Network Models

Viviana Amati Jürgen Lerner Bobo Nick

University of Konstanz, Winter 2011/2012
(Version October 19, 2011)

Contents

Chapter 1

Introduction

Social network analysis, i. e., the joint analysis of actors and relations among them, rapidly gains importance in many scientific and commercial applications. Examples range from studies of organizational and communication networks over to the analysis of Web-based user interaction. Statistical approaches in social network analysis are applied to model, estimate, and predict social interaction and behavior based on empirical data. In this course you will learn mathematical and methodological foundations for modeling social networks.

Statistical network modeling can yield mathematically precise statements about *uncertainty* in social network data. Even under the assumption that a social network has been collected or measured without any error, it is plausible that the data will be different when we collect it again—at a different point in time, on a different set of actors, under different environmental factors, and so on. Thus, we cannot know for sure how the network looks like when we collect the data again (in reality or hypothetically); instead we can specify a probability distribution encoding the information that some network outcomes are just more likely than others. One or several network observations can influence the probability distribution of other networks collected at (real or hypothetical) repetitions of the experiment. A probability distribution for social networks can serve several purposes: to predict future network data, to learn rules that govern social processes, or to simply generate random networks that look like real data. In this lecture we treat several (families of) probability distributions that are more or less realistic as models

for social networks. A variable that has an outstanding effect on the nature of the model is time. This lecture is split in two major parts where in the first part we model networks without any time information and in the second part we model the evolution of networks over time.

Chapter 2

Time-independent Networks

In this chapter, we present and analyze *random graph models* that are increasingly realistic as models for empirical networks. After an introductory section in which some concepts from probability theory are revisited, we present several models of increasing complexity. Statistical properties, such as the expected values of certain graph characteristics, are derived and we present algorithms to (efficiently) sample graphs from these models.

2.1 Preliminaries

In this section we recall some basic definitions from probability theory that are needed later in this chapter.

2.1 Definition (probability space)
A (finite) probability space *is a pair* (Ω, P)*, where* Ω *is a finite set and* $P\colon \Omega \to [0,1]$ *a function satisfying* $\sum_{x\in\Omega} P(x) = 1$*. The real value* $P(x)$ *is called the* probability *of* x*. The probability of a subset* $X \subseteq \Omega$ *is denoted by* $P(X)$ *and defined by* $P(X) = \sum_{x\in X} P(x)$*.*

If φ is a Boolean condition defined on Ω, we often write $P(\varphi(x))$ instead of $P(\{x \in \Omega : \varphi(x)\})$. For instance, if $\Omega = \mathbb{N}$ we write $P(x < 100)$ for $P(\{x \in \mathbb{N} : x < 100\})$.

2.2 Definition (random variable)
Let (Ω, P) *be a probability space. A (real-valued)* random variable *is a*

function $X\colon \Omega \to \mathbb{R}$. *The probability that a random variable* X *takes a real number* $x \in \mathbb{R}$ *is denoted by* $P(X = x)$ *and defined by*

$$P(X = x) = P(X^{-1}(x)) = P(\{\omega \in \Omega \,:\, X(\omega) = x\}) \ .$$

The function $f_X\colon \mathbb{R} \to [0,1];\ x \mapsto P(X = x)$ *is called the* probability mass function *of* X.

The expectation *of a random variable* $X\colon \Omega \to \mathbb{R}$ *is denoted by* $\mathbb{E}(X)$ *and is defined by*

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

where $X(\Omega) = \{X(\omega) \,:\, \omega \in \Omega\}$ *is the set of values of* X.

One of the recurrent questions in this chapter is if and how an observed event (such as the presence of an edge in a network) influences the probability of another event (such as the presence of another edge in the same network). To make such statements precise, we need the following definitions.

2.3 Definition (independence and conditional probability)
Let (Ω, P) *be a probability space and let* $A, A', B \subseteq \Omega$ *be three subsets.*

- A *and* A' *are called* independent *if* $P(A \cap A') = P(A) \cdot P(A')$.
- *If* $P(B) > 0$, *then the* conditional probability of A, given B *is denoted by* $P(A|B)$ *and defined by*

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

- *If* $P(B) > 0$, *then* A *and* A' *are called* conditionally independent, given B *if*

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

2.2 Random Graph Models

A random graph model is a probability space on a set of graphs.

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

2.5 Remark (basic notation)
In this lecture we consider only random graph models $(\mathcal{G}, P)$, where all graphs in $\mathcal{G}$ have the same set of vertices V; usually $V = \{1, \dots, n\}$ for a positive integer n.

The set of dyads *is denoted by D and consists of all elements that can be edges in the respective graph model. For instance, in the case of undirected, simple, loopless graphs with vertex set V it is*

$$D = \{\{u, v\} \,:\, u, v \in V,\ u \neq v\}$$

and in the case of directed, simple, loopless graphs it is

$$D = V \times V \setminus \{(v, v) \,:\, v \in V\} \ .$$

In the case of graphs with loops, the "diagonal elements" (v, v) respectively $\{v, v\} = \{v\}$ are included in the set of dyads. The cardinality of D is usually denoted by M in this lecture.

If no misunderstanding can arise, we denote a dyad connecting the vertices i and j by ij, both for undirected and directed graphs. In the case of undirected graphs ij and ji denote the same dyad and we typically write ij (rather than ji) if $i < j$. In the case of directed graphs ij and ji are different.

2.6 Definition (random variable of a random graph model)
A random graph model $(\mathcal{G}, P)$ is associated with a family of random variables

$$Y = (Y_{ij} \colon \mathcal{G} \to \mathbb{R})_{ij \in D} \ ,$$

defined in the following. For a fixed graph $G \in \mathcal{G}$ the value of the random variable $Y_{ij}(G)$ describes the state of the dyad ij in the graph G. In the case of simple graphs this value is equal to one if ij is an edge in G and zero else. In the case of multi-graphs the value $Y_{ij}(G)$ is a non-negative integer equal to the multiplicity of the edge connecting i and j in G.

It is easy to see that a graph G is uniquely described by the family of values $(y_{ij} = Y_{ij}(G))_{ij \in D}$. Therefore, the random graph model is uniquely described by the joint distribution $P(Y)$ of $Y = (Y_{ij})_{ij \in D}$. Here, the joint distribution

of the family of random variables $Y = (Y_{ij})_{ij \in D}$ is defined in the following way: for a family of real numbers $y = (y_{ij} \in \mathbb{R})_{ij \in D}$, the probability that Y takes the value y is defined by

$$P(Y = y) = P(\{G \in \mathcal{G} \,:\, \forall ij \in D \colon Y_{ij}(G) = y_{ij}\}) \ .$$

The values $(y_{ij} = Y_{ij}(G))_{ij \in D}$ can be thought of as the entries of the adjacency matrix associated with G.

The simplest possible random graph model assigns the same probability to every graph.

2.7 Definition (uniform random graph model)
Let $n \geq 1$ be a positive integer, let $\mathcal{G}$ be the set of undirected, simple, loopless graphs that have exactly n vertices, and let the mapping P be defined by

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

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

Note that $P(G)$ is indeed a real number from the interval $[0, 1]$ and that P is normalized since

$$\sum_{G \in \mathcal{G}} P(G) = |\mathcal{G}| \cdot \frac{1}{2^{\frac{n(n-1)}{2}}} = 1 \ .$$

The last equation holds since the number of graphs in $\mathcal{G}$ is equal to the number of subsets of the set of dyads D. It is $|D| = n(n-1)/2$ and thus $|\{E \subseteq D\}| = 2^{\frac{n(n-1)}{2}}$.

For two integers i, j with $1 \leq i < j \leq n$ we associate the dyad ij with a subset of graphs

$$\mathcal{G}_{ij} = \{G = (V, E) \in \mathcal{G} \,:\, \{i, j\} \in E\} \ ,$$

i. e., the set of all graphs that contain ij in their edge set. The *probability of the edge ij* is defined as the probability of $\mathcal{G}_{ij}$, i. e.,

$$P(\{i, j\} \in E) = P(\mathcal{G}_{ij}) = \sum_{G \in \mathcal{G}_{ij}} P(G) \ .$$

Likewise, two edges $\{i,j\}$ and $\{u,v\}$ are said to be *independent* if $\mathcal{G}_{ij}$ and $\mathcal{G}_{uv}$ are independent, i. e., if

$$P(\mathcal{G}_{ij} \cap \mathcal{G}_{uv}) = P(\mathcal{G}_{ij}) \cdot P(\mathcal{G}_{uv}) \ .$$

The notion of independence of two sets of dyads is made precise in the following definition.

2.8 Definition (independence of sets of dyads)
Let $(\mathcal{G}, P)$ be a random graph model with associated set of dyads D and let $D_1 \subset D$ and $D_2 \subset D$ be two disjoint, non-empty sets of dyads. Then, D_1 is said to be independent *of D_2 if for all partitions $D_1 = D_1^+ \cup D_1^-$ and $D_2 = D_2^+ \cup D_2^-$ the subset*

$$\mathcal{G}_{D_1^+ \cup D_1^-} = \{G \in \mathcal{G} \ : \ D_1^+ \subseteq E_G \text{ and } D_1^- \cap E_G = \emptyset\}$$

is independent of

$$\mathcal{G}_{D_2^+ \cup D_2^-} = \{G \in \mathcal{G} \ : \ D_2^+ \subseteq E_G \text{ and } D_2^- \cap E_G = \emptyset\} \ .$$

Intuitively, the above definition requires that no matter how the dyads in D_1 are restricted to be edges or non-edges, the edge-probabilities for elements of D_2 are left unchanged.

With the above notation we can show that edges are mutually independent in the uniform random graph model $\mathcal{G}(n)$.

2.9 Lemma
Let n be a positive integer and let i, j be two integers with $1 \leq i < j \leq n$. Then, the edge $\{i,j\}$ is contained in a graph drawn from $\mathcal{G}(n)$ with probability $1/2$, independent of any set of dyads.

Proof. The two sets

$$\{(V,E) \in \mathcal{G} \ : \ \{i,j\} \in E\} \quad \text{and} \quad \{(V,E) \in \mathcal{G} \ : \ \{i,j\} \notin E\}$$

have the same cardinality, are disjoint, and their union equals $\mathcal{G}$. Since every graph has the same probability in $\mathcal{G}(n)$, it follows

$$P(\{(V,E) \in \mathcal{G} \ : \ \{i,j\} \in E\}) = 1/2 \ .$$

The same reasoning applies if any two disjoint subsets $D^+, D^- \subseteq D \setminus \{i,j\}$ of dyads are fixed and we compute the conditional probability of $\{(V,E) \in \mathcal{G} \ : \ \{i,j\} \in E\}$, given $D^+ \subseteq E$ (i. e., all dyads in D^+ are edges) and given $D^- \cap E = \emptyset$ (i. e., no dyad in D^- is an edge). □

To get a feeling into the properties of graphs sampled from a random graph model—in particular, to assess whether a random graph model is realistic as a model for empirical networks—it is often insightful to consider the expectation of certain graph properties. In the following we denote the number of edges of a graph $G \in \mathcal{G}$ by $m(G)$.

2.10 Lemma
Let n be a positive integer. The expected number of edges of graphs in $\mathcal{G}(n)$ is

$$\mathbb{E}_{\mathcal{G}(n)}[m] = \frac{n(n-1)}{4}$$

and, thus, the expected density of graphs in $\mathcal{G}(n)$ equals $1/2$.

Proof. Trivially, the number of edges of a graph $G = (V, E)$ can be written as

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

where the characteristic function $\chi_e \colon \mathcal{G} \to \{0, 1\}$ is defined by (let E_G denote the edge set of the graph G)

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

From the linearity of the expectation it follows that

$$\mathbb{E}_{\mathcal{G}(n)}[m] = \sum_{e \in D} \mathbb{E}_{\mathcal{G}(n)}[\chi_e] = \sum_{e \in D} P(e \in E) \cdot 1 + P(e \notin E) \cdot 0 = \frac{1}{2} \frac{n(n-1)}{2}$$

□

The uniform random graph model $\mathcal{G}(n)$ is very unsatisfactory as a model for empirical networks. For instance, it is typically observed that the density of networks tends to zero when the number of vertices increases. (Many more arguments could be given that $\mathcal{G}(n)$ does not realistically model empirical networks.) A random graph model that allows to control the expected density is defined in the following section.

2.3 $\mathcal{G}(n,p)$

2.11 Definition ($\mathcal{G}(n,p)$)
Let n be a positive integer and p be a real number from the interval $[0,1]$. The random graph model $\mathcal{G}(n,p)$ on the set of undirected simple, loopless graphs with exactly n vertices is induced by defining that each dyad $\{u,v\} \in D$ is an edge with probability p, independently on any set of dyads that does not contain $\{u,v\}$.

2.12 Remark

1. *In this section, let $q = 1-p$ denote the probability of a non-edge. To exclude trivial cases, we assume that p is neither zero nor one.*

2. *The model $\mathcal{G}(n,p)$ is equivalently specified by defining the probability of a graph G as*
$$P(G) = p^{m(G)} q^{\frac{n(n-1)}{2} - m(G)} \quad .$$
This follows by multiplying the probabilities for edges and non-edges, respectively, over all dyads. (Multiplication is valid since edge-probabilities are by definition independent in this model; see Lemma 2.13 below.)

3. *The uniform random graph model from the previous section is identical with $\mathcal{G}(n, \frac{1}{2})$.*

4. *The expected density in the model $\mathcal{G}(n,p)$ is equal to p. Consequently, the expected degree of any vertex is $p \cdot (n-1)$ and the expected number of edges is $p \cdot \frac{n(n-1)}{2}$. This follows by the same techniques as for the $\mathcal{G}(n)$ model in Lemma 2.10.*

5. *Sequences of random graph models $\mathcal{G}(n,p_n)$ with growing number of vertices that are sparse with constant expected degree can be generated by choosing $p_n \in \Theta(1/n)$.*

2.13 Lemma
Let $n \geq 1$, $p \in [0,1]$, and $q = 1-p$. The probability of a graph G with $m(G)$ edges in $\mathcal{G}(n,p)$ is
$$P(G) = p^{m(G)} q^{\frac{n(n-1)}{2} - m(G)} \quad .$$

Proof. Let D be the set of dyads and E_G be the set of edges of G. It is

$$\{G\} = \bigcap_{d \in E_G} \mathcal{G}_d \cap \bigcap_{d \in D \setminus E_G} \overline{\mathcal{G}_d} \ .$$

Since every dyad is independent of every set of dyads, the probability of an intersection of subsets is equal to the product of the probabilities of the individual subsets. In particular,

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

□

2.3.1 Generating Graphs from $\mathcal{G}(n, p)$

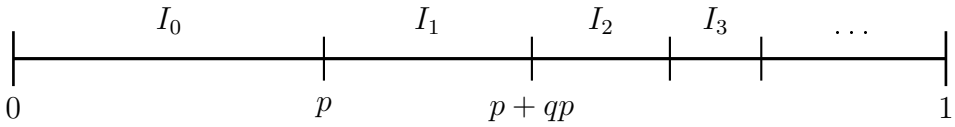
We now turn to the algorithmic question of generating graphs from $\mathcal{G}(n, p)$. This means that we want to design a probabilistic algorithm returning at each call a graph from $\mathcal{G}$ such that, for any graph $G \in \mathcal{G}$, the probability that G is returned by the algorithm is equal to the probability of G in $\mathcal{G}(n, p)$. We assume that probabilistic algorithms can rely on random number generators. Apart from the values returned by the random number generator, all other steps in the algorithm are deterministic.

A naïve algorithm to generate graphs from $\mathcal{G}(n, p)$ iterates over all dyads, generates for each dyad a (uniformly and independently distributed) random number $r \in [0, 1]$, and inserts the dyad in the edge set if $r \leq p$. The runtime of this algorithm is in $\Theta(n^2)$ independent of p. In particular, if p is decreasing with growing n and, thus, the generated graphs are sparse in expectation, then the runtime is asymptotically larger than the expected graph size.

To get an idea about how an efficient algorithm can be designed, we review the abovementioned naïve algorithm in an informal manner. Let p be a probability close to zero, e. g., $p = 1/1000$. The naïve algorithm iterates over all dyads $d_1, \ldots, d_M$ and asks for each dyad d_i the random number generator: *is d_i an edge?* getting most of the times the answer *no* (which can be interpreted that the effort was wasted). On average, the answer is *yes* only once in a thousand questions. Clearly, this is very inefficient.

It would be much more clever to ask the random number generator *what is the index of the next dyad that becomes an edge?* Then, only $\Theta(m)$ questions have to be asked if m is the number of edges of the generated graph. Important question: can this be done in a way that each graph has exactly the probability as in $\mathcal{G}(n,p)$?

A simple observation is that, at each step of the naïve algorithm, the next dyad that becomes an edge is preceeded by exactly k non-edges with probability $q^k p$. Thus we have to generate the number k of failed trials with probability $q^k p$ using only one call to the random number generator. To achieve this, we associate each $k = 0, 1, \ldots$ with a subinterval $I_k \subset [0,1]$ of length $q^k p$ such that these intervals are pairwise disjoint and their union equals $[0,1]$. This is achieved by defining I_k to be the interval ranging from $\sum_{i=0}^{k-1} pq^i$ to $\sum_{i=0}^{k} pq^i$, if $k > 0$ and $I_0 = [0,p]$ (also compare the following drawing).



The union of the I_k indeed equals $[0,1]$, since

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

For a random number $r \leftarrow \mathrm{random}([0,1])$, we need to determine the interval containing r. Note that the interval I_k ends at

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

Thus, for a random number $r \leftarrow \text{random}([0,1])$ we need to compute the minimum k such that I_k ends after r. The following inequalities are equivalent.

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

(For deriving the last inequality, note that $\log q < 0$. Furthermore, note that the probability $P(r = 1) = 0$ so that we can safely assume that r is not equal to one and, hence, $\log(1-r)$ is well-defined.)

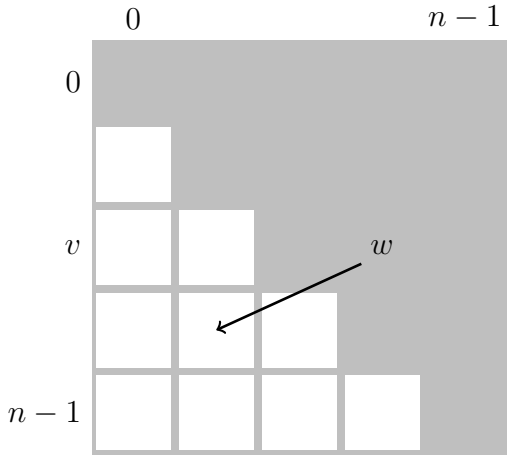
Thus we leave out $k = \left\lfloor \frac{\log(1-r)}{\log q} \right\rfloor$ dyads and insert the $k+1$ dyad in the edge set. Since we draw r from the uniform distribution on $[0,1]$ we can as well take $\log r$ instead of $\log(1-r)$.

Algorithm 1: Generating a graph from $\mathcal{G}(n,p)$

input : number of nodes $n \in \mathbb{N}_{\geq 1}$, edge-probability $0 < p < 1$
output: graph $G = (\{0,\ldots,n-1\}, E)$

```
E ← ∅
c ← 1/log(1−p)
v ← 1;   w ← −1
while v < n do
    r ← random([0,1])
    w ← w + 1 + ⌊c · log r⌋
    while w ≥ v and v < n do
        w ← w − v;  v ← v + 1
    if v < n then E ← E ∪ {{v, w}}
return G
```

The follwing image illustrates the use of the indices v and w. If w gets larger than or equal to v, then w is successively reduced by v and v is incremented by one in the second while loop.



2.14 Theorem
Let n be a positive integer, p from the open, real interval $(0,1)$, $m \in \{0,\ldots,\frac{n(n-1)}{2}\}$ and $G = (\{0,\ldots,n-1\}, E)$ an undirected, simple, loopless graph with exactly m edges. Then with probability

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

i. e., with the probability of G in $\mathcal{G}(n,p)$, Algorithm 1 returns G. In this case Algorithm 1 runs in $\mathcal{O}(n+m)$ time and uses exactly $m+1$ calls to the random number generator.

Proof. Without loss of generality, assume that the (undirected) dyads are represented by pairs of vertices (v,w) such that $v < w$. Using this representation, the dyads are enumerated (starting from one) in lexicographical ordering (first by v, then by w). This ordering of dyads induces also an ordering of the edges of the graph G. For $i = 1,\ldots,m$ let k_i be the number of consecutive non-edges just before the ith edge in G. (More precisely, k_i is the number of dyads between the $(i-1)$th and ith edge, if $i > 1$ and it is the number of dyads before the first edge, if $i = 1$.) Furthermore, let k_{m+1} be the number of non-edges after the last edge in G. Algorithm 1 generates G if and only if the following conditions are satisfied.

1. For all $i = 1,\ldots,m$, the random number generator returns in the ith iteration of the outer `while` loop a random number r such that $\left\lfloor \frac{\log r}{\log(q)} \right\rfloor = k_i$.

2. For $i = m+1$ the random number generator returns in the ith iteration of the outer `while` loop a random number r such that $\left\lfloor \frac{\log r}{\log(q)} \right\rfloor \geq k_i$.

The condition for $i = 1, \ldots, m$ is satisfied with probability pq^{k_i} (compare the derivation of the algorithm). The condition for $i = m+1$ is satisfied with probability $q^{k_{m+1}}$. To see this note that the probability of drawing a random number r such that $\left\lfloor \frac{\log r}{\log(q)} \right\rfloor \geq k_i$ is

$$\sum_{j=k_i}^{\infty} pq^j = \sum_{j=0}^{\infty} pq^j - \sum_{j=0}^{k_i-1} pq^j = 1 - (1 - q^{k_i}) = q^{k_i} \quad .$$

Since the random numbers are independent, it follows that the conditions are simultaneously satisfied for $i = 1, \ldots, m+1$ with probability

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

Thus, the probability of generating G is equal to the probability of G in $\mathcal{G}(n, p)$.

Furthermore, if Algorithm 1 generates G, then the outer `while` loop is executed exactly $m+1$ times, implying that exactly $m+1$ random numbers are generated. During the whole runtime of the algorithm, the second while loop is called n times (since each time v is increased by one). Together it follows that the runtime is in $\mathcal{O}(n+m)$. □

2.3.2 Statistical Inference of the Edge Probability

In this section we treat the following task: suppose that we are given a graph G of which we know that it is generated from some $\mathcal{G}(n, p)$ of which we do not know the parameter p. Can we infer p from G? To exclude trivial cases, assume that G is neither empty nor complete.

From a certain point of view it is impossible to infer p with certainty since G has a positive probability in $\mathcal{G}(n, p)$ for any p that lies between zero and one. Thus, we have to be more specific about what is meant by "inferring p." The specification of the "most likely" value for p that we adopt here

is derived from the so-called *maximum likelihood principle* (see, e. g., [9]). We determine $\hat{p}$ such that the probability of the given graph G in $\mathcal{G}(n, \hat{p})$ is maximized over all $p \in [0, 1]$. We give the following more general definition applying to random graph models that are parameterized by a real parameter vector θ.

2.15 Definition (maximum likelihood)
Let $(\mathcal{G}, P_\theta)$ be a family of random graph models parameterized by a k-dimensional vector of parameters $\theta \in \Theta \subseteq \mathbb{R}^k$ and let $G_{\text{obs}} \in \mathcal{G}$ be a particular graph, called observation. *The function*

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

mapping a parameter vector θ to the probability of G_{obs} in the model defined by P_θ is called likelihood. *A parameter vector $\hat{\theta}$ maximizing L, i. e.,*

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

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

In the case of the $\mathcal{G}(n, p)$ model we have a one-dimensional parameter vector, denoted by $p \in [0, 1]$. Given an observed graph G_{obs} (which is assumed to be neither empty nor complete) it is indeed quite simple to find the maximum likelihood estimate for p. Assume that G_{obs} has exactly m edges. Then the associated likelihood is

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

where M is the number of dyads, i. e., $M = \frac{n(n-1)}{2}$.
Taking the first derivative of L with respect to p yields

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

Searching for p in the open interval $(0, 1)$ such that $L'(p) = 0$ yields

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

so that $L(p)$ can take its only extremal point within the open interval $(0,1)$ at $\hat{p} := \frac{m}{M}$, i. e., at the ratio of the observed number of edges over the number of dyads. Since $L(0) = L(1) = 0$, $L(p) > 0$ for $p \in (0,1)$, $[0,1]$ is a compact set, and L is continuous, we conclude that $L(p)$ assumes a maximum at $\hat{p}$.

Maximum likelihood estimation becomes more insightfull (but also more difficult) for the exponential random graph models, later introduced in this chapter. There, parameter estimates serve to test social science hypotheses (such as homophily or transitivity) on given empirical network data.

2.3.3 Distribution of Degrees in $\mathcal{G}(n,p)$

We noted already that the expected degree $\mathbb{E}_{\mathcal{G}(n,p)}[d(v)]$ of any vertex v is equal to $p(n-1)$. Additionally, it is interesting to look at the distribution of degrees, i. e., how likely is it that the degree of a particular vertex equals a given number (which might be different from the expected degree).

2.16 Theorem

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

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

2. *Let λ be a positive real number, let for $n \geq \lambda + 1$ a sequence of edge probabilities $p_n \in [0,1]$ be defined by*

$$p_n = \frac{\lambda}{n-1}$$

and let $k \in \mathbb{N}_0$.

Further, let $P_n(d(v) = k)$ denote the probability that a given vertex v has degree equal to k in a graph drawn from $\mathcal{G}(n, p_n)$. Then it is

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

Proof. For the first claim, note that there are exactly $\binom{n-1}{k}$ different neighborhoods of v that have cardinality k. Each of them has probability $p^k q^{n-1-k}$.

For the second claim define $q_n = 1-p_n$ and note that it is sufficient to consider the probability $P_n(d(v) = k)$ only when $n \geq k+1$. Then, $P_n(d(v) = k)$ equals

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

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

2.4 Preferential Attachment

The random graph model $\mathcal{G}(n,p)$ has not been designed to model empirical networks (rather it turned out to be very helpful for proofing the existence of certain combinatorial objects in graphs, see [1]). Nevertheless, researchers started to point out that it is indeed a poor model for real-world networks. For instance, Theorem 2.16 shows that if p_n decays linearly in n, then the degree distribution in graphs from $\mathcal{G}(n,p_n)$ approximates a Poisson-distribution when n tends to infinity. In particular, the probability that a graph from $\mathcal{G}(n,p_n)$ has any node with a degree much larger than the average degree is vanishing. Figure 2.1 shows the histogram of 10^7 random draws from a poison distribution $P(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$ with $\lambda = 10$. One can see that the probability mass is highly centered around the mean and larger values become rapidly very unlikely. The maximal observed value is 30 which has been drawn four times out of ten million. In contrast it has been observed that many empirical networks contain some nodes whose degrees are much higher than the average degree.

More specifically, it has been claimed that empirical degree distributions often resemble a power-law, i. e.,

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

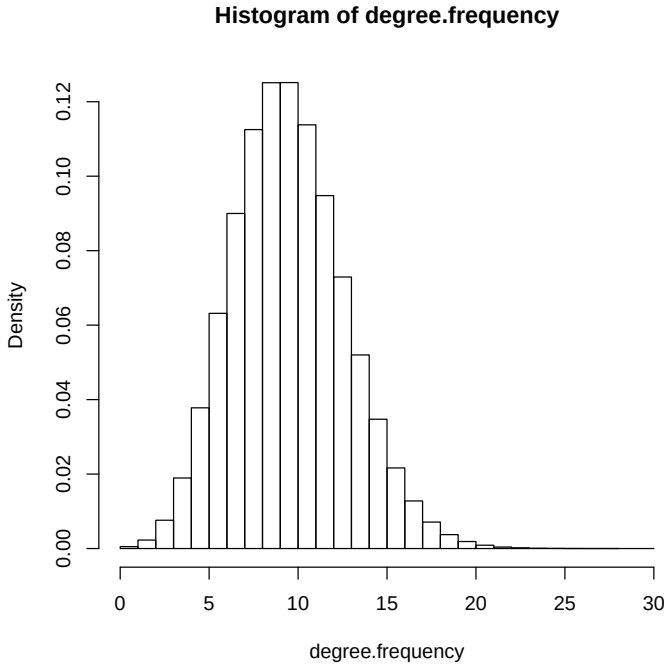


Figure 2.1: Histogram of 10^7 random draws from a poison distribution $P(k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!}$ with $\lambda = 10$.

for constants $c, \gamma > 0$. In such distributions, it is likely that a few nodes have degrees that are much higher than the average degree.

Therefore, efforts have been made to design random graph models that yield a power-law degree distribution. One idea for such a model comes from Barabási und Albert [2]. Roughly, they defined a generative model where vertices are successively added to the network, where the newly introduced vertices create a fixed number of edges to already existing vertices, and where the probability of forming an edge to a particular node v is proportional to the degree of v. Thus, a high-degree vertex is more popular with respect to incomming edges and the vertices entering the network have a preference towards creating edges to popular nodes. Barabási and Albert provide experimental evidence that such a process generates random graphs whose degree distribution is approximately a power-law with exponent $\gamma \approx 3$.

However, the description of the random graph model in Barabási and Albert was claimed to be "rather imprecise [4]" and a more precise model, implementing the ideas of the preferential attachment model, was proposed.

2.17 Definition (preferential attachment [4])
For $n \in \mathbb{N}_{\geq 1}$ a random graph model for directed multi-graphs with loops having exactly n vertices, denoted by $\mathcal{GP}_1^n = (\mathcal{G}, P_n)$, is recursively defined as follows. For a multi-graph $G = (\{1, \ldots, n+1\}, E)$ let G' denote the subgraph of G induced by the vertices $1, \ldots, n$. Then P_{n+1} is defined by

$$P_{n+1}(G) = \begin{cases} 0 & \text{if } d_G^+(n+1) \neq 1 \text{ oder } m(G) - m(G') \neq 1 \\ P_n(G') \cdot \frac{d_{G'}(v)}{2n+1} & \text{if } (n+1, v) \in E,\ v \leq n \\ P_n(G') \cdot \frac{1}{2n+1} & \text{if } (n+1, v) \in E,\ v = n+1\ . \end{cases}$$

For $n = 1$ the probability $P_1(G)$ of a graph G with exactly one vertex is set to one if G is the (unique) directed multi-graph with exactly one edge and $P_1(G)$ is set to zero if the number of edges in G is not equal to one.

Let additionally b be a positive integer. Then the random graph model $\mathcal{GP}_b^n$ is defined as follows. To generate a graph G with n vertices and nb edges from $\mathcal{GP}_b^n$ first a graph G' is drawn from $\mathcal{GP}_1^{nb}$; then repeatedly b successive vertices of G' are contracted to one vertex of G. More precisely, a vertex $i\ :\ (0 \leq i \leq n-1)$ of G is the contraction of vertices $ib, \ldots, ib+b-1$ of G'.

Note that all graphs that get a positive probability in $\mathcal{GP}_b^n$ have the same number of edges. Hence, only a finite number of graphs get a positive probability. Since graphs with probability equal to zero can be left out without loosing any interesting information, the $\mathcal{GP}_b^n$ model can still be treated as a finite probability space.

For the following algorithm to generate graphs from $\mathcal{GP}_b^n$ we use a random number generator that draws (uniformly and independently) integers from $\{0, \ldots, r\}$. The idea of the following algorithm is to record each edge by its two endpoints in an array A. The endpoints of the ith edge (where the counting is assumed to start at zero) are recorded in the $2i$th and $2i+1$th entry of A, respectively. Thus, at each stage of the algorithm the current degree of a vertex is equal to its number of occurrences in A. This last-mentioned property facilitates choosing the targets of edges with the correct

probability, i. e., proportional to their degree.

Algorithm 2: Generation of a graph from $\mathcal{GP}_b^n$

input : number of nodes $n \in \mathbb{N}_0$, out-degree $b \in \mathbb{N}_{\geq 1}$
data : array $A[0 \dots 2nb-1]$
output: multi-graph $G = (\{0, \dots, n-1\}, E)$

$E \leftarrow \emptyset$
$m \leftarrow 0$

foreach $v = 0, \dots, n-1$ **do**
  foreach $j = 0, \dots, b-1$ **do**
    $A[2m] \leftarrow v$
    $w \leftarrow A[\text{random}(\{0, \dots, 2m\})]$
    $A[2m+1] \leftarrow w$
    $E \leftarrow E \cup \{(v, w)\}$
    $m \leftarrow m + 1$

Algorithm 2 generates a directed multi-graph $G = (V, E)$ with n vertices and $m = nb$ edges with probability as specified in $\mathcal{GP}_b^n$. It uses $\mathcal{O}(n+m)$ time and space and draws exactly m random numbers.

The following theorem (which we note without proof) shows that the distribution of in-degrees follows approximately a power-law with exponent equal to three—at least up to degree equal to $n^{\frac{1}{15}}$.

2.18 Theorem ([4])
For $a, b \in \mathbb{N}$ let δ be defined by

$$\delta(a,b) = \frac{2b(b+1)}{(a+b)(a+b+1)(a+b+2)} \ .$$

For $n, a, b \in \mathbb{N}_{\geq n}$ with $0 \leq a \leq n^{\frac{1}{15}}$ and $\varepsilon \in \mathbb{R}_{>0}$ it holds in the model $\mathcal{GP}_b^n$ that

$$P\left((1-\varepsilon)\cdot\delta(a,b) \leq \frac{|\{v \in V_n : d_G^-(v) = a\}|}{n} \leq (1+\varepsilon)\cdot\delta(a,b)\right) \underset{n\to\infty}{\longrightarrow} 1$$

2.5 Exponential Random Graph Models

From a certain point of view, the preferential attachment model is a better model for social networks than the $\mathcal{G}(n,p)$ model. For instance, it can generate graphs whose expected degree distributions resemble those of empirical networks more closely. On the other hand, the preferential attachment model cannot reproduce other structural properties that are commonly observed in empirical data, such as transitivity or homophily. Rather than trying to augment network models by adding more and more structural features, we go the opposite way: we first define a class of random graph models that is very general and then restrict this generality in such a way that the model becomes tractable and can still reproduce important structural traits.

Exponential random graph models (ERGMs) are a class of random graph models whose underlying probability functions have a specific form. Informally, a concrete random graph model belonging to the ERGM class can be defined by specifying two components.

1. A set of network characteristics (called *statistics*) that determine the probability of a graph. Network statistics describe properties of the network such as density, occurrence of high-degree vertices, number of triangles, etc. An ERGM with a fixed set of statistics is still a model class that is parameterized by a real parameter vector.

2. Given a set of statistics, the parameters determine which of the network properties measured by these statistics increase or decrease (or have no influence on) the probability of a network. Parameters can be chosen to obtain a random graph model that incorporates certain structural features or parameters can be inferred from an observed empirical network.

In this section, let $\mathcal{G}$ denote the set of undirected, simple, loopless graphs with exactly n vertices.

2.19 Definition
The class of exponential random graph models (ERGM) *consists of random graph models that assign a graph $G \in \mathcal{G}$ a probability of the form*

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

where

- *the* $\theta_i \in \mathbb{R}$ $(i = 1, \ldots, k)$ *are real numbers, called* parameters*;*
- *the* $g_i \colon \mathcal{G} \to \mathbb{R}$ $(i = 1, \ldots, k)$ *are real-valued functions defined on graphs, called* statistics*;*
- *the* normalizing constant κ *is defined by*

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

Note that κ ensures that P is normalized, i. e., $\sum_{G \in \mathcal{G}} P_\theta(G) = 1$.

2.20 Remark
We remind that the above formula defines a model class rather than a random graph model. A concrete model is defined by specifying k, $(g_1, \ldots, g_k)$, *and* $(\theta_1, \ldots, \theta_k)$.

- *The set of statistics* $(g_1, \ldots, g_k)$ *(and, thus, also the value for* k*) is typically chosen by the researcher, e. g., motivated by social science theory or motivated by the hypotheses that are to be tested.*
- *The parameters are either chosen in order to specify a random graph model with the desired structural features, or they are inferred from an observed empirical network.*
- *Inference of parameters means that for a fixed choice of statistics* $(g_1, \ldots, g_k)$ *and a given empirical network (the* observation*) the parameters* $(\theta_1, \ldots, \theta_k)$ *are estimated by statistical methods such that the resulting model fits in a best-possible manner to the observation. What is meant by "best-possible" is usually specified by the maximum likelihood principle (see Def. 2.15).*
- *The fact whether estimated parameters are (significantly) positive or negative provides answers to empirical questions about structural trends in the network, such as reciprocity, transitivity, homophily, preferential attachment, etc.*

2.21 Example (network statistics)
Many commonly used statistics count the number of specific subgraphs (called configurations*) in the network. The following statistics are of this type. Let* $\mathcal{G}$ *denote the set of undirected, simple, loopless graphs on the vertex set* $V = \{1, \ldots, n\}$ *and let* $G \in \mathcal{G}$.

- *The set of edges of* G *is denoted by* $E(G)$. *The associated statistic* $m(G) = |E(G)|$, *thus, counts the number of edges in the network. If* $m(G)$ *is associated with a positive parameter, then graphs with more edges have higher probability. Conversely, if* $m(G)$ *is associated with a negative parameter, then graphs with less edges have higher probability. The statistic* $m(G)$ *is typically included in every reasonable ERGM (for an argument supporting this rule, see the Hammersley-Clifford Theorem 2.31).*

- *For* $k \in \{2, 3, \ldots, n-1\}$ *the set of* k-stars *of* G *is denoted by* $S_k(G)$ *and defined by*

$$S_k(G) = \{(u, \{v_1, \ldots, v_k\}) : \quad 1 \le u \le n, \\ 1 \le v_1 < v_2 < \cdots < v_k \le n, \\ \forall i = 1, \ldots, k : v_i \ne u, \text{ and} \\ \forall i = 1, \ldots, k : \{u, v_i\} \in E \}$$

  Note that if $2 \le k \le \ell \le n-1$, *then an* ℓ*-star contains* $\binom{\ell}{k}$ k*-stars. The associated* k*-star statistic* s_k *is denoted by* $s_k(G) = |S_k(G)|$.

  Informally, a positive parameter associated with a k*-star statistic implies a tendency to create edges to vertices that already have a high degree. Conversely, if a* k*-star statistic is associated with a negative parameter, then this implies a reluctance to create edges to high-degree vertices.*

- *The set of* triangles *of* G *is denoted by* $T(G)$ *and defined by*

$$T(G) = \{\{u, v, w\} : \quad 1 \le u < v < w \le n, \\ \{u, v\} \in E, \{v, w\} \in E, \{w, u\} \in E \} .$$

  The associated triangle statistic $t(G)$ *is defined by* $t(G) = |T(G)|$.

  Informally, a positive parameter associated with the triangle statistic implies a tendency to create edges between vertices that have common

neighbors. For instance, if the edges encode friendship, then a friend of a friend of a vertex v has a higher probability to become v's friend. Conversely, if the triangle statistic is associated with a negative parameter, then this implies a reluctance to create edges between vertices with common neighbors.

The following lemma clarifies that the $\mathcal{G}(n,p)$ model is a member of the ERGM class.

2.22 Lemma
If the edge probability p is neither zero nor one then $\mathcal{G}(n,p)$ is identical to the ERGM defined by

$$P_1(G) = \frac{\exp(\theta \cdot m(G))}{\kappa(\theta)}$$

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

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

$$\begin{aligned} P_2(G) &= p^{m(G)} \cdot (1-p)^{\binom{n}{2}-m(G)} \\ &= \left(\frac{p}{1-p}\right)^{m(G)} \cdot (1-p)^{\binom{n}{2}} \\ &= \exp[\theta \cdot m(G)] \cdot (1-p)^{\binom{n}{2}} \end{aligned}$$

Note that both the values $(1-p)^{\binom{n}{2}}$ and $\frac{1}{\kappa(\theta)}$ are independent of G (rather they are properties of the graph model). Thus, $\frac{P_1(G)}{P_1(G')} = \frac{P_2(G)}{P_2(G')}$ for any two graphs G, G'.

This implies that for arbitrary but fixed G' it is

$$\begin{aligned} \sum_{G \in \mathcal{G}} P_1(G) &= \sum_{G \in \mathcal{G}} P_2(G) \cdot \frac{P_1(G')}{P_2(G')} \\ 1 &= 1 \cdot \frac{P_1(G')}{P_2(G')} \end{aligned}$$

Hence, $P_1(G') = P_2(G')$. □

2.5.1 Generating Graphs from an ERGM

This section presents a probabilistic algorithm to generate graphs from a completely specified ERGM (i. e., an ERGM with fixed statistics and parameters). The generation of random graphs from ERGMs is much more complicated than the generation of graphs from the $\mathcal{G}(n,p)$ model (where edges are independent and, thus, can be created in an arbitrary order) or the generation of graphs from the preferential attachment model (which is actually defined by its generating algorithm).

The generation of graphs from a given ERGM $(\mathcal{G}, P)$ is done via a Markov chain simulation. Informally, the yet-to-be-defined Markov chain specifies transition probabilities to move from one graph in $\mathcal{G}$ to another in a such way that, when run infinitely long, the relative number of times that the Markov chain is on a given graph G equals the probability of G in $(\mathcal{G}, P)$.

Thus, to generate a graph from an ERGM one has to simulate the Markov chain for a large number of time-steps and then return its current state.

2.23 Definition (Markov chain)
Let $\mathcal{G} = \{G_1, \ldots, G_N\}$ be a set of graphs. A (stationary) Markov chain *on $\mathcal{G}$ is a sequence of random variables $(Y^{(t)})_{t\in\mathbb{N}}$, each taking values in $\mathcal{G}$, with the property that for all $t \in \mathbb{N}$ and all sequences of graphs $(G^{(0)}, \ldots, G^{(t-1)}, G', G) \in \mathcal{G}^{t+2}$ it is*

$$P(Y^{(t+1)} = G | Y^{(0)} = G^{(0)}, \ldots, Y^{(t)} = G') = P(Y^{(1)} = G | Y^{(0)} = G') \ .$$

In words, the conditional probability distribution of $Y^{(t+1)}$, given $Y^{(t)}$, is independent of t (the stationarity condition) and independent of the realizations of the random variables $Y^{(0)}, \ldots, Y^{(t-1)}$ (the Markov condition).

For two graphs $G_i, G_j \in \mathcal{Y}$, the transition probability of a Markov chain to move from G_i to G_j is denoted by

$$\pi_{ij} = P(Y^{(1)} = G_j | Y^{(0)} = G_i) \ .$$

2.24 Definition (stationary distribution)
Let π be the matrix of transition probabilities of a Markov chain on a set of graphs $\mathcal{G} = \{G_1 \ldots, G_N\}$. A probability distribution P on $\mathcal{G}$ is called stationary *if for all j it is $P(G_j) = \sum_{i=1}^{N} P(G_i)\pi_{ij}$.*

Note that P is stationary if and only if (with $P = [P(G_1) \dots P(G_N)] \in \mathbb{R}^N$ written as a row vector) it is

$$P = P\pi \ ,$$

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

2.25 Definition (irreducible, aperiodic)
A Markov chain on $\mathcal{G}$ with transition probabilities π is associated with a directed graph $(\mathcal{G}, E)$ whose vertices are the states and where

$$(G_i, G_j) \in E \Longleftrightarrow \pi_{ij} > 0 \ .$$

The Markov chain is called

- irreducible *if $(\mathcal{G}, E)$ is strongly connected;*
- aperiodic *if the greatest common divisor of all cycles in $(\mathcal{G}, E)$ equals one.*

The following theorem shows how the transition matrix can be defined such that the probability distribution on a set of graphs converges to a given distribution.

2.26 Theorem
Let π be the transition matrix of a Markov chain on $\mathcal{G}$. If a probability distribution P on $\mathcal{G}$ satisfies for all graphs $G_i, G_j \in \mathcal{G}$

$$P(G_i)\pi_{ij} = P(G_j)\pi_{ji}$$

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

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

Proof. Note that the matrix π satisfies the conditions of the Perron-Frobenius theorem. We show that

- the spectral radius ρ of π is one;

- P is an eigenvector of π with eigenvalue one.

Let $G_i \in \mathcal{G}$, then

$$\sum_{j=1}^{N} P(G_j)\pi_{ji} = \sum_{j=1}^{N} P(G_i)\pi_{ij} = P(G_i)$$

(where the latter equation holds since the rows of π sum up to one). Thus, P and π satisfy the matrix equation $P\pi = P$, i. e., P is an eigenvector of π with eigenvalue one.

Now, let x be an eigenvector of π with eigenvalue ρ (the spectral radius of π). For all j it is $\rho x_j = \sum_{i=1}^{N} x_i \pi_{ij}$. Thus

$$\begin{aligned} \rho \sum_{j=1}^{N} x_j &= \sum_{j=1}^{N} \rho x_j = \sum_{j=1}^{N} \sum_{i=1}^{N} x_i \pi_{ij} \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} x_i \pi_{ij} = \sum_{i=1}^{N} x_i \sum_{j=1}^{N} \pi_{ij} = \sum_{i=1}^{N} x_i \end{aligned}$$

Since all entries of x have the same sign it is in particular $\sum_{i=1}^{N} x_i \neq 0$. Thus, it must be $\rho = 1$. □

The transition probabilities are not uniquely determined by the condition of reversibility which leaves some freedom in specifying the Markov chain. One established specification of the transition probabilities π is called Gibbs sampling (compare [8]). The transition probability to jump from graph G_i to G_j is defined to be

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

$$\begin{aligned} \pi_{ij} &= \frac{P(G_j)}{\binom{n}{2}(P(G_i) + P(G_j))} \\ &= \frac{\exp\left(\sum_{\ell=1}^{k} \theta_\ell g_\ell(G_j)\right)}{\binom{n}{2} \exp\left(\sum_{\ell=1}^{k} \theta_\ell g_\ell(G_i)\right) + \exp\left(\sum_{\ell=1}^{k} \theta_\ell g_\ell(G_j)\right)} \quad . \end{aligned}$$

- $\pi_{ii} = \sum \frac{P(G_i)}{\binom{n}{2}(P(G_i)+P(G))}$ where the sum goes over all G that differ from G_i in exactly one dyad.

Note that the computation of the transition probabilities does not require the computation of the normalizing constant κ.

2.27 Theorem
The matrix of transition probabilities π defined by Gibbs sampling is normalized and leads to a Markov chain that is irreducible, aperiodic, and reversible.

Proof.

1. π is normalized since for all $i = 1, \ldots, N$ it is (let the summation index G range over all graphs that differ from G_i in exactly one dyad)

$$\begin{aligned} \sum_{j=1}^{N} \pi_{ij} &= \sum_{G} \frac{P(G_i)}{\binom{n}{2}(P(G_i)+P(G))} + \sum_{G} \frac{P(G)}{\binom{n}{2}(P(G_i)+P(G))} \\ &= \sum_{G} \frac{P(G)+P(G_i)}{\binom{n}{2}(P(G_i)+P(G))} = 1 \ . \end{aligned}$$

2. The so-defined Markov chain is irreducible since every graph G_i can be transformed into every other graph G_j by successively reverting all dyads in which G_i differs from G_j. Since all of these dyad changes have a positive probability, there is a path going from G_i to G_j in the graph associated with the Markov chain.

3. The Markov chain is aperiodic since its associated graph contains all loops and therefore the greatest common divisor of all cycles can only be one.

4. The Markov chain is reversible since for all $i, j = 1, \ldots, N$ it is

$$\frac{\pi_{ij}}{\pi_{ji}} = \frac{P(G_j)(P(G_i)+P(G_j))}{P(G_i)(P(G_i)+P(G_j))} = \frac{P(G_j)}{P(G_i)} \ .$$

□

2.5.2 Markov Random Graphs

The *Markov random graphs* [5] are a sub-class of the ERGM class. Markov random graphs are defined by specifying that the random variables associated with two dyads ij and uv are conditionally independent (given the realization of all other dyads) unless they have a vertex in common. Before we give a formal definition we first introduce the concept of a dependence graph.

2.28 Definition (dependence graph)
Let $Z = (Z_1, \ldots, Z_m)$ be a family of discrete random variables with a finite range and let $M = \{1, \ldots, m\}$ be the index set. The dependence graph *$\mathcal{D} = (M, E)$ of Z has vertex set M and edge set E consisting of all unordered pairs $\{i, j\}$ of indices such that Z_i and Z_j are conditionally dependent, given the rest of Z, i. e., given the values of all Z_ℓ for $\ell \neq i, j$.*

2.29 Remark (dependence graph of a random graph)
The dependence graph $\mathcal{D}$ of a random graph $(\mathcal{G}, P)$ has as vertices all dyads of $\mathcal{G}$ and two dyads $\{i, j\}$ and $\{u, v\}$ are connected by an edge in $\mathcal{D}$ if Y_{ij} and Y_{uv} are conditionally dependent, given all other dyads, i. e., all dyads Y_{rs} with $\{r, s\} \neq \{i, j\}$ and $\{r, s\} \neq \{u, v\}$.

2.30 Definition (Markov graph)
Let $\mathcal{G} = (\mathcal{G}, P)$ be a random graph, where the underlying graph class $\mathcal{G}$ has a common set of vertices $V = \{1, \ldots, n\}$. $\mathcal{G}$ is called a Markov random graph *if for all four pairwise different vertices $i, j, u, v \in V$, the unordered pair of dyads $\{\{i, j\}, \{u, v\}\}$ is not an edge in the associated dependence graph, i. e., Y_{ij} and Y_{uv} are conditionally independent, given the rest of the graph.*

From the definition of Markov random graphs, it is not obvious that they can be expressed as ERGMs. However, the next theorem states that this is true.

2.31 Theorem (Hammersley-Clifford [3, 6]; special case)
Let $(\mathcal{G}, P)$ be a random graph satisfying $P(G) > 0$ for all $G \in \mathcal{G}$ and let D denote the set of dyads of the graphs in $\mathcal{G}$. If $\mathcal{D}$ denotes the dependence graph of $(\mathcal{G}, P)$, then, there are constants $\{\alpha_A \in \mathbb{R} : A \subseteq D\}$, associated with subsets A of D, satisfying $\alpha_A = 0$ if A is not a clique in $\mathcal{D}$, such that

the probability $P(G)$ of each $G \in \mathcal{G}$ can be written as

$$P(G) = \frac{1}{\kappa} \exp\left(\sum_{A \subseteq E(G)} \alpha_A\right) , \qquad \text{where} \tag{2.1}$$

$$\kappa = \sum_{G' \in \mathcal{G}} \exp\left(\sum_{A \subseteq E(G')} \alpha_A\right) . \tag{2.2}$$

Conversely, if the probability P on $\mathcal{G}$ is defined by (2.1) and (2.2), then two dyads d_1 and d_2 are independent unless there is a subset $A \subseteq D$ with $d_1, d_2 \in A$ and $\alpha_A \neq 0$.

Note that a set consisting of only one dyad is always a clique in the dependence graph. For the proof of the Hammersley-Clifford Theorem we need the following theorem.

2.32 Theorem (Möbius inversion theorem [7])
Let S be a finite set and

$$f\colon \mathcal{P}(S) \to \mathbb{R}; \qquad g\colon \mathcal{P}(S) \to \mathbb{R};$$

two functions defined on the set of subsets of S.

Then, for all subsets $A \subseteq S$ it is

$$f(A) = \sum_{B \subseteq A} g(B)$$

if and only if for all subsets $A \subseteq S$ it is

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

We turn to the proof of the Hammersley-Clifford Theorem.

Proof. The following proof is adapted from [6].

For a set of dyads $B \subseteq D$ define $G_B \in \mathcal{G}$ to be the graph (on the same fixed set of vertices) whose edge set is equal to B.

We first note that two dyads $d, d' \in D$ that are not adjacent in the dependence graph $\mathcal{D}$ satisfy for any subset $A \subset D$ with $d \notin A$ and $d' \in A$ the equation

$$\frac{P(G_{A\cup\{d\}})}{P(G_A) + P(G_{A\cup\{d\}})} = \frac{P(G_{A\cup\{d\}\setminus\{d'\}})}{P(G_{A\setminus\{d'\}}) + P(G_{A\cup\{d\}\setminus\{d'\}})} \ . \tag{2.3}$$

(This is a direct formulation of d and d' being conditionally independent.) From (2.3) it can be derived that for two conditionally independent dyads $d, d' \in D$ and all subsets $B \subset D$ with $d, d' \notin B$ it is

$$\frac{P(G_{B\cup\{d,d'\}})}{P(G_{B\cup\{d'\}})} = \frac{P(G_{B\cup\{d\}})}{P(G_B)} \ . \tag{2.4}$$

Define $Q\colon \mathcal{G} \to \mathbb{R}$ by setting for $G \in \mathcal{G}$

$$Q(G) = \sum_{B\subseteq E(G)} (-1)^{|E(G)\setminus B|} \log P(G_B) \ . \tag{2.5}$$

Let $G \in \mathcal{G}$ and $A \subseteq E(G)$. Then, it follows from (2.5) that

$$Q(G_A) = \sum_{B\subseteq A} (-1)^{|A\setminus B|} \log P(G_B) \ . \tag{2.6}$$

For $A \subseteq E(G)$, we claim that $Q(G_A) = 0$ unless A is a clique in $\mathcal{D}$. To show this, assume that A is not a clique in $\mathcal{D}$, i. e., A contains two dyads $d, d' \in D$ which are conditionally independent, given all other dyads. Then, by (2.6) it is

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

where the last equality follows from (2.4).

From (2.6), which holds for all $A \subseteq E(G)$ we can derive from the Möbius inversion theorem that for all $A \subseteq E(G)$ it is

$$\log P(G_A) = \sum_{B \subseteq A} Q(G_B) \ . \tag{2.7}$$

In particular, if we set $A = E(G)$ in (2.7) we get

$$P(G) = P(G_\emptyset) \exp\left(\sum_{\substack{K \subseteq E(G) \\ K \in \mathrm{Cliques}(\mathcal{D})}} Q(G_K) \right) \ . \tag{2.8}$$

Here we used that by (2.6) it is $Q(G_\emptyset) = \log P(G_\emptyset)$. Note that a one-element set of dyads is always a clique in $\mathcal{D}$ but we consider $\emptyset \subseteq D$ not as a clique in $\mathcal{D}$. And we are done with the first implication.

Now suppose that a random graph model is defined by (2.1) and (2.2) and let d, d' be two dyads such that there is no subset $A \subseteq D$ with $\alpha_A \neq 0$ and $d, d' \in A$.

Let $B \subseteq D$ be any subset of dyads with $d, d' \notin B$. We show that

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

i. e., d and d' are conditionally independent. Equivalently,

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

We have that

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

(Note that we sum over all K that are cliques in $\mathcal{D}$.) □

For general Markov graphs the cliques of $\mathcal{D}$ correspond to sets of dyads A such that any pair of dyads within A have a vertex in common, i. e., are incident. These sets are singleton edges, triangles, and k-stars for $k = 2, \ldots, n-1$.

Let $D = \binom{V}{2}$ denote the set of dyads, let $T = T(V, D)$ denote the set of triangles of the complete graph on V, and, for $k = 2, \ldots, n-1$, let $S_k = S_k(V, D)$ denote the set of k-stars of the complete graph on V.

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

$$\begin{aligned} \eta_{uv} &\quad \textit{for all} \quad \{u, v\} \in D \\ \tau_{uvw} &\quad \textit{for all} \quad \{u, v, w\} \in T \\ \sigma_{uv_1 \ldots v_k} &\quad \textit{for all} \quad 2 \le k \le n-1,\ (u, \{v_1, \ldots, v_k\}) \in S_k \end{aligned}$$

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

$$P(G) = \frac{1}{\kappa} \exp\left(\sum_{uv \in E(G)} \eta_{uv} + \sum_{uvw \in T(G)} \tau_{uvw} + \sum_{k=2}^{n-1} \sum_{uv_1 \ldots v_k \in S_k(G)} \sigma_{uv_1 \ldots v_k} \right)$$

Note that the parameter families η_{uv}, τ_{uvw}, and $\sigma_{uv_1 \ldots v_k}$ are defined for all edges, triangles, and k-stars (respectively) of the complete graph; that these parameters are constants determined by the random graph model; and that the summation in the exponent of $P(G)$ goes over all edges, triangles, and k-stars (respectively) of G.

Markov graphs have way too many parameters to yield a parsimonious model. The reason is that in a general Markov graph each edge, triangle, and k-star can have a different probability. A restriction of Markov graphs is given in the following definition.

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

It is easy to see that for homogeneous Markov graphs all edge parameters are equal, all triangle parameters are equal, and, for each k all k-star parameters are equal. Thus the probability of a graph G in a homogeneous Markov model is determined by counting the number of these configurations.

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

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

ERGMs used in empirical research often contain only a few of the k-star statistics for the lower values of k (e. g., for $k = 2, 3$). The reason for this is that the n parameters of a homogeneous Markov graph might still be too many to be estimated with low standard error. Furthermore, the estimated k-star parameters are hard to interpret jointly—especially if some of them assume positive and some negative values.

2.5.3 Near-degeneracy and Multi-modality of ERGMs

Many Markov random graphs (i. e., for many choices of the parameters) give rise to multi-modal probability distributions:

- probability mass centered on a few small sets of graphs
- intermediate graphs are very unlikely.

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

Consider the following ERGM

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

Then, in very **sparse** networks

- there is no possibility to close triangles;

- creation of edges is very unlikely.

In contrast, in very **dense** networks

- an edge can close many triangles (up to $n-2$);
- deletion of edges is very unlikely.

Very unlikely to leave the set of near-complete graphs.

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.

Different proposals to overcome degeneracy have been made. Alternative definition of MC transition probabilities.

- For instance, allow switching to the complement graph.
- Leads to better convergence of the Markov chain.
- However, random graph model stays the same $\Rightarrow$ still unreasonable for empirical networks.
- Additionally, Markov chain simulation becomes unrealistic for empirical network evolution.

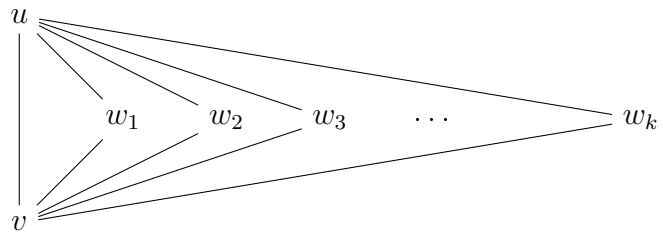
Condition on observed/reasonable number of edges.

- Allow only replacing one edge by another.
- Disjoint union of cliques becomes most likely form of network.

Assumed linear marginal effect of closed triangles:

- closing one triangle contributes τ to the log-probability;

- closing two triangles contributes 2τ; ...



Introduced k-triangle statistic:

- a $2k$-triangle counts more than a k-triangle,
- but less than twice as much.

$\Rightarrow$ no longer a Markov random graph.

Bibliography

[1] N. Alon, J. Spencer, and P. Erdős. *The Probabilistic Method*. Wiley, 1992.

[2] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[3] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society*, 36(2):192–236, 1974.

[4] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures and Algorithms*, 18:279–290, 2001.

[5] O. Frank and D. Strauss. Markov graphs. *Journal of the American Statisticsl Association*, 81(359):832–842, 1986.

[6] G. R. Grimmet. A theorem about random fields. *Bulletin of the London Mathematical Society*, 5(1):81–84, 1973.

[7] G.-C. Rota. On the foundations of combinatorial theory (I. Theory of möbius functions). *Wahrscheinlichkeitstheorie*, 2:340–368, 1964.

[8] T. A. B. Snijders. Markov chain Monte Carlo estimation in exponential random graph models. *Journal of Social Structure*, 2002.

[9] G. A. Young and R. L. Smith. *Essentials of Statistical Inference*. Cambridge University Press, 2005.