left(G^{\prime}\right)=\exp \left(\theta_{i_{0}}\right)^{C} \cdot P(G)$.
It is $\exp \left(\theta_{i_{0}}\right)>1 \Leftrightarrow \theta_{i_{0}}>0$ and $\exp \left(\theta_{i_{0}}\right)<1 \Leftrightarrow \theta_{i_{0}}<0$.

## Relation between statistics and probability.

Let $e \in D$ be a dyad and $G=(V, E) \in \mathcal{G}$ be a graph.
Define $G^{(+e)}=(V, E \cup\{e\})$ and $G^{(-e)}=(V, E \backslash\{e\})$.
Define the change statistic by
$\Delta s(e ; G)=\left[s_{1}\left(G^{(+e)}\right)-s_{1}\left(G^{(-e)}\right), \ldots, s_{k}\left(G^{(+e)}\right)-s_{k}\left(G^{(-e)}\right)\right]$.
Then, it is $\frac{P_{\theta}\left(G^{(+e)}\right)}{P_{\theta}\left(G^{(-\theta)}\right)}=\exp (\theta \cdot \Delta s(e ; G))$.


Note: probability-ratio depends on $e$ and $G$.
Example: $s_{1}(G)=m(G)$ and $s_{2}(G)=\operatorname{triangle}(G)$.

Dyad $\{1,3\}$ has a different change statistic than $\{1,4\}$.

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

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

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

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

Let $P_{2}$ denote the probability function of the $\mathcal{G}(n, p)$ model.

$$
P_{2}(G)=p^{m(G)} \cdot(1-p)^{\binom{n}{2}-m(G)}
$$

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

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

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

where $\theta=\log \left(\frac{p}{1-p}\right)$.
Proof.
Let $P_{2}$ denote the probability function of the $\mathcal{G}(n, p)$ model.

$$
\begin{aligned}
P_{2}(G) & =p^{m(G)} \cdot(1-p)^{\binom{n}{2}-m(G)} \\
& =\left(\frac{p}{1-p}\right)^{m(G)} \cdot(1-p)^{\binom{n}{2}} \\
& =\exp [\theta \cdot m(G)] \cdot(1-p)^{\binom{n}{2}}
\end{aligned}
$$

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

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

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

where $\theta=\log \left(\frac{p}{1-p}\right)$.
Proof.
Let $P_{2}$ denote the probability function of the $\mathcal{G}(n, p)$ model.

$$
\begin{aligned}
P_{2}(G) & =p^{m(G)} \cdot(1-p)^{\binom{n}{2}-m(G)} \\
& =\left(\frac{p}{1-p}\right)^{m(G)} \cdot(1-p)^{\binom{n}{2}} \\
& =\exp [\theta \cdot m(G)] \cdot(1-p)^{\binom{n}{2}}
\end{aligned}
$$

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

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

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

where $\theta=\log \left(\frac{p}{1-p}\right)$.
Proof.
Let $P_{2}$ denote the probability function of the $\mathcal{G}(n, p)$ model.

$$
\begin{aligned}
P_{2}(G) & =p^{m(G)} \cdot(1-p)^{\binom{n}{2}-m(G)} \\
& =\left(\frac{p}{1-p}\right)^{m(G)} \cdot(1-p)^{\binom{n}{2}} \\
& =\exp [\theta \cdot m(G)] \cdot(1-p)^{\binom{n}{2}}
\end{aligned}
$$

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

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

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

where $\theta=\log \left(\frac{p}{1-p}\right)$.
Proof.
Let $P_{2}$ denote the probability function of the $\mathcal{G}(n, p)$ model.

$$
\begin{aligned}
P_{2}(G) & =p^{m(G)} \cdot(1-p){ }^{\binom{n}{2}-m(G)} \\
& =\left(\frac{p}{1-p}\right)^{m(G)} \cdot(1-p)^{\binom{n}{2}} \\
& =\exp [\theta \cdot m(G)] \cdot(1-p)^{\binom{n}{2}}
\end{aligned}
$$

Thus, $P_{2}=c \cdot P_{1}$ for a constant $c ; \Rightarrow c=1$.

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

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

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

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

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


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

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

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

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

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


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

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

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

where $\theta=\log \left(\frac{p}{1-p}\right)$.
Relation between $\theta$ 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 $s$ count the number of specific subgraphs in the network.

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

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.

$$
m(V, E)=|E|
$$

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

## Commonly used network statistics (II).

Assume that actors have attribute values $a: V \rightarrow\{1, \ldots, c\}$, 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 $G$.

$$
t(y)=\sum_{u<v<w} y_{u v} \cdot y_{v w} \cdot y_{w u} \quad ; y \text { adjacency matrix of } G .
$$

A positive (negative) parameter models a preference (reluctance) to close triangles (transitivity).
"A friend of a friend is a friend."


For directed networks: distinguish between transitive triangles and cyclic triangles.

## Commonly used network statistics (IV).

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

$$
s_{\ell}(y)=\sum_{u} \sum_{v_{1}<\cdots<v_{\ell} \neq u} y_{u v_{1}} \cdot \ldots \cdot y_{u v_{\ell}} .
$$

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 $\ell$-star count.
For directed networks: distinguish between out-stars and in-stars.

## Commonly used network statistics (V).

For directed graphs, statistic mutual $(G)$ counts the number of ordered node-pairs $(u, v)$ for which both $(u, v) \in E$ and $(v, u) \in E$.

$$
\operatorname{mutual}(y)=\sum_{u \neq v} y_{u v} \cdot y_{v u}
$$

A positive (negative) parameter models the tendency for (against) reciprocating ties.


## Implication on dyad dependency.

Using some statistics makes dyads dependent.

| statistic | dyads are |  |
| :--- | :--- | ---: |
| number of edges | independent |  |
| edges connecting same attribute |  |  |
| number of triangles |  | dependent |
| number of $\ell$-stars |  |  |

## Dyad dependency (example).

Consider undirected graphs with 3 vertices; 2 -star count $s_{2}$.

| $P(G)=\frac{1}{\kappa} \exp \left[\log (2) \cdot s_{2}(G)\right]$ |  |  |  |  |
| ---: | :---: | :---: | :---: | :---: |
|  | $\bullet$ | $\bullet$ |  |  |
| $s_{2}(G)$ | 0 | 0 | 1 | 3 |
| $P(G) \cdot \kappa$ | 1 | 1 | 2 | $2^{3}=8$ |
| \# isomorphic graphs | 1 | 3 | 3 | 1 |

Let $e, e^{\prime}$ be two different dyads.

$$
\begin{aligned}
P\left(\mathcal{G}_{e} \mid \mathcal{G}_{e^{\prime}}\right) & =(2+8) /(1+2 \cdot 2+8)=10 / 13 \\
P\left(\mathcal{G}_{e}\right) & =(1+2 \cdot 2+8) /(1+3 \cdot 1+3 \cdot 2+8)=13 / 18
\end{aligned}
$$

Thus, dyads $e$ and $e^{\prime}$ are statistically dependent.

## Estimation of ERGM parameters.

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

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

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

$$
L: \mathbb{R}^{k} \rightarrow \mathbb{R} ; \theta \mapsto P_{\theta}\left(G_{\mathrm{obs}}\right)
$$

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

## Outline.

## Introduction.

Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

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

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

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 $D=\left\{d_{1}, \ldots, d_{M}\right\}$ be the set of dyads in an arbitrary but fixed order.

For a given graph $G=(V, E)$ let $E_{i}=E \cap\left\{d_{1}, \ldots, d_{i}\right\}$ and $\bar{E}_{i}=\left\{d_{j} \in\left\{d_{1}, \ldots, d_{i}\right\} ; d_{j} \notin E\right\}$.
For two disjoint subsets $E, \bar{E} \subseteq D$ let

$$
\mathcal{G}_{E, \bar{E}}=\left\{G=\left(V_{G}, E_{G}\right) \in \mathcal{G} ; E \subseteq E_{G} \text { and } \bar{E} \cap E_{G}=\emptyset\right\}
$$

Then, for a given graph $G$ it is


## First try: factorize the probability.

Let $D=\left\{d_{1}, \ldots, d_{M}\right\}$ be the set of dyads in an arbitrary but fixed order.

For a given graph $G=(V, E)$ let $E_{i}=E \cap\left\{d_{1}, \ldots, d_{i}\right\}$ and $\overline{E_{i}}=\left\{d_{j} \in\left\{d_{1}, \ldots, d_{i}\right\} ; d_{j} \notin E\right\}$.
For two disjoint subsets $E, \bar{E} \subseteq D$ let

$$
\mathcal{G}_{E, \bar{E}}=\left\{G=\left(V_{G}, E_{G}\right) \in \mathcal{G} ; E \subseteq E_{G} \text { and } \bar{E} \cap E_{G}=\emptyset\right\}
$$

Then, for a given graph $G$ it is

$$
P(G)=\prod_{d_{i} \in E} P\left(\mathcal{G}_{d_{i}} \mid \mathcal{G}_{E_{i-1}, \overline{E_{i-1}}}\right) \cdot \prod_{d_{i} \in D \backslash E} 1-P\left(\mathcal{G}_{d_{i}} \mid \mathcal{G}_{E_{i-1}, \overline{E_{i-1}}}\right)
$$

## Sampling from an ERGM: first try.

For a given graph $G$ it is

$$
P(G)=\prod_{d_{i} \in E} P\left(\mathcal{G}_{d_{i}} \mid \mathcal{G}_{E_{i-1}, \overline{E_{i-1}}}\right) \cdot \prod_{d_{i} \in D \backslash E} 1-P\left(\mathcal{G}_{d_{i}} \mid \mathcal{G}_{E_{i-1}, \overline{E_{i-1}}}\right)
$$

Sample from a given ERGM:
$E \leftarrow \emptyset ; \bar{E} \leftarrow \emptyset$
for $i=1, \ldots, M$ do
with probability $P\left(\mathcal{G}_{d_{i}} \mid \mathcal{G}_{E, \bar{E}}\right)$
put $d_{i}$ into $E$;
otherwise
put $d_{i}$ into $\bar{E}$.
Problem: probabilities are computationally intractable.

## Sampling from an ERGM: first try.

For a given graph $G$ it is

$$
P(G)=\prod_{d_{i} \in E} P\left(\mathcal{G}_{d_{i}} \mid \mathcal{G}_{E_{i-1}, \overline{E_{i-1}}}\right) \cdot \prod_{d_{i} \in D \backslash E} 1-P\left(\mathcal{G}_{d_{i}} \mid \mathcal{G}_{E_{i-1}, \overline{E_{i-1}}}\right)
$$

Sample from a given ERGM:
$E \leftarrow \emptyset ; \bar{E} \leftarrow \emptyset$
for $i=1, \ldots, M$ do
with probability $P\left(\mathcal{G}_{d_{i}} \mid \mathcal{G}_{E, \bar{E}}\right)$
put $d_{i}$ into $E$;
otherwise put $d_{i}$ into $\bar{E}$.

Problem: probabilities are computationally intractable.

## More details: conditional probabilities are computationally intractable in general.

Probability of a graph $G$ in an ERGM:

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

Normalizing constant $\kappa$ cancels out when computing conditional probabilities

$$
P\left(\mathcal{G}_{d_{i}} \mid \mathcal{G}_{E_{i-1}, \overline{E_{i-1}}}\right) .
$$

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

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

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

Sample from a given ERGM:


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

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

(this is 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 $P\left(\mathcal{G}_{d_{i}} \mid \mathcal{G}\left[(V, E),-d_{i}\right]\right)$ put $d_{i}$ into $E$;
otherwise
remove $d_{i}$ from $E$ (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

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

(this is 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 $P\left(\mathcal{G}_{d_{i}} \mid \mathcal{G}\left[(V, E),-d_{i}\right]\right)$ put $d_{i}$ into $E$;
otherwise
remove $d_{i}$ from $E$ (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 $P\left(\mathcal{G}_{d} \mid \mathcal{G}[(V, E),-d]\right)$
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
$T \rightarrow \infty$.
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 $P\left(\mathcal{G}_{d} \mid \mathcal{G}[(V, E),-d]\right)$ 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 $T \rightarrow \infty$.

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);


## 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);
 $\Rightarrow$ 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 $\pi$ are a function of $P$
- in such a way that
- the probability to be on graph $G$ converges to $P(G)$, when the number of simulation steps tends to $\infty$. $\Rightarrow$ 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 $\pi$ are a function of $P$
- in such a way that
- the probability to be on graph $G$ converges to $P(G)$, when the number of simulation steps tends to $\infty$.
$\Rightarrow$ 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 $\pi$ are a function of $P$
- in such a way that
- the probability to be on graph $G$ converges to $P(G)$, when the number of simulation steps tends to $\infty$.
$\Rightarrow$ 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.

$\pi_{i j}$ 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 $\mathcal{G}=\left\{G_{1}, \ldots, G_{N}\right\}$ (state space);
- $\pi$ is a matrix $\pi \in \mathbb{R}^{N \times N}$ (transition matrix) satisfying
- for all $i, j$ it is $\pi_{i j} \in[0,1]$;
- for all $i$ it is $\sum_{j=1}^{N} \pi_{i j}=1$.
$\pi_{i j}$ 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 $\pi$ 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 $\infty$.


## An important observation.

$\pi_{i j}$ interpreted as the probability to jump from state $G_{i}$ to $G_{j}$

$$
\begin{gathered}
P\left(G_{1}\right) \xrightarrow{\pi_{12}} P\left(G_{2}\right) \stackrel{\pi_{32}}{\Perp} P\left(G_{3}\right) \\
P\left(G_{4}\right) \xrightarrow{\pi_{42}} \underset{P\left(G_{5}\right)}{ } \quad \ldots
\end{gathered}
$$

If the probability to be on a graph $G$ after $t$ iteration steps is denoted by $P^{(t)}(G)$, then (with $\left.P^{(t)}=\left[P^{(t)}\left(G_{1}\right), \ldots, P^{(t)}\left(G_{N}\right)\right] \in \mathbb{R}^{N}\right)$ it is

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

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

$$
P^{(t+1)}\left(G_{j}\right)=\sum_{i=1}^{N} P^{(t)}\left(G_{i}\right) \pi_{i j}
$$

## Stationary state space distributions.

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

That is, for the desired limit $P$ it must hold

$$
P=P \pi .
$$

Such a $P$ is called a stationary distribution for the Markov chain.

## Irreducible and aperiodic Markov chains.

$\pi_{i j}$ 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 $\pi_{i_{j} j_{j+1}}>0$.

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

- irreducible if for any two states $G_{j}, G_{j} \in \mathcal{G}$ 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.

$\pi_{i j}$ 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 $\pi_{i_{j} j_{j+1}}>0$.

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.


## Theorem

Let $P$ be a probability distribution on $\mathcal{G}$ and $\pi$ be the transition matrix of a Markov chain on $\mathcal{G}$.
If for all graphs $G_{i}, G_{j}$ it is

$$
P\left(G_{i}\right) \pi_{i j}=P\left(G_{j}\right) \pi_{j i}
$$

(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 $P_{0}$ it is

$$
\lim _{K \rightarrow \infty} P_{0} \pi^{K}=P
$$

## Gibbs sampling (transition matrix).

Given $P$, define $\pi$ such that

$$
P\left(G_{i}\right) \pi_{i j}=P\left(G_{j}\right) \pi_{j i}
$$

Gibbs sampling: define $\pi$ as follows

- $\pi_{i j}=0$ if $G_{i}$ and $G_{j}$ differ in more than one dyad;
- if $G_{i}$ and $G_{j}$ differ in exactly one dyad, then

$$
\pi_{i j}=\frac{P\left(G_{j}\right)}{\binom{n}{2}\left(P\left(G_{i}\right)+P\left(G_{j}\right)\right)} .
$$

- $\pi_{i i}=\sum \frac{P\left(G_{i}\right)}{\binom{n}{2}\left(P\left(G_{i}\right)+P(G)\right)}$
(sum over all $G$ that differ from $G_{i}$ in exactly one dyad)
Show: $\pi$ is normalized, irreducible, aperiodic, reversible.


## Gibbs sampling (algorithm).

initialize $\mathcal{G}$ by any graph from $\mathcal{G}$;
repeat many times

- select a dyad e uniformly at random;
- with probability $\frac{P(G(+e))}{P\left(G^{(+e)}\right)+P(G(-e))}$
- replace $G=(V, E)$ by $G^{(+e)}=(V, E \cup\{e\})$
- otherwise replace $G=(V, E)$ by $G^{(-e)}=(V, E \backslash\{e\})$;
return $G$;
Note: The probability at each step is just dependent on the change statistic $\Delta s(e ; G)$.


## Sampling from an ERGM in R.

The function simulate in the ergm package can sample from an ERGM with given statistics and parameters.

Function gof (goodness-of-fit) compares several statistics of the observed network with the distribution in a given ERGM (estimated from sampled networks).

## Outline.

Introduction.
Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## Recall: maximum likelihood parameters.

Definition (maximum likelihood)
$\left(\mathcal{G}, P_{\theta}\right)$ random graph model parameterized by $\theta \in \Theta \subseteq \mathbb{R}^{k}$;
$G^{*} \in \mathcal{G}$ a graph (observation).
Likelihood function associated with $G^{*}$

$$
L: \Theta \rightarrow \mathbb{R} ; \theta \mapsto P_{\theta}\left(G^{*}\right)
$$

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

$$
\hat{\theta}=\arg \max _{\theta} L(\theta)
$$

is called a maximum likelihood estimate (MLE) for $\theta$.

## Testing hypotheses with ERGMs.

To test a hypothesis (e. g., the friend of a friend is a friend) with observed network data $G^{*}$.

- Decide on a reasonable set of statistics $s_{i}, i=1, \ldots, k$

$$
P(G)=\frac{1}{\kappa} \exp \left(\sum_{i=1}^{k} \theta_{i} \cdot s_{i}(G)\right)
$$

including a statistic related to the hypothesis,
e. g., $s_{k}=$ number of triangles.

- Compute maximum likelihood estimates $\hat{\theta}=\left(\hat{\theta}_{1}, \ldots, \hat{\theta}_{k}\right)$.
- Check whether $\hat{\theta}_{k}$ is significantly positive: compute probability of observing a network in the null model defined by $\left(\hat{\theta}_{1}, \ldots, \hat{\theta}_{k-1}, 0\right)$ that gives rise to $\theta_{k}$ as large as $\hat{\theta}_{k}$.

Maximizing the likelihood function

$$
L(\theta)=\frac{1}{\kappa(\theta)} \exp \left(\sum_{i=1}^{k} \theta_{i} \cdot s_{i}\left(G^{*}\right)\right)
$$

is computationally intractable (since the normalizing constant $\kappa$ has too many terms).

Even though it is analytically simple (formulas for the partial derivatives of any order can be given).

## Partial derivatives of the likelihood function.

Likelihood function

$$
\begin{aligned}
& L(\theta)=\frac{1}{\kappa(\theta)} \exp \left(\sum_{i=1}^{k} \theta_{i} \cdot s_{i}\left(G^{*}\right)\right) \text { with } \\
& \kappa(\theta)=\sum_{G^{\prime} \in \mathcal{G}} \exp \left(\sum_{i=1}^{k} \theta_{i} \cdot s_{i}\left(G^{\prime}\right)\right) .
\end{aligned}
$$

Log-likelihood function

$$
\ell(\theta)=\log (L(\theta))=\left(\sum_{i=1}^{k} \theta_{i} \cdot s_{i}\left(G^{*}\right)\right)-\log (\kappa(\theta)) .
$$

Partial derivative by $\theta_{j}$ for $j=1, \ldots, k$

$$
\begin{aligned}
\frac{\partial}{\partial \theta_{j}} \ell(\theta) & =s_{j}\left(G^{*}\right)-\frac{1}{\kappa(\theta)} \cdot \sum_{G^{\prime} \in \mathcal{G}} \exp \left(\sum_{i=1}^{k} \theta_{i} \cdot s_{i}\left(G^{\prime}\right)\right) \cdot s_{j}\left(G^{\prime}\right) \\
& =s_{j}\left(G^{*}\right)-\mathbb{E}_{\theta}\left(s_{j}\right) .
\end{aligned}
$$

## Maximizing the likelihood function.

Likelihood function

$$
L(\theta)=\frac{1}{\kappa(\theta)} \exp \left(\sum_{i=1}^{k} \theta_{i} \cdot s_{i}\left(G^{*}\right)\right)
$$

is maximized by computing $\hat{\theta}$ such that

$$
\frac{\partial}{\partial \theta_{j}} \ell(\hat{\theta})=s_{j}\left(G^{*}\right)-\mathbb{E}_{\hat{\theta}}\left(s_{j}\right)=0, \text { for } j=1, \ldots, k .
$$

That is: compute those parameters $\hat{\theta}$ that make the expected values of all statistics equal to the statistics of the observed network.

Expected values are computationally intractable as well, but can be estimated from sampled graphs:

- sample $N$ graphs $G_{1}, \ldots, G_{N}$ from ( $\mathcal{G}, P_{\theta}$ );
- take $\bar{s}=\frac{1}{N} \cdot \sum_{j=1}^{N} s\left(G_{j}\right)$ as an estimate for $\mathbb{E}_{\theta}(s)$.


## Newton-Raphson method.

Let $\left(\mathcal{G}, P_{\theta}\right)$ be an ERGM parameterized by $\theta$ and $G^{*}$ an observed network. To compute the maximum likelihood estimates $\hat{\theta}$

1. Choose initial parameter values $\theta^{(0)}=\left(\theta_{1}^{(0)}, \ldots, \theta_{k}^{(0)}\right)$;
2. For $i=0,1, \ldots$ until convergence
2.1 Sample $G_{1}, \ldots, G_{N} \sim\left(\mathcal{G}, P_{\theta^{(1)}}\right)$;
$2.2 \bar{s} \leftarrow \frac{1}{N} \cdot \sum_{j=1}^{N} s\left(G_{j}\right)$;
2.3 $C \leftarrow \frac{1}{N} \cdot \sum_{j=1}^{N} s\left(G_{j}\right) \cdot s\left(G_{j}\right)^{T}$;
2.4 $H \leftarrow \bar{s} \cdot \bar{s}^{T}-C$;
$2.5 \theta^{(i+1)} \leftarrow \theta^{(i)}-H^{-1} \cdot\left(s\left(G^{*}\right)-\bar{s}\right)$
The inverse of the matrix $H$ from the last iteration is an estimate for the covariance matrix. $\Rightarrow$ yields standard errors

Function ergm in the ergm package estimates parameters (using a different, more sophisticated method).

## Outline.

Introduction.
Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## Near-degeneracy and multi-modality of ERGMs.

Many ERGMs give rise to multi-modal probability distributions:

- probability mass centered on a small set of graphs
- other graphs are very unlikely.

For instance, only near-empty or near-complete graphs have a non-vanishing probability.

## Near-degeneracy and multi-modality of ERGMs

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

$$
P_{\theta}(G)=\frac{1}{\kappa(\theta)} \exp \left(\eta m(G)+\sigma_{2} s_{2}(G)\right) \quad\left(\eta, \sigma_{2}\right)=(-2,-0.2)
$$



## Near-degeneracy and multi-modality of ERGMs

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

$$
P_{\theta}(G)=\frac{1}{\kappa(\theta)} \exp \left(\eta m(G)+\sigma_{2} s_{2}(G)\right) \quad\left(\eta, \sigma_{2}\right)=(1,0.2)
$$



## Near-degeneracy and multi-modality of ERGMs

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

## Examples

$$
P_{\theta}(G)=\frac{1}{\kappa(\theta)} \exp \left(\eta m(G)+\sigma_{2} s_{2}(G)\right) \quad\left(\eta, \sigma_{2}\right)=(-2,0.4)
$$



## Near-degeneracy and multi-modality of ERGMs.

Consider the following ERGM

$$
P(G)=\frac{1}{\kappa} \exp (\eta \cdot m(G)+\tau \cdot t(G)) \text { with } \eta<0, \tau>0 .
$$

Then, in very sparse networks

- there are few possibilities to close triangles;
- creation of edges is unlikely;
$\Rightarrow$ very unlikely to leave the set of near-empty graphs.
In contrast, in very dense networks
- an edge can close many triangles (up to $n-2$ );
- deletion of edges is unlikely;
$\Rightarrow$ very unlikely to leave the set of near-complete graphs.


## Near-degeneracy and multi-modality of ERGMs.

Degeneracy is undesirable for two reasons.

1. Convergence of the Markov chain towards the stationary distribution is very slow.
2. Degenerate models seem to be unreasonable models for empirical networks.

## Avoiding near-degeneracy of ERGMs.

Triangle statistic implies linear marginal effect of closed triangles:

- closing one triangle contributes $\tau$ to the log-probability;
- closing two triangles contributes $2 \tau \ldots$


Geometrically-weighted edgewise shared partner (gwesp) statistic:

- a $k$-triangle counts more than a single triangle,
- but less than $k$-times as much.

Typically leads to less degenerate models.

## Outline.

## Introduction.

Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

## Recall: dyad dependency.

Using some statistics makes dyads dependent.

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



## 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 $P\left(d_{1} \in E\right)$.


## 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 $D^{+} \uplus D^{-}=D \backslash\left\{d_{1}, d_{2}\right\}$ of the set of dyads different from $d_{1}$ and $d_{2}$ let the subset $\mathcal{G}_{D^{+}, D^{-}}$be defined by

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

We say that $d_{1}$ and $d_{2}$ are conditionally independent (given the rest of the graph) if for all partitions $D^{+} \uplus D^{-}=D \backslash\left\{d_{1}, d_{2}\right\}$ it is

$$
P\left(\mathcal{G}_{d_{1}} \mid \mathcal{G}_{D^{+}, D^{-}}\right)=P\left(\mathcal{G}_{d_{1}} \mid \mathcal{G}_{D^{+}, D^{-}} \cap \mathcal{G}_{d_{2}}\right) .
$$

Informally: if we know the state of all dyads in $D \backslash\left\{d_{1}, d_{2}\right\}$, the state of the dyad $d_{2}$ does not provide any additional information about the probability $P\left(d_{1} \in E\right)$.

## Hammersley-Clifford Theorem; special case.

Theorem (first part)
Let $D$ be a set of dyads and $\mathcal{G}$ be the set of all graphs defined on $D$. Let $(\mathcal{G}, P)$ be a random graph satisfying $P(G)>0$ for all $G \in \mathcal{G}$.

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


## Hammersley-Clifford Theorem; special case.

## Theorem (first part)

Let $D$ be a set of dyads and $\mathcal{G}$ be the set of all graphs defined on $D$. Let $(\mathcal{G}, P)$ be a random graph satisfying $P(G)>0$ for all $G \in \mathcal{G}$.

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

$$
\begin{align*}
P(G) & =\frac{1}{\kappa} \exp \left(\sum_{A \subseteq E(G)} \alpha_{A}\right), \quad \text { where }  \tag{1}\\
\kappa & =\sum_{G^{\prime} \in \mathcal{G}} \exp \left(\sum_{A \subseteq E\left(G^{\prime}\right)} \alpha_{A}\right) . \tag{2}
\end{align*}
$$

## Hammersley-Clifford Theorem; special case.

Theorem (second part)
Conversely, if the probability $P$ on $\mathcal{G}$ is defined by

$$
\begin{aligned}
P(G) & =\frac{1}{\kappa} \exp \left(\sum_{A \subseteq E(G)} \alpha_{A}\right), \quad \text { where } \\
\kappa & =\sum_{G^{\prime} \in \mathcal{G}} \exp \left(\sum_{A \subseteq E\left(G^{\prime}\right)} \alpha_{A}\right),
\end{aligned}
$$

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

$$
\begin{aligned}
P(G) & =\frac{1}{\kappa} \exp \left(\sum_{A \subseteq E(G)} \alpha_{A}\right), \quad \text { where } \\
\kappa & =\sum_{G^{\prime} \in \mathcal{G}} \exp \left(\sum_{A \subseteq E\left(G^{\prime}\right)} \alpha_{A}\right),
\end{aligned}
$$

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 $\left\{\alpha_{A} \in \mathbb{R} ; A \subseteq D\right\}$, 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 $s_{A}(G)= \begin{cases}1 & A \subseteq E(G) \\ 0 & \text { else }\end{cases}$
- parameters: $\alpha_{A}$



## Conclusion from the Hammersley-Clifford Theorem.

There are constants $\left\{\alpha_{A} \in \mathbb{R} ; A \subseteq D\right\}$, 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 $s_{A}(G)= \begin{cases}1 & A \subseteq E(G) \\ 0 & \text { else }\end{cases}$
- parameters: $\alpha_{A}$

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

## 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 $D$, $\left\{d_{i}, d_{j}\right\} \in E$ 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


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

$$
\begin{array}{rll}
\eta_{u v} & \text { for all } & \text { dyads }\{u, v\} \\
\tau_{u v w} & \text { for all } & \text { triangles }\{u, v, w\} \\
\sigma_{u v_{1} \ldots v_{\ell}} & \text { for all } & 2 \leq \ell \leq n-1, \text { and all } \\
& & \ell \text {-stars }\left(u,\left\{v_{1}, \ldots, v_{\ell}\right\}\right)
\end{array}
$$

such that the probability of a graph $\mathcal{G} \in \mathcal{G}$ can be written as

$$
P(G)=\frac{1}{\kappa} \exp \left(\sum_{u v \in E(G)} \eta_{u v}+\sum_{u v w \in T(G)} \tau_{u v w}+\sum_{\ell=2}^{n-1} \sum_{u v_{1} \ldots v_{\ell} \in S_{\ell}(G)} \sigma_{u v_{1} \ldots v_{\ell}}\right)
$$

## ERGM of general Markov graphs (remarks).

$$
P(G)=\frac{1}{\kappa} \exp \left(\sum_{u v \in E(G)} \eta_{u v}+\sum_{u w \in T(G)} \tau_{u u w}+\sum_{l=2}^{n-1} \sum_{u v_{1} \ldots v_{\ell} \in S_{\ell}(G)} \sigma_{u v_{1} \ldots v_{l}}\right)
$$

Each dyad, triangle, $\ell$-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 $\varphi: V \rightarrow W$ such that

$$
\forall u, v \in V:\{u, v\} \in E \Leftrightarrow\{\varphi(u), \varphi(v)\} \in F
$$

## 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 $\eta, \tau$, and $\sigma_{\ell}$ 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 $\ell$-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 $\eta, \tau$, and $\sigma_{\ell}$ for $\ell=2, \ldots, n-1$ such that the probability of a graph $\mathcal{G} \in \mathcal{G}$ 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 $\ell$-star parameters are equal...
Show that any two triangle parameters are equal. . .

## ERGMs of fully independent models.

The dependence graph of a fully independent model has no edges. Thus, the only cliques are single dyads and

$$
P(G)=\frac{1}{\kappa} \exp \left(\sum_{u v \in E(G)} \eta_{u v}\right)
$$

for constants $\eta_{u v}$ associated with dyads $u v \in D$.
$\mathcal{G}(n, p)$ is a homogeneous fully independent model. Thus,

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

## Outline.

## Introduction.

Running example: data, questions, and simple answers.
Random graph models.
$\mathcal{G}(n, p)$.
Definition of $\mathcal{G}(n, p)$.
Sampling from $\mathcal{G}(n, p)$.
Plausibility of $\mathcal{G}(n, p)$ as a model for social networks.
Towards more structured models.
Planted partition models.
Preferential attachment.
Exponential random graph models.
Definition and examples.
Sampling from an ERGM.
Hypothesis testing and parameter estimation.
Near-degeneracy and multi-modality of ERGMs.
Hammersley-Clifford Theorem.
Miscellaneous.

Estimating parameters from several observed networks.

## ERGM of a family of networks.

Suppose you have observed $N>1$ networks $G_{1}, \ldots, G_{N}$ (for instance, friendship networks in $N$ school classes).

If you can assume that

1. the networks are independent draws
2. from an identical ERGM (meaning: same graphspace, same statistics, and same parameters)
then everything is fine!
To estimate maximum likelihood parameters $\hat{\theta}$

- the expected statistics $\mathbb{E}_{\hat{\theta}}(s)$ must be equal to the average observed statistics of the $N$ networks
$\Rightarrow$ use the target. stats argument of ergm.
- Divide the standard errors by $\sqrt{N}$.

Similar: time-homogeneous TERGM (note: $N=T-1$ ).

## What if assumptions are unrealistic?

Suppose you have observed $N>1$ networks $G_{1}, \ldots, G_{N}$.

1. The networks are independent draws
2. from an identical ERGM.

If the above assumptions are cannot be made, different approaches might be chosen.

1. If dependencies across networks cannot be excluded, put the $N$ adjacency matrices in the diagonal blocks of a joint adjacency matrix; use appropriate cross-network statistics.
2. Methods to model variation of parameters across networks exist (beyond the scope of this lecture).

## Maximum pseudolikelihood estimation.

(Faster computation of the wrong parameter estimates.)

## Pseudolikelihood.

Let $P_{\theta}$ be the probability of an ERGM on $\mathcal{G}$, let $d \in D$ be a dyad and $G=(V, E) \in \mathcal{G}$ be a graph.
Define $G^{(+d)}=(V, E \cup\{d\})$ and $G^{(-d)}=(V, E \backslash\{d\})$ and

$$
\mathcal{G}[G,-d]=\left\{G^{(+d)}, G^{(-d)}\right\},
$$

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

The pseudolikelihood associated with an observation $G^{*}=(V, E)$ is
$L_{G^{*}}^{\text {(pseudo })}(\theta)=\prod_{d \in E} P_{\theta}\left(\mathcal{G}_{d} \mid \mathcal{G}\left[\mathcal{G}^{*},-d\right]\right) \cdot \prod_{d \in D \backslash E} 1-P_{\theta}\left(\mathcal{G}_{d} \mid \mathcal{G}\left[\mathcal{G}^{*},-d\right]\right)$.
Note that the associated "probability" is not a proper probability on $\mathcal{G}$.

## Interpreting pseudolikelihood.

The pseudolikelihood "probability" of a graph $G=(V, E)$
$P_{\theta}^{(\text {pseudo })}(G)=\prod_{d \in E} P_{\theta}\left(\mathcal{G}_{d} \mid \mathcal{G}[G,-d]\right) . \prod_{d \in D \backslash E} 1-P_{\theta}\left(\mathcal{G}_{d} \mid \mathcal{G}[G,-d]\right)$,

- pretends that the graph $G$ is the result of independent decisions for all dyads $d \in D$.
- When deciding about whether $d \in E$ we fix all other dyads to their value (edge or non-edge) in $G$.
- Dyad $d$ is turned into an edge with the conditional probability $P_{\theta}\left(\mathcal{G}_{d} \mid \mathcal{G}[G,-d]\right)$.
Note that this process to draw a graph is not well-defined, since we need to know $\mathcal{G}[G,-d])$ before we know it.


## Summary of pseudolikelihood estimation.

Pseudolikelihood estimation of ERGMs is logistic regression where

- the binary outcome variables are the dyads $d \in D$ for which either $d \in E$ (variable equal to 1 ) or $d \notin E$ (variable equal to 0 );
- the explanatory variables for the dyads $d \in D$ are the $k$ change statistics

$$
\Delta s(d ; G)=\left[s_{1}\left(G^{(+d)}\right)-s_{1}\left(G^{(-d)}\right), \ldots, s_{k}\left(G^{(+d)}\right)-s_{k}\left(G^{(-d)}\right)\right]
$$

associated with the observed graph $G$.

## Remarks about pseudolikelihood.

Pseudolikelihood estimation has the following properties.

- It is fast and does not lead to degenerate models.
- Results suggest that parameters from pseudolikelyhood converge to the MLE parameters when the network size increases.
- Results suggest that standard errors are likely to be too small.
That is, the null-hypothesis is rejected too often.
That is, you might think you've found a significant effect when in reality there is none.
- You cannot use such a pseudolikelihood model to sample a graph from scratch because the explanatory variables can only be computed once you have a graph.


## Bootstrap sampling of confidence intervals.

(Correcting the damage done by pseudolikelihood estimation.)

## Bootstrap sampling of confidence intervals: idea.

Situation: have $N$ graphs assumed to be drawn from one ERGM and want to estimate parameters and their standard errors.

Compute parameter estimates by logistic regression (MPLE) but don't use the standard errors from logistic regression.

Bootstrap sampling: repeatedly sample $N$ graphs from the observed graphs $G_{1}, \ldots, G_{N}$ (with replacement) and use the resulting distribution of the parameters.

## Bootstrap sampling of confidence intervals.

Given $N$ observed graphs $G_{1}, \ldots, G_{N}$ from an ERGM with unknown parameters.

1. For $j=1, \ldots, q$
1.1 sample uniformly, with replacement, $N$ indices $i_{1}^{(j)}, \ldots, i_{N}^{(j)}$ from $1, \ldots, N$;
1.2 compute $\theta^{(j)}=\operatorname{MPLE}\left(G_{i_{1}^{(j)}}, \ldots, G_{i_{N}^{(j)}}\right)$;
2. compute confidence intervals from $\theta^{(1)}, \ldots, \theta^{(q)}$.

Reject the null hypothesis if, for instance, the 95\% confidence interval of a parameter is positive.

Is implemented in the xergm package.

Constrained ERGMs.

## Constrained ERGMs.

ERGMs can be defined on constrained graph spaces, e. g.,

- only graphs with exactly $m$ edges;
- only graphs with bounded maximum or minimum degree;
- only graphs with given degree distribution; ...

Constraints might result from specific data collection.
(Almost) everything stays the same as for unconstrained ERGMs (use the constraints argument of ergm).

- Markov chain simulation has to be adapted to never leave the graph space.
- Hammersley-Clifford Theorem is no longer valid.

