Network Modeling

Viviana Amati Jürgen Lerner

Dept. Computer & Information Science University of Konstanz

Winter 2015/2016 (last updated: January 11, 2016)

▲□▶ ▲□▶ ▲ 三▶ ▲ 三▶ - 三 - のへぐ

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.

Statistical models for social networks.

◆□▶ ◆□▶ ◆ □▶ ◆ □▶ ● □ ● ● ● ●

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, ...



(日) (日) (日) (日) (日) (日) (日)

Actors and ties may have associated data (attributes).

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.

(ロ) (同) (三) (三) (三) (○) (○)

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.

(ロ) (同) (三) (三) (三) (○) (○)

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);

(日) (日) (日) (日) (日) (日) (日)

- behavior of A's friends (social influence);
- A's friendship ties; …

Running hypotheses.

- H_1 Pupils chose friends with the same gender.
- H₂ Pupils reciprocate friendship.
- H_3 The friend of a friend is a friend.
- H_4 Pupils chose friends with similar delinquency behavior.

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

 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

 $D_1 = \{(A, B); \text{ gender}(A) = \text{gender}(B)\}$

 $D_2 = \{(A, B); \text{ gender}(A) \neq \text{gender}(B)\}$

Compare ratio of friendship ties in the two groups.

$$\frac{\# \text{ ties in } D_1}{\# \text{ dyads in } D_1} \quad \text{vs.} \quad \frac{\# \text{ ties in } D_2}{\# \text{ dyads in } D_2}$$
sult:
$$\frac{105}{312} = 0.3365 \quad \text{vs.} \quad \frac{31}{288} = 0.1076$$

<ロ> <0</p>

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

 $D_1 = \{(A, B); gender(A) = gender(B)\}$

$$D_2 = \{(A, B); \text{ gender}(A) \neq \text{gender}(B)\}$$

Compare ratio of friendship ties in the two groups.

$$\frac{\# \text{ ties in } D_1}{\# \text{ dyads in } D_1} \quad \text{vs.} \quad \frac{\# \text{ ties in } D_2}{\# \text{ dyads in } D_2}$$

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

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

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

 $D_1 = \{(A, B); gender(A) = gender(B)\}$

 $D_2 = \{(A, B); gender(A) \neq gender(B)\}$

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 gender0.34 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.

(ロ) (同) (三) (三) (三) (三) (○) (○)

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 gender0.34 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.

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

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.

(ロ) (同) (三) (三) (三) (三) (○) (○)

Modeling the occurence of ties by logistic regression.

Random variable Y_{uv} for the from u to v

$$Y_{uv} = \begin{cases} 1 & \text{with probability } p_{uv} ; \\ 0 & \text{with probability } 1 - p_{uv} \end{cases}$$

p_uv = someFunctionOf(statistics, parameters)

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 occurrence of ties by logistic regression.

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

$$p_{uv} = \log it^{-1}(\theta \cdot s) = rac{\exp(\theta \cdot s)}{\exp(\theta \cdot s) + 1}$$
, where
 $s = (s_1, \dots, s_k) \in \mathbb{R}^k$ statistics
 $\theta = (\theta_1, \dots, \theta_k) \in \mathbb{R}^k$ parameters
 $\theta \cdot s = \sum_{i=1}^k \theta_i \cdot s_i$

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*:

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

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

Gender model: friendship ties explained by gender-equality

 $p_{uv} = \text{logit}^{-1}(\theta_1 + \theta_2 \cdot \text{sameGender}(u, v))$.

Results:

statistic	parameter	Std. Error	Pr(> z)
(Intercept)	-2.1151	0.1901	< 2e-16 ***
sameGender	1.4363	0.2247	1.64e-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

Delinquency model: friendship ties explained by similar behavior

 $p_{uv} = \text{logit}^{-1}(\theta_1 + \theta_2 \cdot \text{similarDelinquency}(u, v))$.

Results:

statistic	parameter	Std. Error	Pr(> z)
(Intercept)	-1.5880	0.1796	<2e-16 ***
similarDelinquency	0.6568	0.2619	0.0121 *

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

More complex model: control for alternative explanations:

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

$s_i(u, v; y)$	interpretation
1	constant (intercept)
sameGender(<i>u</i> , <i>v</i>)	gender homophily
similarDelinquency (u, v)	behavior homophily
Уvu	reciprocity
$\sum_{w} y_{uw} \cdot y_{wv}$	transitivity (friend of friend is friend)

▲□▶ ▲□▶ ▲ 三▶ ▲ 三▶ - 三 - のへぐ

More complex model:

$$p_{uv} = \text{logit}^{-1}\left(\sum_{i=1}^{k} \theta_i \cdot s_i(u, v; y)\right)$$

Results:

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

The analysis so far is invalid.

Logistic regression is only valid for independent observations.

Yuv	sameGender	simDelinquency	Уvи	$\sum_{w} y_{uw} \cdot y_{wv}$
1	1	0.5	1	3
1	1	0.5	1	0
0	0	0.1	1	1
1	0	0.1	0	2

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

In our case, the different observations (rows) are not independent.

This is even implied by the model itself.

The analysis so far is invalid.

Logistic regression is only valid for *independent observations*.

Y _{uv}	sameGender	simDelinquency	У _{vu}	$\sum_{w} y_{uw} \cdot y_{wv}$
1	1	0.5	1	3
1	1	0.5	1	0
0	0	0.1	1	1
1	0	0.1	0	2

▲□▶ ▲□▶ ▲ 三▶ ▲ 三▶ - 三 - のへぐ

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.



observed network



simulated network

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </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.

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 (Ω, P) , where

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

Notation

- $P(\omega)$ is called the *probability* of $\omega \in \Omega$.
- The probability of a subset Ω' ⊆ Ω is defined by
 P(Ω') = ∑_{ω∈Ω'} P(ω).

◆□▶ ◆□▶ ▲□▶ ▲□▶ ■ ののの

Background: finite probability space.

Definition

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

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

Notation

- $P(\omega)$ is called the *probability* of $\omega \in \Omega$.
- The probability of a subset Ω' ⊆ Ω is defined by
 P(Ω') = ∑_{ω∈Ω'} P(ω).

Example (dice)

$$egin{aligned} \Omega &= \{1,2,3,4,5,6\} \ P(\omega) &= 1/6 ext{ for all } \omega \in \Omega \ \Omega' &= \{1,3,5\} \end{aligned}$$

(possible outcomes when throwing a die) (uniform probability) (throwing an odd number) Background: finite probability space.

Definition

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

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

Notation

- $P(\omega)$ is called the *probability* of $\omega \in \Omega$.
- The probability of a subset Ω' ⊆ Ω is defined by
 P(Ω') = ∑_{ω∈Ω'} P(ω).

Example (lottery)

$$\Omega = \{X \subset \{1, \dots, 49\}; |X| = 6\}$$
 (sets of 6 different numbers)
 $P(\omega) = {49 \choose 6}^{-1} = \frac{6!43!}{49!}$ for all $\omega \in \Omega$ (uniform probability)
Background: graphs.

Definition

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

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

The elements that can be edges are called *dyads*.

Interpretation:

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



◆□▶ ◆□▶ ▲□▶ ▲□▶ ■ ののの

Background: adjacency matrices.

Definition Let G = (V, E) be a graph with an ordered set of vertices $V = \{v_1, ..., v_n\}$. The **adjacency matrix** associated with *G* is a binary $n \times n$ matrix $y = (y_{ij})_{i,j \in 1,...n}$ with

$$y_{ij} = \left\{ egin{array}{cc} 1 & ext{if } (v_i,v_j) \in E \ 0 & ext{if } (v_i,v_j)
ot\in E \end{array}
ight.$$



	Α	В	С	D
Α	0	0	1	0
В	1	0	0	0
С	1	1	0	0
D	1	0	0	0

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

(ロ)、

Assign probabilities to individual ties.

What are the possible outcomes of a network experiment?

We have a set of actors $V = \{v_1, \ldots, v_n\}$

▶ defining a set of dyads (pairs of actors)
D = {(v_i, v_j) ∈ V × V}.

Each of these dyads (v_i, v_j) is associated with a observation y_{ij} that has two possible outcomes:

- $y_{ij} = 1$ if there is a tie from v_i to v_j ;
- $y_{ij} = 0$ if there is no tie from v_i to v_j .

Each of the two outcomes happens with a certain probability where

$$P(Y_{ij} = 1) + P(Y_{ij} = 0) = 1$$

Assign probabilities to individual ties.

What are the possible outcomes of a network experiment?

Suppose we are given tie probabilities $p_{ij} = P(Y_{ij} = 1)$ for all dyads $(v_i, v_j) \in V \times V$.

Does this imply probabilities for higher-order structures?



Four different possible outcomes with the probabilities:

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 $p = p_{ij} = P(Y_{ij} = 1) = 136/(25 \cdot 24) = 0.23$

	predicted	empirical
$P(Y_{ij} = 1 \land Y_{ji} = 1)$	0.05	0.13
$P(Y_{ij} = 1 \land Y_{ji} = 0)$	0.18	0.10
$P(Y_{ij} = 0 \land Y_{ji} = 1)$	0.18	0.10
$P(Y_{ij} = 0 \land Y_{ji} = 0)$	0.59	0.67

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

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 (v_i, v_j) together with its reverse (v_j, v_i) .



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

$$\begin{array}{ll} Y_{ij} = 1 \land Y_{ji} = 1 & \text{both ties are there} \\ Y_{ij} = 1 \land Y_{ji} = 0 & (v_i, v_j) \in E \text{ but } (v_j, v_i) \notin E \\ Y_{ij} = 0 \land Y_{ji} = 1 & (v_i, v_j) \notin E \text{ but } (v_j, v_i) \in E \\ Y_{ij} = 0 \land Y_{ji} = 0 & \text{none of the two ties is there} \end{array}$$

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 (v_i, v_j) can also depend on ties with other actors.



Specifying different probabilities to all possible outcomes on a triplet (v_i , v_i , v_k) doesn't help either.

・ロト ・ 同 ・ ・ ヨ ・ ・ ヨ ・ うへつ

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

▲□▶ ▲□▶ ▲□▶ ▲□▶ ▲□ ● のへぐ

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 G be the set of all undirected, loopless graphs with vertex set $V = \{1, ..., n\}$ and let P be defined by

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

・ロト ・ 同 ・ ・ ヨ ・ ・ ヨ ・ うへつ

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 G be the set of all undirected, loopless graphs with vertex set $V = \{1, ..., n\}$ and let P be defined by

$$P\colon \mathcal{G} o \mathbb{R}; \ P(G) = rac{1}{2^{rac{n(n-1)}{2}}}$$

・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・

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 $V = \{1, ..., n\}$.

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

・ロト ・ 同 ・ ・ ヨ ・ ・ ヨ ・ うへつ

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

 $\mathcal{G}_{\boldsymbol{e}} = \{ \boldsymbol{G} \in \mathcal{G} \; ; \; \boldsymbol{e} \in \boldsymbol{E}_{\boldsymbol{G}} \}$.

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

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

・ロト ・ 同 ・ ・ ヨ ・ ・ ヨ ・ うへつ

Note: this does not hold the other way round.

Random graph models: edge probability.

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

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

 $\mathcal{G}_{\pmb{e}} = \{\pmb{G} \in \mathcal{G} \; ; \; \pmb{e} \in \pmb{E}_{\pmb{G}} \}$.

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

Thus, assigning a probability to each graph 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. $P: \mathcal{G} \to [0, 1]$ defines a probability for each graph.

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

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

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

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

A D F A 同 F A E F A E F A Q A

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 P₂ by

$$P_2(G) = \left\{ egin{array}{cc} 1/2 & ext{if } E = \emptyset ext{ or } E = D; \\ 0 & ext{else} \end{array}
ight.$$

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

A D F A 同 F A E F A E F A Q A

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.

◆□▶ ◆□▶ ▲□▶ ▲□▶ ■ ののの

Independence and non-independence of edges.

small facebook network

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



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

(日)

Independence and non-independence of edges.

small facebook network

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



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

Independence and non-independence of edges. Knecht Classroom Data

$$p_{uv} = \text{logit}^{-1}\left(\sum_{i=1}^{k} \theta_i \cdot s_i(u, v; y)\right)$$

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

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

▲□▶ ▲□▶ ▲ 三▶ ▲ 三▶ - 三■ - のへぐ

Definition

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

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

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

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

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

Definition

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

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

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

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

A D F A 同 F A E F A E F A Q A

Example (probability space: dice) $A_{odd} = \{1, 3, 5\}$ and $A_{\leq 4} = \{1, 2, 3, 4\}$ are independent.

Definition

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

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

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

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

A D F A 同 F A E F A E F A Q A

Example (probability space: dice) $A_{odd} = \{1, 3, 5\}$ and $A_{\leq 3} = \{1, 2, 3\}$ are **not** independent.

Definition

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

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

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

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

▲□▶ ▲□▶ ▲ 三▶ ▲ 三▶ - 三 - のへぐ

Definition

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

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

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

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

▲□▶ ▲□▶ ▲□▶ ▲□▶ ▲□ ● のへで

Example (probability space: dice) $P(A_{odd}|A_{\leq 4}) = 1/2$, but $P(A_{odd}|A_{\leq 3}) = 2/3$ Independence of dyads in random graph models.

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

$$\mathcal{G}_{e} = \{ \textit{G} \in \mathcal{G} \; ; \; \textit{e} \in \textit{E}_{G} \}$$
 .

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

If G_{e1} and G_{e2} are independent, we say that "the dyads e₁ and e₂ are independent"

Example: two random graph models revisited.

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

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

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

Let

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

It is for two different dyads e_1 and e_2

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

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

Structural balance theory (illustrating dependence).

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



not balanced

SBT claims that actors prefer balanced networks.

In an appropriate random graph model, it holds that

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





Structural balance theory (illustrating dependence).

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





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

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

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

Definition

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

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

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

Fully independent random graph models.

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

$$\mathcal{G}_{m{e}} = \{m{G} \in \mathcal{G} \ ; \ m{e} \in m{E}_{m{G}} \}$$
 .

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

Definition

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

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

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

Fully independent random graph models.

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

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

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

Definition

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

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

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

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

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

$$P(G) = P({G})$$

$$= P\left(\bigcap_{d \in E} \mathcal{G}_d \cap \bigcap_{d \in D \setminus E} \overline{\mathcal{G}}_d\right)$$

$$= \prod_{d \in E} P(\mathcal{G}_d) \cdot \prod_{d \in D \setminus E} P(\overline{\mathcal{G}}_d)$$

$$= \prod_{d \in E} P(d \in E) \cdot \prod_{d \in D \setminus E} 1 - P(d \in E) .$$

▲□▶▲圖▶▲≣▶▲≣▶ = ● のへで

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, ...

(ロ) (同) (三) (三) (三) (三) (○) (○)

Background: random variable and expectation.

Let (Ω, P) be a finite probability space. Definition A random variable is a function $X \colon \Omega \to \mathbb{R}$.

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

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

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

Example

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

Background: random variable and expectation.

Let (Ω, P) be a finite probability space. Definition A random variable is a function $X \colon \Omega \to \mathbb{R}$.

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

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

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

Example

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

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

Lemma

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

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

▲□▶▲□▶▲□▶▲□▶ □ のQ@

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

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



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 $\chi_{e} \colon \mathcal{G} \to \{0,1\}$ is defined by

$$\chi_{e}(G) = \left\{ egin{array}{c} 1 & ext{if } e \in E_{G} \ 0 & ext{else.} \end{array}
ight.$$

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

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



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

$$\mathbb{E}[m] = \sum_{e \in D} \mathbb{E}[\chi_e] = \sum_{e \in D} P(e \in E) \cdot 1 + P(e \notin E) \cdot 0$$
$$= \sum_{e \in D} P(e \in E)$$

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

$$\begin{array}{rcl} \mathcal{G}_{e} &=& \{G \in \mathcal{G} \; ; \; e \in E_{G} \}, \\ \overline{\mathcal{G}_{e}} &=& \{G \in \mathcal{G} \; ; \; e \notin E_{G} \} \end{array}$$

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

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

Uniform graph model: edge probability.

Claim

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

Proof.

The two sets

$$\begin{array}{rcl} \mathcal{G}_{\boldsymbol{e}} &=& \{\boldsymbol{G} \in \mathcal{G} \; ; \; \boldsymbol{e} \in \boldsymbol{E}_{\boldsymbol{G}} \}, \\ \overline{\mathcal{G}_{\boldsymbol{e}}} &=& \{\boldsymbol{G} \in \mathcal{G} \; ; \; \boldsymbol{e} \notin \boldsymbol{E}_{\boldsymbol{G}} \} \end{array}$$

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

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

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

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

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

Uniform graph model: independence.

Claim

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

Proof.

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

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

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

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).

(ロ) (同) (三) (三) (三) (○) (○)

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.

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.

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.

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.

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

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

G(n, p) is the random graph model on the set of undirected, loopless graphs with vertex set $V = \{1, ..., n\}$ 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)

$$\sum_{G \in \mathcal{G}} P(G) = \sum_{m=0}^{M} {M \choose m} p^m (1-p)^{M-m}$$
$$= (p+(1-p))^M = 1^M = 1$$

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

(日) (日) (日) (日) (日) (日) (日)

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

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

G(n, p) is the random graph model on the set of undirected, loopless graphs with vertex set $V = \{1, ..., n\}$ 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)

$$\sum_{G \in \mathcal{G}} P(G) = \sum_{m=0}^{M} {M \choose m} p^{m} (1-p)^{M-m}$$
$$= (p + (1-p))^{M} = 1^{M} = 1$$

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

(日) (日) (日) (日) (日) (日) (日)

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

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

G(n, p) is the random graph model on the set of undirected, loopless graphs with vertex set $V = \{1, ..., n\}$ 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)

$$\sum_{G \in \mathcal{G}} P(G) = \sum_{m=0}^{M} {M \choose m} p^{m} (1-p)^{M-m}$$
$$= (p + (1-p))^{M} = 1^{M} = 1$$

Remark

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

(日) (日) (日) (日) (日) (日) (日)

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).

◆□▶ ◆□▶ ▲□▶ ▲□▶ ■ ののの

Proof. See next exercise sheet.

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

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

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

Expected density is p.

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.

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*.



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

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

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

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 $r \le p$ add e to the edge set.

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

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

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

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 $\Theta(n^2)$ (independent of *p*)

 \Rightarrow inefficient for small p (i.e., sparse graphs).

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

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

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

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

enumerate dyads

<i>d</i> ₁				
d ₂	d ₃			
d ₄	<i>d</i> 5	<i>d</i> ₆		
d 7	<i>d</i> ₈	d ₉	<i>d</i> ₁₀	

(日) (日) (日) (日) (日) (日) (日)

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

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

*Is d*₁ *an edge?* (draw a random number...)

<i>d</i> ₁ ?				
d ₂	d ₃			
d ₄	<i>d</i> 5	<i>d</i> ₆		
d 7	d ₈	d ₉	<i>d</i> ₁₀	

(日) (日) (日) (日) (日) (日) (日)

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

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

 $\begin{array}{l} \textit{Is } \textit{d}_1 \textit{ an edge?} \\ \rightarrow \textit{NO (for instance)} \end{array}$

<i>d</i> ₁				
d ₂	d ₃			
d ₄	<i>d</i> 5	<i>d</i> ₆		
d 7	<i>d</i> ₈	d ₉	<i>d</i> ₁₀	

(日) (日) (日) (日) (日) (日) (日)

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

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

*Is d*₂ *an edge?* (draw a random number...)



◆□▶ ◆□▶ ▲□▶ ▲□▶ ■ ののの

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

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

Is d_2 *an edge*? → YES (for instance) ⇒ turn d_2 into the first edge



◆□▶ ◆□▶ ▲□▶ ▲□▶ ■ ののの

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

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

*Is d*₃ *an edge?* (draw a random number...)

<i>d</i> ₁				
<i>e</i> 1	<i>d</i> ₃ ?			
<i>d</i> ₄	d 5	<i>d</i> ₆		
d 7	<i>d</i> ₈	d ₉	<i>d</i> ₁₀	

◆□▶ ◆□▶ ▲□▶ ▲□▶ ■ ののの

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

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

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

<i>d</i> ₁				
<i>e</i> 1	d ₃			
d ₄	d 5	<i>d</i> ₆		
d 7	d ₈	d ₉	<i>d</i> ₁₀	

(日) (日) (日) (日) (日) (日) (日)

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

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

go on . . .

<i>d</i> ₁				
<i>e</i> 1	d ₃			
d ₄	<i>d</i> 5	<i>d</i> ₆		
d ₇	<i>d</i> ₈	d ₉	<i>d</i> ₁₀	

(ロ) (同) (三) (三) (三) (○) (○)

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

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

*Is d*₆ *an edge*? (draw a random number...)



◆□▶ ◆□▶ ▲□▶ ▲□▶ ■ ののの

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

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

 $\begin{array}{l} \textit{Is } \textit{d}_6 \textit{ an edge?} \\ \rightarrow \texttt{YES} \textit{ (for instance)} \\ \Rightarrow \textit{turn } \textit{d}_6 \textit{ into the second edge} \end{array}$

<i>d</i> ₁				
<i>e</i> 1	d ₃			
d ₄	d 5	<i>e</i> ₂		
d 7	d ₈	d ₉	<i>d</i> ₁₀	

◆□▶ ◆□▶ ▲□▶ ▲□▶ ■ ののの

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

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

to be continued ...

<i>d</i> ₁				
<i>e</i> 1	d ₃			
d ₄	d 5	<i>e</i> ₂		
d 7	<i>d</i> ₈	<i>d</i> 9	<i>d</i> ₁₀	

(日) (日) (日) (日) (日) (日) (日)

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

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

to be continued ...

<i>d</i> ₁				
<i>e</i> 1	d ₃			
d ₄	<i>d</i> 5	<i>e</i> ₂		
d 7	<i>d</i> ₈	<i>d</i> 9	<i>d</i> ₁₀	

(日) (日) (日) (日) (日) (日) (日)

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

Better ask the question:

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

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

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

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

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


Better ask the question:

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

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

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

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

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



Better ask the question:

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

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

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

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

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



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 $q^k p$.

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

draw k = 1 (happens with probability qp)

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

<i>d</i> ₁				
d ₂	d ₃			
d ₄	d 5	<i>d</i> ₆		
d ₇	<i>d</i> 8	d ₉	<i>d</i> ₁₀	

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 $q^k p$.

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

draw k = 1 (happens with probability qp)

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

<i>d</i> ₁				
d ₂	d ₃			
d ₄	<i>d</i> 5	<i>d</i> ₆		
<i>d</i> ₇	<i>d</i> 8	d ₉	<i>d</i> ₁₀	

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 $q^k p$.

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

draw k = 1 (happens with probability qp)

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



(ロ) (同) (三) (三) (三) (○) (○)

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 $q^k p$.

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

draw k = 1 (happens with probability qp)

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

<i>d</i> ₁				
<i>e</i> 1	d ₃			
d ₄	<i>d</i> 5	<i>e</i> ₂		
d 7	<i>d</i> 8	d ₉	<i>d</i> ₁₀	

・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・

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, ... is called the **geometric distribution**.

Such a random number generator is implemented in R (function $\tt rgeom$).

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

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

・ロト ・ 同 ・ ・ ヨ ・ ・ ヨ ・ うへつ

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



◆□▶ ◆□▶ ▲□▶ ▲□▶ ■ ののの

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

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



◆□▶ ◆□▶ ▲□▶ ▲□▶ ■ ののの

If $w \ge v$ 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 $v < n \ do$

$$k \leftarrow rgeom(p)$$

$$w \leftarrow w + k + 1$$
while $w \ge v \ and \ v < n \ do$

$$w \leftarrow w - v$$

$$v \leftarrow v + 1$$
if $v < n \ then$

$$E \leftarrow E \cup \{\{v, w\}\}$$
return $G = (V, E)$

 \Rightarrow runtime in $\Theta(m + n)$.

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.

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

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

For i = 1, ..., 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, ..., 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 + 1th iteration satisfies $k \ge k_{m+1}$.

・ロト ・ 同 ・ ・ ヨ ・ ・ ヨ ・ うへつ

Algorithm returns G if and only if

- For all i = 1,..., m, the random number k in the *i*th iteration satisfies k = k_i. Happens with probability pa^{k_i}.
- For *i* = *m* + 1 the random number *k* in the *m* + 1th iteration satisfies *k* ≥ *k*_{*m*+1}.
 Happens with probability *a*^k_{*m*+1}.

・ロト ・ 同 ・ ・ ヨ ・ ・ ヨ ・ うへつ

Algorithm returns G if and only if

- For all *i* = 1,..., *m*, the random number *k* in the *i*th iteration satisfies *k* = *k_i*. Happens with probability *pq^{k_i}*.
- 2. For i = m + 1 the random number k in the m + 1th iteration satisfies $k \ge k_{m+1}$.

・ロト ・ 同 ・ ・ ヨ ・ ・ ヨ ・ うへつ

Happens with probability $q^{K_{m+1}}$

Algorithm returns G if and only if

- For all *i* = 1,..., *m*, the random number *k* in the *i*th iteration satisfies *k* = *k_i*. Happens with probability *pq^{k_i}*.
- For *i* = *m* + 1 the random number *k* in the *m* + 1th iteration satisfies *k* ≥ *k*_{*m*+1}. Happens with probability *q*^{*k*_{*m*+1}}.
 </sup>

・ロト ・ 同 ・ ・ ヨ ・ ・ ヨ ・ うへつ

Algorithm returns G if and only if

- For all *i* = 1,..., *m*, the random number *k* in the *i*th iteration satisfies *k* = *k_i*. Happens with probability *pq^{k_i}*.
- For *i* = *m* + 1 the random number *k* in the *m* + 1th iteration satisfies *k* ≥ *k*_{*m*+1}. Happens with probability *q*^{*k*_{*m*+1}}.
 </sup>

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

・ロト ・ 同 ・ ・ ヨ ・ ・ ヨ ・ うへつ

Algorithm returns G if and only if

- For all *i* = 1,..., *m*, the random number *k* in the *i*th iteration satisfies *k* = *k_i*. Happens with probability *pq^{k_i}*.
- For *i* = *m* + 1 the random number *k* in the *m* + 1th iteration satisfies *k* ≥ *k*_{*m*+1}. Happens with probability *q*^{*k*_{*m*+1}}.
 </sup>

All conditions for i = 1, ..., m + 1 are satisfied with probability

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

(日) (日) (日) (日) (日) (日) (日)

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 16 600 edges.



ヘロト 人間 とくほとくほとう

Which G(n,p)? What is the most likely value for the parameter p? Can such a network be drawn from a $\mathcal{G}(n, p)$ model?

Graph has 769 vertices and about 16 600 edges.



・ロト ・ 四ト ・ ヨト ・ ヨト

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

Background: maximum likelihood.

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

What is the most likely parameter value?

Definition (maximum likelihood)

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

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

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

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

(日) (日) (日) (日) (日) (日) (日)

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

Background: maximum likelihood.

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

What is the most likely parameter value?

Definition (maximum likelihood)

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

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

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

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

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

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

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

$$L(p) = P_p(G_{obs}) = p^m (1-p)^{M-m} .$$

$$L'(p) = m \cdot p^{m-1} \cdot (1-p)^{M-m} - p^m \cdot (M-m) \cdot (1-p)^{M-m-1}$$

Setting L'(p) = 0 for 0 yields

$$m \cdot p^{m-1} \cdot (1-p)^{M-m} = p^m \cdot (M-m) \cdot (1-p)^{M-m-1}$$

$$m \cdot (1-p) = p \cdot (M-m)$$

$$m-pm = pM-pm$$

$$\frac{m}{M} = p$$

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:

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

- 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).

・ロット (雪) (日) (日)

- -

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.

・ロン・(部・・市・・日・

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.

・ コット (雪) (小田) (コット 日)

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



▲□▶ ▲□▶ ▲三▶ ▲三▶ - 三 - のへで

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).



max degree is 65 min degree is 21

max degree is 248 min degree is 1

ъ

・ロット (雪) (日) (日)

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

Lemma

Let $v \in \{1, ..., n\}$ be any vertex. The probability that v has degree equal to $k \in \{0, ..., n-1\}$ 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

・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・

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

Lemma

Let $v \in \{1, ..., n\}$ be any vertex. The probability that v has degree equal to $k \in \{0, ..., n-1\}$ 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

・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・



Histogram of degree2

Histogram of degree1





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

Theorem

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

Then it is

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

◆□▶ ◆□▶ ◆□▶ ◆□▶ ● ● ● ●

(Is called Poisson distribution.)

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

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

Histogram of degree.frequency



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



Histogram of degree2








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



Comparing number of triangles.

expected: 13,000

observed: 119,000

-

A D > A P > A D > A D >

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.

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

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.

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.

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

(ロ) (同) (三) (三) (三) (○) (○)

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.

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.

・ロト ・ 理 ト ・ 理 ト ・ 理 ト

Planted partition models.

Definition

A planted partition model is defined by

- ► A partition of the vertex set $V = V_1 \cup \cdots \cup V_k$ into *k* disjoint classes.
- ► Probabilities p_{ij} ∈ (0, 1) associated with each unordered pair of classes V_i and V_j.
- ► Two vertices u ∈ V_i and v ∈ V_j are connected by an edge with probability p_{ij}.

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

The model is fully independent.

Planted partition models.

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

$p_1 \cdots p_1$	$p_2 \cdots p_2$	$p_3 \cdots p_3$
: :	: :	: :
$p_1 \cdots p_1$	$p_2 \cdots p_2$	$p_3 \cdots p_3$
$p_2 \cdots p_2$	$p_4 \cdots p_4$	$p_5 \cdots p_5$
: :	: :	: :
$p_2 \cdots p_2$	$p_4 \cdots p_4$	$p_5 \cdots p_5$
$p_3 \cdots p_3$	$p_5 \cdots p_5$	$p_6 \cdots p_6$
: :	: :	: :
$p_3 \cdots p_3$	$p_5 \cdots p_5$	$p_6 \cdots p_6$

▲□▶ ▲□▶ ▲□▶ ▲□▶ ▲□ ● のへぐ

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.

Recall: degree distributions.

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

empirical network





Histogram of degree1

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ ─臣 ─のへで

Other empirical dist. (Barabasi and Albert, 1999).

Note: logarithmic scaling of axes.



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

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

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

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

Model idea (Barabási and Albert, 1999):

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

Interpretation high-degree vertices are more popular.

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

◆□ → ◆□ → ◆ = → ◆ = → のへで

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

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

Model idea (Barabási and Albert, 1999):

- 1. vertices are successively added to the network;
- 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 $\gamma \approx 3$.

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

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

Model idea (Barabási and Albert, 1999):

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

Interpretation high-degree vertices are more popular.

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

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

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

Model idea (Barabási and Albert, 1999):

- 1. vertices are successively added to the network;
- 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 $n \ge 1$ vertices and constant outdegree equal to $b \ge 1$.

```
foreach i = 0, ..., b - 1 do
```

くして 「「」 (山下 (山下 (山下 (山下

Preferential attachment model.

Definition (Bollobás, Riordan, Spencer, and Tusnády)

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

(日) (日) (日) (日) (日) (日) (日)

Iterative definition:

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

```
foreach v = 0, ..., 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

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

put e = (v, w) into E;
```

Preferential attachment (algorithm).

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

input : number of nodes $n \in \mathbb{N}_{>1}$, out-degree $b \in \mathbb{N}_{>1}$ data : array A[0...2nb-1] //collects endpoints of edges output multi-graph $G = (\{0, \ldots, n-1\}, E)$ $E \leftarrow \emptyset; m \leftarrow 0$ *//edge set and edge counter* foreach v = 0, ..., n - 1 do foreach i = 0, ..., b - 1 do $A[2m] \leftarrow v$ //v is source of next edge $w \leftarrow A[random(\{0, \dots, 2m\})]$ //randomly select target $A[2m+1] \leftarrow w;$ $A[2m + 1] \leftarrow w,$ $E \leftarrow E \cup \{(v, w)\}; m \leftarrow m + 1$ //put target in A //update edges

Note: number of occurences of *v* in *A* equals degree of $v \Rightarrow$ target node gets selected with the correct probability.

Some remarks.

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

But different properties might be interrelated:

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

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

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

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.

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.

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.

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 $(\mathcal{G}, P_{\theta})$ whose probability function P_{θ} can be written as

$$\mathcal{P}_{ heta}(G) = rac{1}{\kappa(heta)} \exp\left(\sum_{i=1}^k heta_i \cdot s_i(G)
ight)$$

with

- $s_i : \mathcal{G} \to \mathbb{R}$ for i = 1, ..., k (*statistics*);
- $\theta_i \in \mathbb{R}$ for i = 1, ..., k (parameters); $\theta = (\theta_1, ..., \theta_k)$;

normalizing constant κ defined by

$$\kappa(heta) = \sum_{G' \in \mathcal{G}} \exp\left(\sum_{i=1}^k heta_i \cdot s_i(G')\right)$$

・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・

ERGM (example).

Consider undirected, loopless graphs with 3 vertices.

 $P(G) = \frac{1}{\kappa} \exp\left[-\log(2) \cdot m(G) + \log(16) \cdot \text{triangles}(G)\right]$

	• •	•		
<i>m</i> (<i>G</i>)	0	1	2	3
triangles(G)	0	0	0	1
$P(G) \cdot \kappa$	1	$\frac{1}{2}$	$\frac{1}{2^2}$	$\frac{16}{2^3}$
# isomorphic graphs	1	3	3	1

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

Probability of a graph G

$$\mathcal{P}_{ heta}(G) = rac{1}{\kappa(heta)} \exp\left(\sum_{i=1}^k heta_i \cdot oldsymbol{s}_i(G)
ight)$$

What happens if one single statistic s_{i_0} changes?

$$\mathcal{P}_{ heta}(G) = \exp[heta_{i_0} \cdot s_{i_0}(G)] \cdot rac{1}{\kappa(heta)} \exp\left(\sum_{i
eq i_0} heta_i \cdot s_i(G)
ight)$$

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

It is $\exp(\theta_{i_0}) > 1 \Leftrightarrow \theta_{i_0} > 0$ and $\exp(\theta_{i_0}) < 1 \Leftrightarrow \theta_{i_0} < 0$.

Probability of a graph G

$$\mathcal{P}_{ heta}(G) = rac{1}{\kappa(heta)} \exp\left(\sum_{i=1}^k heta_i \cdot oldsymbol{s}_i(G)
ight)$$

What happens if one single statistic s_{i_0} changes?

$$egin{split} m{\mathcal{P}}_{ heta}(G) = \exp[heta_{i_0}\cdotm{s}_{i_0}(G)]\cdotrac{1}{\kappa(heta)}\exp\left(\sum_{i
eq i_0} heta_i\cdotm{s}_i(G)
ight) \end{split}$$

A D F A 同 F A E F A E F A Q A

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

It is $\exp(\theta_{i_0}) > 1 \Leftrightarrow \theta_{i_0} > 0$ and $\exp(\theta_{i_0}) < 1 \Leftrightarrow \theta_{i_0} < 0$.

Probability of a graph G

$$\mathcal{P}_{ heta}(G) = rac{1}{\kappa(heta)} \exp\left(\sum_{i=1}^k heta_i \cdot oldsymbol{s}_i(G)
ight)$$

What happens if one single statistic s_{i_0} changes?

$$P_{ heta}(G) = \exp[heta_{i_0} \cdot s_{i_0}(G)] \cdot rac{1}{\kappa(heta)} \exp\left(\sum_{i
eq i_0} heta_i \cdot s_i(G)
ight)$$

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

It is $\exp(\theta_{i_0}) > 1 \Leftrightarrow \theta_{i_0} > 0$ and $\exp(\theta_{i_0}) < 1 \Leftrightarrow \theta_{i_0} < 0$.

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 $G^{(-e)} = (V, E \setminus \{e\})$.

Define the change statistic by

$$\Delta s(e; G) = [s_1(G^{(+e)}) - s_1(G^{(-e)}), \dots, s_k(G^{(+e)}) - s_k(G^{(-e)})]$$

Then, it is $\frac{P_{\theta}(G^{(+e)})}{P_{\theta}(G^{(-e)})} = \exp(\theta \cdot \Delta s(e; G)).$



Note: probability-ratio depends on *e* and *G*. Example: $s_1(G) = m(G)$ and $s_2(G) = \text{triangle}(G)$.

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

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

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

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

Proof.

Let P_2 denote the probability function of the $\mathcal{G}(n, p)$ model.

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

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

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

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

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

Proof.

Let P_2 denote the probability function of the $\mathcal{G}(n, p)$ model.

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

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

・ロット (雪) (日) (日) (日)

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

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

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

Proof.

Let P_2 denote the probability function of the $\mathcal{G}(n, p)$ model.

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

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

・ロット (雪) (日) (日) (日)

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

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

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

Proof.

Let P_2 denote the probability function of the $\mathcal{G}(n, p)$ model.

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

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

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

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

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

Proof.

Let P_2 denote the probability function of the $\mathcal{G}(n, p)$ model.

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

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

(日) (日) (日) (日) (日) (日) (日)

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

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

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

where $heta = \log\left(rac{p}{1-p}
ight)$.

Relation between θ and p

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

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

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

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

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

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

Relation between θ and p

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

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

(日) (日) (日) (日) (日) (日) (日)

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

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

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

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

Relation between θ and p

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

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

Commonly used network statistics.

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

$$P_{ heta}(G) = rac{1}{\kappa(heta)} \exp\left(\sum_{i=1}^{k} heta_i \cdot s_i(G)
ight)$$

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

◆□▶ ◆□▶ ▲□▶ ▲□▶ □ のQ@

Commonly used network statistics (I).

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

m(V, E) = |E|.

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

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

Commonly used network statistics (II).

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

$$t(y) = \sum_{u < v < w} y_{uv} \cdot y_{vw} \cdot y_{wu}$$
; y 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.

(ロ) (同) (三) (三) (三) (三) (○) (○)

Commonly used network statistics (IV).

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

$$s_{\ell}(y) = \sum_{u} \sum_{v_1 < \cdots < v_{\ell} \neq u} y_{uv_1} \cdot \ldots \cdot y_{uv_{\ell}}$$
.

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 ℓ -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 $(v, u) \in E$.

$$\mathsf{mutual}(y) = \sum_{u \neq v} y_{uv} \cdot y_{vu}$$

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

$$u \xrightarrow{} v$$

◆□▶ ◆□▶ ▲□▶ ▲□▶ ■ ののの

Implication on dyad dependency.

Using some statistics makes dyads dependent.



Dyad dependency (example).

Consider undirected graphs with 3 vertices; 2-star count s_2 .

$$P(G) = \frac{1}{\kappa} \exp \left[\log(2) \cdot s_2(G) \right]$$

$$s_2(G) = \frac{1}{\kappa} \exp \left[\log(2) \cdot s_2(G) \right]$$

$$s_2(G) = \frac{1}{\kappa} \exp \left[\log(2) \cdot s_2(G) \right]$$

$$s_2(G) = \frac{1}{\kappa} \exp \left[\log(2) \cdot s_2(G) \right]$$

$$s_2(G) = \frac{1}{\kappa} \exp \left[\log(2) \cdot s_2(G) \right]$$

$$s_2(G) = \frac{1}{\kappa} \exp \left[\log(2) \cdot s_2(G) \right]$$

$$s_2(G) = \frac{1}{\kappa} \exp \left[\log(2) \cdot s_2(G) \right]$$

$$s_2(G) = \frac{1}{\kappa} \exp \left[\log(2) \cdot s_2(G) \right]$$

$$s_2(G) = \frac{1}{\kappa} \exp \left[\log(2) \cdot s_2(G) \right]$$

$$s_2(G) = \frac{1}{\kappa} \exp \left[\log(2) \cdot s_2(G) \right]$$

$$s_2(G) = \frac{1}{\kappa} \exp \left[\log(2) \cdot s_2(G) \right]$$

$$s_2(G) = \frac{1}{\kappa} \exp \left[\log(2) \cdot s_2(G) \right]$$

$$s_2(G) = \frac{1}{\kappa} \exp \left[\log(2) \cdot s_2(G) \right]$$

$$s_2(G) = \frac{1}{\kappa} \exp \left[\log(2) \cdot s_2(G) \right]$$

$$s_2(G) = \frac{1}{\kappa} \exp \left[\log(2) \cdot s_2(G) \right]$$

$$s_2(G) = \frac{1}{\kappa} \exp \left[\log(2) \cdot s_2(G) \right]$$

$$s_2(G) = \frac{1}{\kappa} \exp \left[\log(2) \cdot s_2(G) \right]$$

Let e, e' be two different dyads.

$$\begin{array}{rcl} P(\mathcal{G}_{e}|\mathcal{G}_{e'}) &=& (2+8)/(1+2\cdot 2+8) = 10/13 \\ P(\mathcal{G}_{e}) &=& (1+2\cdot 2+8)/(1+3\cdot 1+3\cdot 2+8) = 13/18 \end{array}$$

Thus, dyads e and e' are statistically dependent.

Estimation of ERGM parameters.

Given an observed network G_{obs} and a set of statistics s_i , i = 1, ..., k.

$$\mathcal{P}_{ heta}(G) = rac{1}{\kappa(heta)} \exp\left(\sum_{i=1}^k heta_i \cdot oldsymbol{s}_i(G)
ight)$$

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

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

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

$$P_{ heta}(G) = rac{1}{\kappa(heta)} \exp\left(\sum_{i=1}^k heta_i \cdot s_i(G)
ight) \;\;,$$

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

we want to design a probabilitstic algorithm

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

First try: factorize the probability.

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

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

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

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

Then, for a given graph G it is

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

・ロト ・ 同 ・ ・ ヨ ・ ・ ヨ ・ うへつ

First try: factorize the probability.

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

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

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

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

Then, for a given graph G it is

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

Sampling from an ERGM: first try.

For a given graph G it is

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

Sample from a given ERGM:

$$E \leftarrow \emptyset; \ \overline{E} \leftarrow \emptyset$$

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

Problem: probabilities are computationally intractable.

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

Sampling from an ERGM: first try.

For a given graph G it is

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

Sample from a given ERGM:

$$E \leftarrow \emptyset; \ \overline{E} \leftarrow \emptyset$$

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

Problem: probabilities are computationally intractable.

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

More details: conditional probabilities are computationally intractable in general.

Probability of a graph G in an ERGM:

$$P_{ heta}(G) = rac{1}{\kappa(heta)} \exp\left(\sum_{i=1}^{k} heta_i \cdot s_i(G)
ight)$$

Normalizing constant κ cancels out when computing conditional probabilities

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

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

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

Sampling from an ERGM: second try.

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

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

(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, ..., M do

with probability P(\mathcal{G}_{d_i} | \mathcal{G}[(V, E), -d_i])

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

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

(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, ..., M do
with probability P(\mathcal{G}_{d_i} | \mathcal{G}[(V, E), -d_i])
put d_i into E;
otherwise
remove d_i from E (i. e., do nothing).
```

Problem: graphs are not returned with the correct probabilities

・ロト ・ 同 ・ ・ ヨ ・ ・ ヨ ・ うへつ

Sampling from an ERGM: second try.

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

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

(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, ..., M do
with probability P(\mathcal{G}_{d_i} | \mathcal{G}[(V, E), -d_i])
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(\mathcal{G}_d | \mathcal{G}[(V, E), -d])

put d into E;

otherwise

remove d from E.
```

Fact: graphs are still not returned with the correct probabilities. But probability converges to the correct probability when $T \to \infty$.

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

Sampling from an ERGM: yet another try.

```
Sample from a given ERGM:

start with some arbitrary graph (V, E)

for some number of steps T do

draw a random dyad d \in D

with probability P(\mathcal{G}_d | \mathcal{G}[(V, E), -d])

put d into E;

otherwise

remove d from E.
```

Fact: graphs are still not returned with the correct probabilities. But probability converges to the correct probability when $T \rightarrow \infty$.

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

Informally, a Markov chain consists of a set of states and transition probabilities to jump from one state to another.

Here, given an ERGM (\mathcal{G}, P)

- the set of states is \mathcal{G} (all graphs);
- transition probabilities π are a function of P
- in such a way that
 - It the probability to be on graph G converges to P(G), when the number of simulation steps tends to ∞.
- \Rightarrow Simulate many steps and return the current graph.



・ コ ト ・ 雪 ト ・ 目 ト ・

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 G (all graphs);
- transition probabilities π 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 ∞ .
- \Rightarrow Simulate many steps and return the current graph.



◆□▶ ◆□▶ ▲□▶ ▲□▶ ■ ののの

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 G (all graphs);
- transition probabilities π 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 ∞ .

 \Rightarrow Simulate many steps and return the current graph.



◆□▶ ◆□▶ ▲□▶ ▲□▶ ■ ののの

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 G (all graphs);
- transition probabilities π 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 ∞.

 \Rightarrow Simulate many steps and return the current graph.



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 G (all graphs);
- transition probabilities π 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 ∞.
- \Rightarrow Simulate many steps and return the current graph.



◆□▶ ◆□▶ ◆□▶ ◆□▶ ● ● ● ●

Finite stationary Markov chain (simplified definition).

Note: Markov chains are usually defined as *random processes* that satisfy certain properties. The following is a more intuitive definition for stationary Markov chains.

Definition

A (finite stationary) Markov chain is a pair (\mathcal{G}, π), where

- \mathcal{G} is a finite set $\mathcal{G} = \{G_1, \ldots, G_N\}$ (*state space*);
- π is a matrix $\pi \in \mathbb{R}^{N \times N}$ (*transition matrix*) satisfying
 - For all *i*, *j* it is π_{ij} ∈ [0, 1];

• for all *i* it is
$$\sum_{j=1}^{N} \pi_{ij} = 1$$
.

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



◆□▶ ◆□▶ ◆三≯ ◆三≯ ● ● ● ●

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}, π) , where

- \mathcal{G} is a finite set $\mathcal{G} = \{G_1, \ldots, G_N\}$ (*state space*);
- π is a matrix $\pi \in \mathbb{R}^{N \times N}$ (*transition matrix*) satisfying
 - ▶ for all *i*, *j* it is π_{ij} ∈ [0, 1];
 - for all *i* it is $\sum_{j=1}^{N} \pi_{ij} = 1$.

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



◆□▶ ◆□▶ ◆ □▶ ◆ □▶ ● □ ● ● ●

How to define the transition probabilities.

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

▶ define transition probabilities π on the set of graphs 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 ∞.

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

An important observation.

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



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

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

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

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

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

$${m P}={m P}\pi$$
 .

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

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

Irreducible and aperiodic Markov chains.

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



A sequence of states $G_{i_1}, G_{i_2}, \ldots, G_{i_k}$ is called a *(directed) path* if for all $j = 1, \ldots, k - 1$ it is $\pi_{i_j i_{j+1}} > 0$.

Definition

The Markov chain (\mathcal{G}, π) is called

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

Irreducible and aperiodic Markov chains.

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



A sequence of states $G_{i_1}, G_{i_2}, \ldots, G_{i_k}$ is called a *(directed) path* if for all $j = 1, \ldots, k - 1$ it is $\pi_{i_j i_{j+1}} > 0$.

Definition

The Markov chain (\mathcal{G}, π) is called

- *irreducible* if for any two states G_i, G_j ∈ 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 π be the transition matrix of a Markov chain on \mathcal{G} . If for all graphs G_i , G_i it is

$$\mathsf{P}(\mathsf{G}_{i})\pi_{ij}=\mathsf{P}(\mathsf{G}_{j})\pi_{ji}$$

(Markov chain is then called **reversible**) and the Markov chain is **irreducible** and **aperiodic** then P is the unique stationary distribution of the Markov chain and for any initial distribution P_0 it is

$$\lim_{K\to\infty} P_0 \pi^K = P$$

◆□▶ ◆□▶ ◆□▶ ◆□▶ ● ● ● ●
Gibbs sampling (transition matrix).

Given *P*, define π such that

$$\mathsf{P}(\mathsf{G}_{i})\pi_{ij}=\mathsf{P}(\mathsf{G}_{j})\pi_{ji}$$
 .

Gibbs sampling: define π as follows

- $\pi_{ij} = 0$ if G_i and G_j differ in more than one dyad;
- ▶ if G_i and G_i differ in exactly one dyad, then

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

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

(sum over all G that differ from G_i in exactly one dyad)

Show: π is normalized, irreducible, aperiodic, reversible.

Gibbs sampling (algorithm).

initialize G by any graph from G; repeat *many* times

- select a dyad e uniformly at random;
- with probability $\frac{P(G^{(+e)})}{P(G^{(+e)})+P(G^{(-e)})}$
 - replace G = (V, E) by $G^{(+e)} = (V, E \cup \{e\})$
 - otherwise replace G = (V, E) by $G^{(-e)} = (V, E \setminus \{e\});$

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).

・ロト ・ 同 ・ ・ ヨ ・ ・ ヨ ・ うへつ

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)

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

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

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

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

・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・
 ・

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

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, ..., 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 $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_k)$.
- Check whether θ̂_k is significantly positive: compute probability of observing a network in the **null model** defined by (θ̂₁,..., θ̂_{k-1}, 0) that gives rise to θ_k as large as θ̂_k.

Maximizing the likelihood function

$$L(\theta) = rac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} heta_i \cdot s_i(G^*)
ight)$$

is computationally intractable (since the normalizing constant κ has too many terms).

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

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

Partial derivatives of the likelihood function. Likelihood function

$$L(\theta) = \frac{1}{\kappa(\theta)} \exp\left(\sum_{i=1}^{k} \theta_i \cdot s_i(G^*)\right) \text{ with}$$

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

Log-likelihood function

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

Partial derivative by θ_j for $j = 1, \ldots, k$

$$\begin{array}{lll} \frac{\partial}{\partial \theta_j} \ell(\theta) &=& s_j(G^*) - \frac{1}{\kappa(\theta)} \cdot \sum_{G' \in \mathcal{G}} \exp\left(\sum_{i=1}^k \theta_i \cdot s_i(G')\right) \cdot s_j(G') \\ &=& s_j(G^*) - \mathbb{E}_{\theta}(s_j) \ . \end{array}$$

Maximizing the likelihood function. Likelihood function

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

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

$$rac{\partial}{\partial heta_j} \ell(\hat{ heta}) = m{s}_j(m{G}^*) - \mathbb{E}_{\hat{ heta}}(m{s}_j) = m{0}, ext{ for } j = 1, \dots, k$$

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_θ) ;
- take $\overline{s} = \frac{1}{N} \cdot \sum_{j=1}^{N} s(G_j)$ as an estimate for $\mathbb{E}_{\theta}(s)$.

Newton-Raphson method.

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

- 1. Choose initial parameter values $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_k^{(0)});$
- 2. For $i = 0, 1, \ldots$ until convergence

2.1 Sample
$$G_1, \ldots, G_N \sim (\mathcal{G}, P_{\theta^{(j)}});$$

2.2 $\overline{s} \leftarrow \frac{1}{N} \cdot \sum_{j=1}^N s(G_j);$
2.3 $C \leftarrow \frac{1}{N} \cdot \sum_{j=1}^N s(G_j) \cdot s(G_j)^T;$
2.4 $H \leftarrow \overline{s} \cdot \overline{s}^T - C;$
2.5 $\theta^{(i+1)} \leftarrow \theta^{(i)} - H^{-1} \cdot (s(G^*) - \overline{s})$

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.

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.

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

Near-degeneracy and multi-modality of ERGMs

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

Examples

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



Near-degeneracy and multi-modality of ERGMs

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

Examples

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



Near-degeneracy and multi-modality of ERGMs

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

Examples

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



・ロット (雪) ・ (日) ・ (日)

ъ

Near-degeneracy and multi-modality of ERGMs.

Consider the following ERGM

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

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

Avoiding near-degeneracy of ERGMs.

Triangle statistic implies linear marginal effect of closed triangles:

- closing one triangle contributes τ to the log-probability;
- closing two triangles contributes 2τ ...



Geometrically-weighted edgewise shared partner (gwesp) statistic:

・ロト ・ 日 ・ ・ 日 ・ ・ 日 ・ ・ 日 ・

- ▶ 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.

$$P_{ heta}(G) = rac{1}{\kappa(heta)} \exp\left(\sum_{i=1}^k heta_i \cdot s_i(G)
ight)$$



Conditional independence of edges (informally).

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



・ ロ ト ・ 雪 ト ・ 雪 ト ・ 日 ト

Conditional independence of edges.

Let (\mathcal{G}, P) be a random graph model where *D* is the set of dyads of graphs in \mathcal{G} and assume that P(G) > 0 for all $G \in \mathcal{G}$. Let $d_1, d_2 \in D$ be two different dyads.

For a partition $D^+ \uplus D^- = D \setminus \{d_1, d_2\}$ of the set of dyads different from d_1 and d_2 let the subset \mathcal{G}_{D^+,D^-} be defined by

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

We say that d_1 and d_2 are *conditionally independent* (given the rest of the graph) if for all partitions $D^+ \uplus D^- = D \setminus \{d_1, d_2\}$ it is

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

Informally: if we know the state of all dyads in $D \setminus \{d_1, d_2\}$, the state of the dyad d_2 does not provide any additional information about the probability $P(d_1 \in E)$.

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

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

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

・ロト ・ 同 ・ ・ ヨ ・ ・ ヨ ・ うへつ

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

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

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

◆□▶ ◆□▶ ▲□▶ ▲□▶ ■ ののの

Theorem (second part)

Conversely, if the probability P on G is defined by

$$\begin{split} \mathcal{P}(G) &= \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A\right) \ , \qquad \textit{where} \\ \kappa &= \sum_{G' \in \mathcal{G}} \exp\left(\sum_{A \subseteq E(G')} \alpha_A\right) \ , \end{split}$$

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 $\alpha_A \neq 0$.

・ロト ・ 同 ・ ・ ヨ ・ ・ ヨ ・ うへつ

Theorem (second part)

Conversely, if the probability P on G is defined by

$$\begin{split} \mathcal{P}(G) &= \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A\right) \ , \qquad \textit{where} \\ \kappa &= \sum_{G' \in \mathcal{G}} \exp\left(\sum_{A \subseteq E(G')} \alpha_A\right) \ , \end{split}$$

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 $\alpha_A \neq 0$.

・ロト ・ 同 ・ ・ ヨ ・ ・ ヨ ・ うへつ

Conclusion from the Hammersley-Clifford Theorem.

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

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

 \Rightarrow every random graph (\mathcal{G}, P) with P > 0 is an ERGM:

▶ statistics: for $A \subseteq D$ define $s_A(G) = \begin{cases} 1 & A \subseteq E(G) \\ 0 & \text{else} \end{cases}$

• parameters: α_A

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

◆□ ▶ ◆□ ▶ ◆三 ▶ ◆□ ▶ ◆□ ●

Conclusion from the Hammersley-Clifford Theorem.

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

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

 \Rightarrow every random graph (\mathcal{G} , P) with P > 0 is an ERGM:

- ► statistics: for $A \subseteq D$ define $s_A(G) = \begin{cases} 1 & A \subseteq E(G) \\ 0 & \text{else} \end{cases}$
- parameters: α_A

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

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, $\{d_i, d_j\} \in E$ if d_i and d_j are not conditionally independent, given the rest of the graph.

Example

the dependence graph of a Markov graph on vertices $V = \{1, 2, 3, 4\}$ is



A subset $A \subseteq D$ is a clique in the dependence graph if A does not contain two conditionally independent dyads.

Cliques in the dependence graph of a Markov graph.

Markov random graphs: edges $\{i, j\}$ and $\{u, v\}$ are conditionally independent, unless they have a vertex in common.

Cliques in the dependence graph of a Markov graph are



No other subgraphs are cliques in the dependence graph.

◆□▶ ◆□▶ ▲□▶ ▲□▶ ■ ののの

ERGM of general Markov graphs.

Corollary Let (\mathcal{G}, P) be a Markov random graph on vertices $V = \{1, ..., n\}$. Then there are real constants

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

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

$$P(G) = \frac{1}{\kappa} \exp\left(\sum_{uv \in E(G)} \eta_{uv} + \sum_{uvw \in T(G)} \tau_{uvw} + \sum_{\ell=2}^{n-1} \sum_{uv_1 \dots v_\ell \in S_\ell(G)} \sigma_{uv_1 \dots v_\ell}\right)$$

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

ERGM of general Markov graphs (remarks).

$$P(G) = \frac{1}{\kappa} \exp\left(\sum_{uv \in E(G)} \eta_{uv} + \sum_{uvw \in T(G)} \tau_{uvw} + \sum_{\ell=2}^{n-1} \sum_{uv_1 \dots v_\ell \in S_\ell(G)} \sigma_{uv_1 \dots v_\ell}\right)$$

▲□▶ ▲□▶ ▲ 三▶ ▲ 三▶ - 三 - のへぐ

Each dyad, triangle, ℓ -star can contribute differently to the probability of a graph.

 \Rightarrow unreasonably high number of parameters.

Homogeneous random graph model.

Two graphs G = (V, E) and H = (W, F) are called *isomorphic* if there is a bijection $\varphi \colon V \to W$ such that

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

Definition

A random graph model (\mathcal{G}, P) is called *homogeneous* if for any pair of isomorphic graphs *G* and *H* it is P(G) = P(H).

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </p>

ERGM of homogeneous Markov graphs.

Corollary

Let (\mathcal{G}, P) be a homogeneous Markov random graph. Then there are real constants η , τ , and σ_{ℓ} for $\ell = 2, ..., n-1$ such that the probability of a graph $G \in \mathcal{G}$ can be written as

$$P(G) = \frac{1}{\kappa} \exp\left(\eta \cdot m(G) + \tau \cdot t(G) + \sum_{\ell=2}^{n-1} \sigma_{\ell} \cdot s_{\ell}(G)\right)$$

Proof.

Start from the ERGM of a general Markov graph.

Show that any two edge parameters are equal...

For $\ell = 2, \ldots, n-1$, show that any two ℓ -star parameters are equal...

・ロト ・ 同 ・ ・ ヨ ・ ・ ヨ ・ うへつ

Show that any two triangle parameters are equal...

ERGM of homogeneous Markov graphs.

Corollary

Let (\mathcal{G}, P) be a homogeneous Markov random graph. Then there are real constants η , τ , and σ_{ℓ} for $\ell = 2, ..., n-1$ such that the probability of a graph $G \in \mathcal{G}$ can be written as

$$P(G) = \frac{1}{\kappa} \exp\left(\eta \cdot m(G) + \tau \cdot t(G) + \sum_{\ell=2}^{n-1} \sigma_{\ell} \cdot s_{\ell}(G)\right)$$

Proof.

Start from the ERGM of a general Markov graph.

Show that any two edge parameters are equal...

For $\ell = 2, \ldots, n-1$, show that any two ℓ -star parameters are equal...

Show that any two triangle parameters are equal...
ERGMs of fully independent models.

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

$$\mathcal{P}(\mathcal{G}) = rac{1}{\kappa} \exp \left(\sum_{\mathit{uv} \in \mathcal{E}(\mathcal{G})} \eta_{\mathit{uv}}
ight) \; \; ,$$

for constants η_{uv} associated with dyads $uv \in D$.

 $\mathcal{G}(n, p)$ is a *homogeneous* fully independent model. Thus,

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

< □ > < 同 > < 三 > < 三 > < 三 > < ○ < ○ </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.

Estimating parameters from several observed networks.

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 の�?

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 E_∂(s) must be equal to the average observed statistics of the N networks

 \Rightarrow use the <code>target.stats</code> argument of <code>ergm.</code>

• 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.

- 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.)

(ロ)、(型)、(E)、(E)、 E) のQの

Pseudolikelihood.

Let P_{θ} 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 $G^{(-d)} = (V, E \setminus \{d\})$ and $\mathcal{G}[G, -d] = \{G^{(+d)}, G^{(-d)}\}$,

(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

$$\mathcal{L}_{G^*}^{(\mathsf{pseudo})}(heta) = \prod_{d \in E} \mathcal{P}_{ heta}(\mathcal{G}_d | \mathcal{G}[G^*, -d]) \cdot \prod_{d \in D \setminus E} 1 - \mathcal{P}_{ heta}(\mathcal{G}_d | \mathcal{G}[G^*, -d]) + \mathcal{O}_{G^*}(\mathcal{G}_d | \mathcal{G}[G^*, -d])$$

Note that the associated "probability" is not a proper probability on \mathcal{G} .

Interpreting pseudolikelihood.

The **pseudolikelihood** "probability" of a graph G = (V, E)

$$\mathcal{P}^{(\mathsf{pseudo})}_{ heta}(G) = \prod_{d \in E} \mathcal{P}_{ heta}(\mathcal{G}_d | \mathcal{G}[G, -d]) \cdot \prod_{d \in D \setminus E} 1 - \mathcal{P}_{ heta}(\mathcal{G}_d | \mathcal{G}[G, -d]) \; \; ,$$

- ▶ pretends that the graph *G* is the result of independent decisions for all dyads *d* ∈ *D*.
- When deciding about whether *d* ∈ *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_θ(G_d|G[G, -d]).

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* ∈ *D* for which either *d* ∈ *E* (variable equal to 1) or *d* ∉ *E* (variable equal to 0);
- ► the explanatory variables for the dyads d ∈ D are the k change statistics

$$\Delta s(d;G) = [s_1(G^{(+d)}) - s_1(G^{(-d)}), \dots, s_k(G^{(+d)}) - s_k(G^{(-d)})] ,$$

(ロ) (同) (三) (三) (三) (○) (○)

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.)

▲□▶ ▲□▶ ▲□▶ ▲□▶ ▲□ ● のへぐ

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.

(ロ) (同) (三) (三) (三) (○) (○)

Bootstrap sampling of confidence intervals.

Given *N* observed graphs G_1, \ldots, G_N from an ERGM with unknown parameters.

- 1. For j = 1, ..., q
 - 1.1 sample uniformly, with replacement, *N* indices $i_1^{(j)}, \ldots, i_N^{(j)}$ from 1, ..., *N*;

1.2 compute
$$\theta^{(j)} = \mathsf{MPLE}(G_{i_1^{(j)}}, \dots, G_{i_N^{(j)}});$$

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.

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三 のへで

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.

(ロ) (同) (三) (三) (三) (○) (○)

Hammersley-Clifford Theorem is no longer valid.