RANDOM GRAPHS AND SIMPLICIAL COMPLEXES

CATHERINE LI

Abstract. We define two models of random graphs, the latter of which we use throughout the paper. Then we examine threshold functions, which are proven using the first and second moment methods. We also define monotone graph properties and prove the existence of thresholds for these properties. We then move on to a series of specific graph properties and their thresholds: the existence of cycles and cliques, connectivity, and the emergence of the giant component. While exploring components, we use a branching process and approximate our model with probability distributions. Finally, we discuss the generalization to random simplicial complexes, which include higher dimensional objects beyond vertices and edges.

1. Introduction

The study of the field of random graphs began with Paul Erdős and Alfréd Rényi's 1959 paper *On Random Graphs* and their series of papers on the random graph in later years. In these texts, the authors defined the concept of random graphs and identified the phenomenon of thresholds, where the probability of having a certain graph property jumps from 0 to 1 below and above a certain value.

Definition 1.1. Erdős and Rényi's model denotes $G(n,m)$ as the probability space of graphs with n vertices and m edges, such that each of $\binom{\binom{n}{2}}{m}$ possible graphs has an equal probability, $\binom{\binom{n}{2}}{m}^{-1}$, of being chosen.

Definition 1.2. Another model, introduced by Edgar Gilbert, which we will use throughout this paper, denotes $G(n,p)$ as the probability space of graphs with n vertices such that each of the $\binom{n}{2}$ possible edges has probability p of being in a graph.

These two models can be related by the result that, if a graph in $G(n,p)$ has m edges, it is equally likely to be any graph in $G(n,m)$.

One way to study these graphs is the evolution of random graphs, which characterizes the stages of the graph $G(n,p)$ as we increase p : first, cycles arise, then the unique giant component emerges (accompanied by small components), followed by connectivity, and roughly equal degrees for all vertices.

2. Thresholds

As p increases, certain properties of the graphs often appear. For example, if edges are more likely to be in the graph, it is much more likely that the graph will be connected or contain a triangle. We define threshold functions, which describe when properties of a graph are likely or unlikely to appear as n increases to infinity.

Date: July 2022.

Definition 2.1. A function $r(n)$ is a *threshold function* for a property A of the graph if, as $n \to \infty$,

$$\Pr[G(n,p) \in A] \to \begin{cases} 0 & \text{if } p \ll r(n), \\ 1 & \text{if } p \gg r(n), \end{cases}$$

where $\Pr[G(n,p) \in A]$ denotes the probability that a graph in $G(n,p)$ satisfies A.

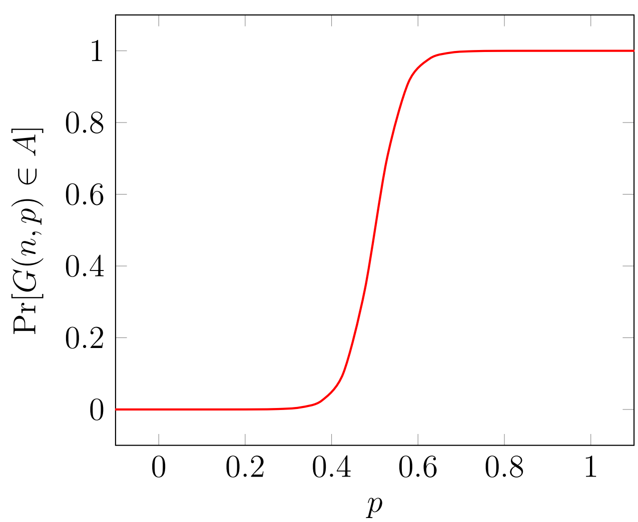
Note: We say that $p \ll r(n)$ or $p = o(r(n))$ if

$$\lim_{n\to\infty} \frac{p}{r(n)} = 0,$$

and $p \gg r(n)$ if

$$\lim_{n\to\infty} \frac{p}{r(n)} = \infty.$$

Interestingly, the interval of p during which the probability of having a certain property jumps from 0 to 1 can be quite small, creating the following relationship between p and $\Pr[G(n,p) \in A]$:



To prove the validity of thresholds, we use two methods: the first and second moment methods. Note that the kth moment of X is $\mathbb{E}\left[X^k\right]$.

When $p \ll r(n)$, we generally use the first moment method with Markov's Inequality, which involves $\mathbb{E}[X]$, the first moment of X.

Theorem 2.2 (Markov's Inequality)**.** *For nonnegative X and positive a,*

$$\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}.$$

Proof. We have

$$\begin{aligned}\mathbb{E}[X] &= \sum_i i\Pr[X=i] \\ &\geq \sum_{i\geq a} i\Pr[X=i] \\ &\geq \sum_{i\geq a} a\Pr[X=i] \\ &= a\Pr[X\geq a].\end{aligned}$$

Rearranging, we have

$$\Pr[X\geq a]\leq \frac{\mathbb{E}[X]}{a},$$

as desired. ■

To consider what happens when $p \gg r(n)$, we often use the second moment method with Chebyshev's Inequality, of which a special case involves $\mathbb{E}[X^2]$, the second moment of X. This represents the number of pairs of the structures counted by X.

Theorem 2.3 (Chebyshev's Inequality)**.** *Let μ and σ denote the mean (expected value) and standard deviation of a nonnegative variable X. For positive λ,*

$$\Pr[|X-\mu|\geq \lambda\sigma]\leq \frac{1}{\lambda^2}.$$

Proof. An equivalent statement is $\Pr[(|X-\mu|)^2\geq(\lambda\sigma)^2]\leq\frac{1}{\lambda^2}$. Applying Markov's inequality, we have

$$\begin{aligned}\Pr[(|X-\mu|)^2\geq(\lambda\sigma)^2] &\leq \frac{\mathbb{E}[|X-\mu|^2]}{\lambda^2\sigma^2} \\ &= \frac{\sigma^2}{\lambda^2\sigma^2} \\ &= \frac{1}{\lambda^2},\end{aligned}$$

and we are done. ■

Corollary 2.4. *For nonnegative X,*

$$\Pr[X=0]\leq\frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2},$$

where

$$\begin{aligned}\mathrm{Var}[X] &= \mathbb{E}[(X-\mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2-2X\mathbb{E}[X]+\mathbb{E}[X]^2] \\ &= \mathbb{E}[X^2]-2\mathbb{E}[X]^2+\mathbb{E}[X]^2 \\ &= \mathbb{E}[X^2]-\mathbb{E}[X]^2.\end{aligned}$$

By definition, $\mathrm{Var}[X]$ *is also equal to* σ^2.

Proof. Plugging in $\lambda = \mu/\sigma$ to Chebyshev's Inequality, we have

$$\begin{aligned}\Pr[X = 0] &\leq \Pr[|X - \mu| \geq \lambda\sigma = \mu] \\ &\leq \frac{1}{\lambda^2} \\ &= \frac{\sigma^2}{\mu^2} \\ &= \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2}.\end{aligned}$$

■

This corollary is the form of Chebyshev's Inequality that we will use in later proofs.

Now, we consider a specific type of graph property: monotone graph properties.

Definition 2.5. A graph property A is *monotone increasing* if, given a graph G that satisfies A, adding any edge to G maintains property A. In other words, adding edges cannot violate the property. One example of a monotone increasing graph property is connectivity: if a graph is already connected, adding edges cannot "unconnect" it.

Definition 2.6. Similarly, a *monotone decreasing* graph property A is one such that, given a graph G that satisfies A, removing any edge from G maintains the property.

We say that any property that is either monotone increasing or monotone decreasing is *monotone*.

Theorem 2.7. *Consider a monotone increasing graph property A and probabilities p_1, p_2. If $p_1 \leq p_2$, then*

$$\Pr[G(n, p_1) \in A] \leq \Pr[G(n, p_2) \in A].$$

Proof. Let p_0 be the nonnegative value such that

$$p_2 = p_1 + p_0(1 - p_1).$$

We take the graphs $G_0 \in G(n, p_0)$ and $G_1 \in G(n, p_1)$, randomly and independently chosen. Let

$$G_2 = G_0 \cup G_1.$$

Each edge in G_2 is put into the graph with probability

$$p_0 + p_1 - p_0 p_1 = p_2,$$

by the Principle of Inclusion-Exclusion, because the edge must be in at least one of G_0 and G_1. Thus

$$G_2 \in G(n, p_2).$$

Since $G_1 \in A$ implies $G_2 \in A$ (because G_2 is G_1 with a nonzero number of added edges), we have

$$\Pr[G(n, p_1) \in A] \leq \Pr[G(n, p_2) \in A],$$

as claimed. ■

Theorem 2.8. *Every monotone graph property has a threshold, given that for large enough n, the property is neither always satisfied nor never satisfied (such a property is called* non-trivial *by Béla Bollobás).*

Proof. Let A be the monotone graph property. Without loss of generality, we assume that A is monotone increasing. We define the function $p(\varepsilon)$ such that, for $0 \leq \varepsilon \leq 1$,

$$\Pr[G(n, p(\varepsilon)) \in A] = \varepsilon.$$

This is like an inverse function of $\Pr[G(n,p) \in A]$ that returns the probability $p(\varepsilon)$ we need to have probability ε of satisfying property A. We know that this value exists for all such ε because

$$\Pr[G(n,p) \in A] = \sum_{G \in A} p^{|E(G)|}(1-p)^{n-|E(G)|},$$

where $E(G)$ is the set of edges of G, since we sum the probabilities of getting each graph that satisfies A in $G(n,p)$.

Since the above probability is a polynomial in p, with value 0 when $p = 0$ and 1 when $p = 1$, all possible values of $\Pr[G(n,p) \in A]$ between 0 and 1 are realized, by the Intermediate Value Theorem.

Let p_0 represent $p\left(\frac{1}{2}\right)$. We claim that p_0 is a threshold for property A, meaning that as $n \to \infty$,

$$\Pr[G(n,p) \in A] \to \begin{cases} 0 & \text{if } p \ll p_0, \\ 1 & \text{if } p \gg p_0. \end{cases}$$

Consider k independent copies $G_1, G_2, \ldots, G_k$ of $G(n,p)$. Taking the union of these k copied graphs, there is $1-(1-p)^k$ probability that an edge is in at least one of the k copies. Thus, symbolically we write

$$G_1 \cup G_2 \cup \cdots \cup G_k \equiv G(n, 1-(1-p)^k).$$

By the binomial theorem,

$$1-(1-p)^k \leq kp,$$

so if any G_i satisfies property A, so does $G(n,kp)$. The contrapositive of this statement gives us that, if $G(n,kp)$ does not satisfy A, then no G_i does. Equivalently, by Theorem 2.7, we write

$$\Pr[G(n,kp) \notin A] \leq \left(\Pr[G(n,p) \notin A]\right)^k. \tag{2.1}$$

Now consider a function $\omega(n) \ll \log\log n$ such that as $n \to \infty$, $\omega \to \infty$ as well. As $n \to \infty$, we have $\frac{p_0}{\omega(n)} \to 0$ and $p_0\omega(n) \to \infty$, so

$$\frac{p_0}{\omega(n)} \ll p_0 \quad \text{and} \quad p_0 f(n) \gg p_0.$$

Plugging in $k = \omega$ and $p = \frac{p_0}{\omega}$ into inequality 2.1, we have, as $n \to \infty$,

$$\begin{aligned} \Pr\left[G\left(n, p_0\right) \notin A\right] &= \frac{1}{2} \\ &\leq \left(\Pr\left[G\left(n, \frac{p_0}{\omega}\right) \notin A\right]\right)^\omega. \end{aligned}$$

Rearranging, we get

$$\begin{aligned}\Pr\left[G\left(n, \frac{p_0}{\omega}\right) \notin A\right] &\geq \left(\frac{1}{2}\right)^{1/\omega} \\ &= 1 - o(1).\end{aligned}$$

This means $\Pr[G(n, p_0/\omega) \in A] = 1 - \Pr[G(n, p_0/\omega) \notin A] \to 0$, as claimed.

Similarly, plugging in $k = \omega$ and $p = p_0$ into inequality 2.1, we have, as $n \to \infty$,

$$\begin{aligned}\Pr[G(n, \omega p_0) \notin A] &\leq (\Pr[G(n, p_0) \notin A])^{\omega} \\ &= \left(\frac{1}{2}\right)^{\omega} \\ &= o(1).\end{aligned}$$

Thus $\Pr[G(n, \omega p_0) \in A] \to 1$, as desired. ■

3. Probability Distributions

Before we find the thresholds for specific graph properties, we introduce several probability distributions that can model the number of edges in random graphs.

Definition 3.1. A *binomial distribution* $\text{Bin}(n, p)$ is the discrete probability distribution over $0, 1, \ldots, n$ such that

$$\Pr[x \text{ successes}] = B(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}.$$

The binomial distribution is used to model the number of successes in n trials, where each trial has probability p of success and thus probability $1 - p$ of failure.

Definition 3.2. A *Poisson distribution* $\text{Pois}(\lambda)$ is the discrete probability distribution such that

$$\Pr[x \text{ occurrences}] = P(x; n, \lambda) = \frac{\lambda^x}{x!} e^{-\lambda},$$

where $\lambda = \mathbb{E}[x]$. This distribution models the number of times x that a certain event occurs over a time interval of length n, given that the event has a mean rate of λ.

We note that the Poisson distribution is the limiting case of the binomial distribution. Letting $\lambda = np$, we have

$$\begin{aligned}B(x; n, p) &= \frac{n!}{x!(n-x)!} \cdot p^x (1-p)^{n-x} \\ &= \frac{\lambda^x}{x!} \cdot \frac{n!}{(n-x)! \cdot n^x} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x}.\end{aligned}$$

Taking limits as $n \to \infty$, we have

$$\lim_{n\to\infty} \frac{n!}{(n-x)! \cdot n^x} = \lim_{n\to\infty} \frac{n(n-1)\ldots(n-x+1)}{n^x} = 1,$$

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda},$$
$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-x} = 1.$$

Therefore,

$$\lim_{n\to\infty} B(x;n,p) = \frac{\lambda^x}{x!}e^{-\lambda} = P(x;n,\lambda) = P(x;n-c,\lambda)$$

for any $c \ll n$.

4. Cycles

The first property we will look at is the existence of cycles; let C_k denote the cycle of k vertices that form a loop. We will see that the threshold for containing cycles (with any number of vertices ≥ 3) is $p = \frac{1}{n}$.

Theorem 4.1. *As $n \to \infty$,*

$$\Pr[G(n,p) \supseteq C_{k\geq 3}] \to \begin{cases} 0 & \text{if } p \ll \frac{1}{n}, \\ 1 & \text{if } p \gg \frac{1}{n}. \end{cases}$$

Proof. Let X represent the number of cycles in $G(n,p)$. We first find $\mathbb{E}[X]$. For each cycle size $k \geq 3$, we have $\binom{n}{k}$ ways to choose the k vertices and $\frac{k!}{2k} = \frac{(k-1)!}{2}$ ways to arrange them (rotating a cycle or changing its direction keeps it the same). There is also p^k probability that the k edges forming the loop are all present. Thus, by linearity of expectation,

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k\geq 3}\mathbb{E}[\text{cycles of size } k] \\ &= \sum_{k\geq 3}\binom{n}{k}\frac{k!}{2k}p^k \\ &= \sum_{k\geq 3}\frac{n(n-1)\dots(n-k+1)}{k!}\cdot\frac{k!}{2k}p^k \\ &\leq \sum_{k\geq 3} n^k p^k.\end{aligned}$$

By the geometric series formula,

$$\sum_{k\geq 3} n^k p^k = \frac{n^3p^3}{1-np}.$$

Next, Markov's Inequality gives us

$$\Pr[X \geq 1] \leq \mathbb{E}[X] \leq \frac{n^3p^3}{1-np}.$$

Since $p \ll \frac{1}{n}$, the numerator $n^3p^3 = o(1)$ and the denominator $1-np = 1-o(1)$. Therefore, the probability that we have at least one cycle when $p \ll \frac{1}{n}$ approaches 0 as $n \to \infty$.

Next, we will show that $\Pr[G(n,p) \supseteq C_{k\geq 3}] \to 1$ as $n \to \infty$. In particular, if this holds for $p = \frac{3}{n}$, then since having cycles is a monotone increasing property, this will hold for any $p \geq \frac{3}{n}$, and thus for all $p \gg \frac{1}{n}$.

Let $E(G)$ denote the set of edges in $G(n,p)$, and let S denote $|E(G)|$, the number of edges. It is well-known that a graph with n vertices and at least n edges must have at least one cycle (this can be shown by proof by contradiction and induction). Since each of $\binom{n}{2}$ edges has probability p of being in the graph, S follows the binomial distribution $\text{Bin}\left(\binom{n}{2}, p\right)$. This means

$$\mathbb{E}[S] = \binom{n}{2}p,$$
$$\text{Var}[S] = \binom{n}{2}p(1-p).$$

Letting $p = \frac{3}{n}$ and taking $n \to \infty$, we get

$$\mathbb{E}[S] = \frac{n(n-1)}{2} \cdot \frac{3}{n} \sim \frac{3n}{2}$$
$$\text{Var}[S] = \binom{n}{2} \cdot \frac{3}{n} \cdot \left(1 - \frac{3}{n}\right) \sim \frac{n^2}{2} \cdot \frac{3}{n} \sim \frac{3n}{2}.$$

Since we want to prove that with high probability $S \geq n$, it suffices to show that with high probability $|S - \mathbb{E}[S]| < \left|n - \frac{3n}{2}\right| = \frac{n}{2}$. Taking the complement, we plug in $\lambda\sigma = \frac{n}{2}$ to Chebyshev's Inequality, giving us

$$\begin{aligned}\Pr\left[|S - \mathbb{E}[S]| \geq \frac{n}{2}\right] &\leq \frac{\sigma^2}{\left(\frac{n}{2}\right)^2} \\ &= \frac{\text{Var}[S]}{\left(\frac{n}{2}\right)^2} \\ &\sim \frac{3n/2}{n^2/4} \\ &= \frac{6}{n},\end{aligned}$$

which tends to 0 when n tends to ∞. Therefore, $\Pr\left[|S - \mathbb{E}[S]| < \frac{n}{2}\right] \to 1$, so we have at least n edges and at least one cycle, with high probability.

■

5. Cliques

Another specific property of graphs that we will examine is containing cliques. Let K_r denote the r-clique, which is a graph on r vertices, all of which are connected to each other with a total of $\binom{r}{2}$ edges. This is also called the complete graph on r vertices.

A well-known result shows that the threshold for containing a triangle (3-clique) is $\frac{1}{n}$. We start with finding the threshold for containing K_4, then generalize to K_r for all $r \geq 3$.

Theorem 5.1. *Let K_4 be the complete graph on 4 vertices. As $n \to \infty$,*

$$\Pr[G(n,p) \supseteq K_4] \to \begin{cases} 0 & \text{if } p \ll n^{-\frac{2}{3}}, \\ 1 & \text{if } p \gg n^{-\frac{2}{3}}. \end{cases}$$

Proof. Let X be the number of K_4 subgraphs in $G(n,p)$. By linearity of expectation,

$$\mathbb{E}[X] = \binom{n}{4} p^6 = O(n^4)p^6,$$

since there are $\binom{n}{4}$ sets of 4 vertices and p^6 probability that we have all $\binom{4}{2} = 6$ edges.

If $p \ll n^{-\frac{2}{3}}$, we have

$$\mathbb{E}[X] = O(n^4)p^6 \ll 1.$$

By Markov's Inequality,

$$\Pr[G(n,p) \supseteq K_4] = \Pr[X \geq 1] \leq \mathbb{E}[X] = o(1),$$

so as $n \to \infty, \Pr[G(n,p) \supseteq K_4] \to 0$, meaning $G(n,p)$ very likely does not contain K_4.

Next, suppose $p \gg n^{-\frac{2}{3}}$. Chebyshev's Inequality tells us that

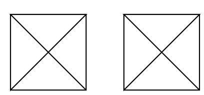
$$\Pr[X = 0] \leq \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2},$$

so it suffices to show that $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \ll \mathbb{E}[X]^2$.

We will find $\mathbb{E}[X^2]$, which represents the number of pairs of K_4 subgraphs in $G(n,p)$. We have 5 cases: the two 4-cliques can share 0, 1, 2, 3, or 4 vertices.

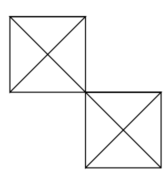
Case 1: 0 shared vertices

If two 4-cliques share no vertices, there are $\binom{n}{4}\binom{n-4}{4}$ ways to choose the eight vertices, and there are $\binom{4}{2} \cdot 2 = 12$ edges total, each of which appears with probability p. Thus this case contributes an expected $O(n^8)p^{12}$ pairs of 4-cliques.



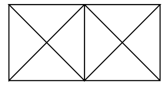
Case 2: 1 shared vertex

There are $\binom{n}{4}$ ways to choose the vertices of the first K_4, and $4 \cdot \binom{n-4}{3}$ ways to choose the shared vertex and 3 other vertices for the second K_4. We still have 12 total edges, so this case contributes an expected $O(n^7)p^{12}$ pairs.

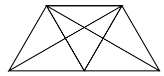


Case 3: 2 shared vertices

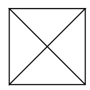
The two 4-cliques have 6 vertices and 11 edges total, so similarly to the previous cases, we have $O(n^6)p^{11}$ pairs in this case.



Case 4: 3 shared vertices
If the 4-cliques share 3 vertices (and 3 edges), they have 5 vertices and 9 edges total, so this case gives us $O(n^5)p^9$ pairs.



Case 5: 4 shared vertices
Finally, if the 4-cliques share all 4 of their vertices, they are the exact same graphs, and we have $O(n^4)p^6$ pairs of K_4 in this case.



Adding up the contributions from each case, we have

$$\begin{aligned}\mathbb{E}\left[X^2\right] &= O(n^8)p^{12} + O(n^7)p^{12} + O(n^6)p^{11} + O(n^5)p^9 + O(n^4)p^6,\\ \mathrm{Var}[X] &= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = O(n^7)p^{12} + O(n^6)p^{11} + O(n^5)p^9 + O(n^4)p^6\\ &= O(n^8p^12)\left(\frac{1}{n} + \frac{1}{n^2p} + \frac{1}{n^3p^3} + \frac{1}{n^4p^6}\right).\end{aligned}$$

Since $p \gg n^{-\frac{2}{3}}$, each term of $\left(\frac{1}{n} + \frac{1}{n^2p} + \frac{1}{n^3p^3} + \frac{1}{n^4p^6}\right)$ is $o(1)$, so $\mathrm{Var}[X] = \mathbb{E}[X]^2 o(1)$. Therefore, $\mathrm{Var}[X] \ll \mathbb{E}[X]^2$, so $G(n,p)$ contains K_4 with high probability, as desired. ∎

Now, we tackle the general case. The threshold for triangles is n^{-1}, and the threshold for K_4 is $n^{-\frac{2}{3}}$. We can reasonably guess that the threshold for containing K_r is $n^{-\frac{2}{r-1}}$, which is in fact correct.

Theorem 5.2. *Let K_r be the complete graph on r vertices, such that $r \geq 3$. As $n \to \infty$,*

$$\Pr[G(n,p) \supseteq K_r] \to \begin{cases} 0 & \text{if } p \ll n^{-\frac{2}{r-1}}, \\ 1 & \text{if } p \gg n^{-\frac{2}{r-1}}. \end{cases}$$

Proof. Again, let X be the number of subgraphs K_r in a graph in $G(n,p)$. Linearity of expectation gives us

$$\mathbb{E}[X] = \binom{n}{r} p^{\binom{r}{2}} = O(n^r)p^{\binom{r}{2}}.$$

If $p \ll n^{-\frac{2}{r-1}}$, we have

$$\mathbb{E}[X] = O(n^r)p^{\binom{r}{2}} \ll n^r n^{\frac{-2}{r-1} \cdot \frac{r(r-1)}{2}} = 1.$$

By Markov's Inequality,

$$\Pr[G(n,p) \supseteq K_r] = \Pr[X \geq 1] \leq \mathbb{E}[X] = o(1),$$

so as $n \to \infty, \Pr[G(n,p) \supseteq K_r] \to 0$.

Next, suppose $p \gg n^{-\frac{2}{r-1}}$. Since

$$\Pr[X=0] \leq \frac{\mathrm{Var}[X]}{\mathbb{E}[X]^2},$$

it suffices to show that $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \ll \mathbb{E}[X]^2$.

We find $\mathbb{E}[X^2]$, which represents the number of pairs of K_r subgraphs in $G(n,p)$. We have $r+1$ cases: the two r-cliques share $0, 1, 2, \ldots, r-1$, or r vertices.

Consider the case where the two r-cliques share k vertices. Between the two cliques, there are $2r-k$ vertices and $2\binom{r}{2}-\binom{k}{2}$ edges. Thus this case contributes $O(n^{2r-k})p^{2\binom{r}{2}-\binom{k}{2}}$ to $\mathbb{E}[X^2]$.

Summing for $k=0$ to r, we have

$$\begin{aligned}
\mathbb{E}[X^2] &= \sum_{k=0}^{r} O(n^{2r-k})p^{2\binom{r}{2}-\binom{k}{2}} \\
&= O(n^{2r})p^{2\binom{r}{2}} + \sum_{k=1}^{r} O(n^{2r-k})p^{2\binom{r}{2}-\binom{k}{2}}, \\
\mathrm{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= O(n^{2r})p^{2\binom{r}{2}} + \sum_{k=1}^{r} O(n^{2r-k})p^{2\binom{r}{2}-\binom{k}{2}} - \left(O(n^r)p^{\binom{r}{2}}\right)^2 \\
&= O(n^{2r})p^{2\binom{r}{2}} \left(\sum_{k=1}^{r} \frac{1}{n^k p^{\binom{k}{2}}} \right).
\end{aligned}$$

Since $p \gg n^{-\frac{2}{r-1}} \geq \frac{-2}{k-1}$ for each $1 \leq k \leq r$, each term

$$\frac{1}{n^k p^{\binom{k}{2}}} \ll \frac{1}{n^k n^{-\frac{2}{k-1}\cdot\binom{k}{2}}} = 1.$$

Thus

$$\mathrm{Var}[X] = O\left(n^{2r}p^{2\binom{r}{2}}\right) o(1) = \mathbb{E}[X]^2 o(1),$$

so $\mathrm{Var}[X] \ll \mathbb{E}[X]^2$, as needed.

■

6. Connectivity

We now examine another property: connectivity, or connectedness.

A graph is connected if and only if each vertex can be reached from every other vertex by traveling on a series of edges—the graph is a single connected component.

We define C to be the property of connectedness. The following theorem on connectivity is from Erdos and Renyi.

Theorem 6.1. *Consider a function $\omega(n) \geq 0$ such that $\omega(n) \leq \log\log n$ and as $n \to \infty$, the function $\omega(n)$ also tends to ∞. Then, as $n \to \infty$,*

$$\Pr[G(n,p) \in C] \to \begin{cases} 0 & \text{if } p = \frac{\log n - \omega(n)}{n}, \\ 1 & \text{if } p = \frac{\log n + \omega(n)}{n}. \end{cases}$$

In other words, $\frac{\log n}{n}$ is a threshold for connectivity.

Proof. First, let $p = \frac{\log n - \omega(n)}{n} = \frac{\log n}{n}(1 - o(1))$. We claim that $G(n,p)$ is not connected with high probability; it suffices to prove that there is at least one isolated vertex, because then the graph cannot be connected. Let X_1 be the number of isolated vertices (components of order 1, where order is the number of vertices). We have

$$\mathbb{E}[X_1] = n(1-p)^{n-1},$$

since each of the n vertices has $(1-p)^{n-1}$ probability of not being connected to any of the $n-1$ other vertices.

We note that $p \to 0$ for sufficiently large n, so

$$\begin{aligned} \mathbb{E}[X_1] &= n(1-p)^{n-1} \\ &\geq n(1-p)^n \\ &\geq ne^{(-p-p^2)n} \\ &= ne^{-\log n}e^{-\omega(n)}e^{-p^2 n} \\ &= e^{-\omega(n)}e^{-p^2 n} \to \infty, \end{aligned}$$

because $p^2 n = o(1)$. This means $\mathbb{E}[X_1] \to \infty$.

We will use the second moment method. With Chebyshev's Inequality, we have

$$\Pr[X_1 = 0] \leq \frac{\text{Var}[X_1]}{\mathbb{E}[X]^2},$$

so it suffices to show that $\text{Var}[X_1] = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 \ll \mathbb{E}[X_1]^2$.

To find $\mathbb{E}[X_1^2]$, we use $\mathbb{E}[X_1(X_1-1)]$, which represents the number of ordered pairs of isolated vertices in our graph. There are $n(n-1)$ ways to choose an ordered pair of vertices, and probability $(1-p)^{2(n-2)+1}$ that there are no edges between the isolated vertices themselves or between an isolated vertex and the rest of the graph. This gives us

$$\begin{aligned} \mathbb{E}[X_1(X_1-1)] &= n(n-1)(1-p)^{2(n-2)+1} \\ &\leq \frac{n^2(1-p)^{2n-2}}{1-p} \\ &= \frac{\mathbb{E}[X_1]^2}{1-p} \\ &\leq \mathbb{E}[X_1]^2 + 1, \end{aligned}$$

since $p \sim \frac{\log n}{n} \to 0$ as $n \to \infty$.

Thus

$$\begin{aligned}\mathrm{Var}[X_1] &= \mathbb{E}\left[X_1^2\right] - \mathbb{E}[X_1]^2 \\ &= \mathbb{E}[X_1(X_1-1)] + \mathbb{E}[X_1] - \mathbb{E}[X_1]^2 \\ &\leq \mathbb{E}[X_1]^2 + 1 + \mathbb{E}[X_1] - \mathbb{E}[X_1]^2 \\ &= \mathbb{E}[X_1] + 1.\end{aligned}$$

Since $\mathbb{E}[X_1] \to \infty$, we have

$$\mathrm{Var}[X_1] \leq \mathbb{E}[X_1] + 1 \ll \mathbb{E}[X_1]^2,$$

as desired.

Next, we claim that if $p = \frac{\log n + \omega(n)}{n} = \frac{\log n}{n}(1 + o(1))$, $G(n,p)$ is connected with high probability.

Again, we can characterize $\mathbb{E}[X_1]$,

$$\begin{aligned}\mathbb{E}[X_1] &= n(1-p)^{n-1} \\ &\leq ne^{-pn} \\ &= ne^{-\log n}e^{-\omega(n)} \\ &= e^{-\omega(n)} \to 0,\end{aligned}$$

as $n \to \infty$. Thus $\mathbb{E}[X_1] = o(1)$.

We let X_k denote the number of connected components that have k vertices (order k). Then our graph is disconnected if at least one of $X_1, X_2, \ldots, X_{\lfloor n/2 \rfloor}$ is greater than 0.

We find $\mathbb{E}[X_k]$. Each connected component has a spanning tree, and there are k^{k-2} ways to label a tree on k vertices. Thus, with Stirling's formula about factorials,

$$\begin{aligned}\mathbb{E}[X_k] &\leq \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} \\ &\leq \left(\frac{ne}{k}\right)^k k^{k-2} p^{k-1} (1-p)^{k(n-k)} \\ &\leq n^k e^k p^{k-1} e^{-pk(n-k)},\end{aligned}$$

where the last step follows from $1 - p \leq e^{-p}$.

It suffices to show that $\Pr[G(n,p) \notin C] \to 0$ as $n \to \infty$. We have

$$\begin{aligned}\Pr[G(n,p) \notin C] &= \bigcup_{k=1}^{\lfloor n/2 \rfloor} \Pr[X_k \geq 1] \\ &\leq \sum_{k=1}^{\lfloor n/2 \rfloor} \Pr[X_k \geq 1] \\ &\leq \sum_{k=1}^{\lfloor n/2 \rfloor} \mathbb{E}[X_k],\end{aligned}$$

by Markov's Inequality. Since $\mathbb{E}[X_1] = o(1)$, we only need to show that $\sum_{k=2}^{\lfloor n/2 \rfloor} \mathbb{E}[X_k] = o(1)$.

We split the sum into two parts: $k \leq \log\log n$ and $k > \log\log n$.

First we consider $k \leq \log\log n$. Since $\frac{\log n}{n} < p < \frac{2\log n}{n}$,

$$\begin{aligned}\mathbb{E}[X_k] &\leq n^k e^k p^{k-1} e^{-pk(n-k)} \\ &\leq n^k e^k \left(\frac{2\log n}{n}\right)^{k-1} e^{-k\log n} e^{pk^2}.\end{aligned}$$

We know $pk^2 \ll 1$ because $k \ll \log\log n$, so $e^{pk^2} < e$. Thus, we bound our expression

$$n^k e^k \left(\frac{2\log n}{n}\right)^{k-1} e^{-k\log n} e^{pk^2} \leq e^2 \left(\frac{2e\log n}{n}\right)^{k-1}.$$

Using the geometric series formula, we have

$$\sum_{k=2}^{\lfloor \log\log n \rfloor} \mathbb{E}[X_k] \leq e^2 \sum_{k=2}^{\infty} \left(\frac{2e\log n}{n}\right)^{k-1} = e^2 \frac{\frac{2e\log n}{n}}{1 - \frac{2e\log n}{n}} = o(1),$$

as desired.

For the second part of the summation, we take $\log\log n < k \leq \frac{n}{2}$. This means $n - k \geq \frac{n}{2}$. We have

$$\begin{aligned}\mathbb{E}[X_k] &\leq n^k e^k p^{k-1} e^{-pk(n-k)} \\ &\leq n^k e^k \left(\frac{2\log n}{n}\right)^{k-1} e^{-pkn/2} \\ &\leq ne(2e\log n)^{k-1} e^{-\frac{\log n}{n}\cdot\frac{kn}{2}} \\ &= ne(2e\log n)^{k-1} n^{-\frac{k}{2}} \\ &= n^{1/2} e \left(2en^{-1/2}\log n\right)^{k-1}.\end{aligned}$$

Summing over $\log\log n < k \leq \frac{n}{2}$ and using the geometric series formula, we get

$$\sum_{k=\lfloor \log\log n\rfloor+1}^{\lfloor n/2\rfloor} \mathbb{E}[X_k] \leq n^{1/2}e \sum_{k=\lfloor \log\log n\rfloor}^{\infty} \left(2en^{-1/2}\log n\right)^k$$

$$= n^{1/2}e \cdot \frac{\left(2en^{-1/2}\log n\right)^{\lfloor \log\log n\rfloor}}{1-2en^{-1/2}\log n}$$

.

Since $\log n \ll n^{1/2}$, the above expression is $o(1)$. Therefore, adding the two parts of the summation gives

$$\sum_{k=1}^{\lfloor n/2\rfloor} \mathbb{E}[X_k] = o(1),$$

so when $p = \frac{\log n+\omega(n)}{n}$, $G(n, p \in A)$ with high probability, and we are done.

■

7. Branching Processes and the Giant Component

An important and surprising discovery that Erdos and Renyi made about the components of a graph is that for small p, graphs generally have many small components, each with order $O(\log n)$. On the other hand, for larger p, a "giant component" emerges with order greater than $n^{2/3}$, along with small components of order $O(\log n)$.

We can examine the emergence of different components by a breadth-first traversal similar to the Galton-Watson branching process.

Now that we have defined some key probability distributions, we turn to the Galton-Watson branching process, which simply models the growth of a population with a tree. We start with a root node, expand to the root node's children, then to the children's children, and so on.

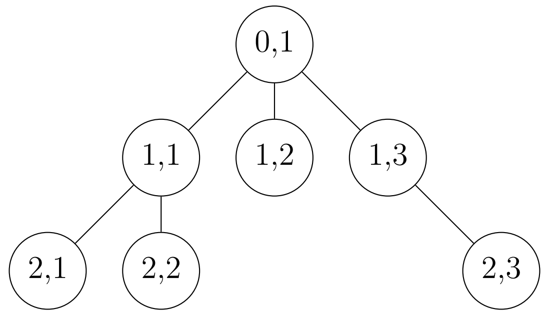
Let X be a random variable with some fixed probability distribution on the natural numbers $0, 1, 2 \dots$. We consider one organism (the 0th generation) that reproduces by dividing into a random number X of children, which make up the 1st generation. Each of the first-generation children then reproduces a random number of its own children, according to the distribution of X, to make the second generation, and so on.

We define Z_n to be the number of organisms in the nth generation, and let $\xi_{i,j}$ denote the number of children of the jth organism in the ith generation. By definition, $Z_0 = 1$. Also, we have the recursive formula

$$Z_{n+1} = \sum_{j=1}^{Z_n} \xi_{n,j},$$

because the number of organisms in the $(n+1)$th generation is the sum of the number of children of the organisms in the nth generation.

Consider the following example:



We have labeled a node (organism) "i, j" if it is the jth organism of the ith generation. Then in this example,

$$Z_0 = 1, Z_1 = 3, Z_2 = 3.$$

Also,

$$\begin{aligned} \xi_{0,1} &= 3, \\ \xi_{1,1} &= 2, \\ \xi_{1,3} &= 1, \\ \xi_{1,2} = \xi_{2,1} = \xi_{2,2} = \xi_{2,3} &= 0. \end{aligned}$$

Thus, we confirm our recursive formula to get

$$Z_2 = \sum_{j=1}^{Z_1} \xi_{1,j} = 2 + 0 + 1 = 3,$$

as expected.

Now that we are comfortable with the structure of our branching process, we consider the question of whether the population of organisms becomes extinct at some point, depending on the distribution of X.

Lemma 7.1. *Suppose $\mu = \mathbb{E}[\xi_{i,j}] < 1$. Then the population becomes extinct. Equivalently,*

$$\lim_{n\to\infty} Z_n = 0.$$

Proof. We know that

$$\begin{aligned} \mathbb{E}[Z_{n+1}] &= \mathbb{E}\left[\sum_{j=1}^{Z_n} \xi_{n,j}\right] \\ &= \sum_{j=1}^{Z_n} \mathbb{E}[\xi_{n,j}] \\ &= \sum_{j=1}^{Z_n} \mu \\ &= \mu Z_n. \end{aligned}$$

Since $Z_0 = 1$ and $Z_{n+1} = \mu Z_n$, we have

$$Z_n = \mu^n.$$

Because $\mu < 1$, we get

$$\lim_{n\to\infty} Z_n = \lim_{n\to\infty} \mu^n = 0,$$

as desired. ■

Next, if $\mu > 1$, we note that

$$\lim_{n\to\infty} Z_n = \lim_{n\to\infty} \mu^n \to \infty,$$

but this does not guarantee that the population never dies out, since there is still a probability that the organisms stop reproducing. Instead, we have a fixed nonzero probability that the organisms keep reproducing forever.

Lemma 7.2. *Suppose X, the random variable for an organism's number of children, follows the distribution $Pois(\lambda)$. We have $\lambda = \mathbb{E}[\xi_{i,j}] > 1$. Then the organism population becomes extinct with probability q, where q is the smallest nonnegative solution of $q = e^{\lambda(q-1)}$.*

Proof. We consider a generating function $f(t) = \mathbb{E}[t^X]$, where X is the random variable for the number of children each organism will have.

We are interested in the probability that the population dies out before the kth generation, where we start with generation 0.

For $k = 1$, this probability is

$$\Pr[X = 0] = \mathbb{E}\left[0^X\right] = f(0).$$

For $k = 2$, the answer is

$$(\Pr[X = 0])^X = f(f(0)).$$

Inductively, we have that the probability of the population dying out in k generations is

$$f(f(\dots f(0)\dots)) = f^k(0).$$

By the definition of q in the theorem statement,

$$\lim_{k\to\infty} f^k(0) = q.$$

We claim that q is the smallest fixed point of f; let this value be q_{fixed}. We have

$$\begin{aligned} 0 &\leq q_{\text{fixed}}, \\ f(0) &\leq f(q_{\text{fixed}}) = q_{\text{fixed}}, \\ &\vdots \\ f^k(0) &\leq f^k(q_{\text{fixed}}) = q_{\text{fixed}}. \end{aligned}$$

Thus $q = q_{\text{fixed}}$.

In particular,

$$q = \mathbb{E}[q^X] = \sum_{k=0}^{\infty} q^k \Pr[X = k].$$

Since X is a Poisson random variable with mean λ, this gives us

$$\begin{aligned} q &= \sum_{k=0}^{\infty} q^k \Pr[X = k] \\ &= \sum_{k=0}^{\infty} q^k \cdot \frac{\lambda^k}{k!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(q\lambda)^k}{k!}. \end{aligned}$$

The sum $\sum_{k=0}^{\infty} \frac{(q\lambda)^k}{k!}$ is the power series for $e^{q\lambda}$, so we have

$$q = e^{-\lambda + q\lambda} = e^{q(1-\lambda)}.$$

Therefore, the probability that the organisms go extinct eventually is the smallest nonnegative solution to $q = e^{\lambda(q-1)}$.

■

Now let's look at the application of this branching process in our random graph $G(n, p)$.

On our component, we conduct a breadth-first search similar to the branching process, starting from a root node u. First, suppose $p = \frac{\lambda}{n}$.

Let $N(u)$ denote the set of vertex u's neighbors in $G(n, p)$. Since each edge to the other $n - 1$ vertices appears with probability p, the number of "children" that u is connected to is

$$|N(u)| = \text{Bin}(n - 1, p) \to \text{Pois}(\lambda).$$

Consider one vertex v that is a child of u. We have $n - 1 - |N(u)|$ other vertices (not counting children of u) that can be v's children. Thus

$$|N(v)| = \text{Bin}(n - 1 - |N(u)|, p) \to \text{Pois}(\lambda)$$

as well.

Our approximations of $|N|$ as Poisson hold as long as we have only discovered a small number of vertices (negligible compared to n) through the branching process. The behavior of our branching process then allows us to determine whether our component is large or small, and also roughly bound the component order. Small components correspond to the organisms dying out in our branching process analogy, and the giant component corresponds to the organisms living "forever." We will find that the probability of having components of certain orders often decreases exponentially, meaning that smaller components have order at most $O(\log n)$.

Let C be a small component containing some vertex u. We will evaluate the behavior of $\Pr[|C| = k]$. There is a high probability that C is a tree, because we need to choose $k - 1$ other vertices in $O(n^{k-1})$ ways; in order for this component to exist with nonzero probability, we have at most $k - 1$ edges (with probability p^{k-1}).

For fixed k, there are $\binom{n-1}{k-1}$ ways to choose the other k vertices in the component, k^{k-2} to arrange the k vertices into a tree. Then, there is

$$p^{k-1}(1-p)^{k(n-k)+\binom{k}{2}-(k-1)}$$

probability of having the correct edges in the graph: $k-1$ edges in C, $k(n-k)$ non-edges between vertices in C and vertices not in C, and $\binom{k}{2}-(k-1)$ non-edges in C.

Therefore, if $n \gg k$,

$$\begin{aligned}\Pr[|C|=k] &= \binom{n-1}{k-1}k^{k-2}p^{k-1}(1-p)^{k(n-k)+\binom{k}{2}-k+1} \\ &\sim \frac{n^{k-1}}{(k-1)!}k^{k-2}\left(\frac{\lambda}{n}\right)^{k-1}\left(1-\frac{\lambda}{n}\right)^{kn} \\ &\sim e^{-\lambda k}\cdot\frac{\lambda^{k-1}k^{k-2}}{(k-1)!} \\ &= e^{-\lambda k}\cdot\frac{(\lambda k)^{k-1}}{k!}.\end{aligned}$$

Also, $k! = \left(\frac{k}{e}\right)^k \cdot P(k)$, where P is a polynomial in k. As $k \to \infty$, we consider only exponential terms, giving us

$$\begin{aligned}\Pr[|C|=k] &\sim e^{-\lambda k}\cdot\frac{(\lambda k)^{k-1}}{k^k/e^k} \\ &\sim \left(\lambda e^{-\lambda+1}\right)^k.\end{aligned}$$

The graph of $\lambda e^{-\lambda+1}$ shows that the function is < 1 unless $\lambda = 1$, in which case this evaluates to 1. Thus, if $\lambda \neq 1$, the probability of having component of order k decreases exponentially, so all small components have order at most $O(\log n)$, as desired.

Next, if $\lambda > 1$, the number of vertices in small components is at most cn for $c < 1$, so there must be a giant component with at least $O(n)$ vertices. We claim that this giant component, with εn vertices for $\varepsilon > 0$, is unique.

Let $p = \frac{\lambda}{n} = (1-p_1)(1-p_2)$, where $p_1 = n^{-3/2}$ and p_2 is roughly $p = \frac{\lambda}{n}$. Then

$$G(n,p) = G(n,p_1) \cup G(n,p_2).$$

Since $p_1 = n^{-3/2} \ll \frac{1}{n}$, with high probability $G(n,p_1)$ has no giant components. Also, $G(n,p_2)$ has a giant component, but we must show that $G(n,p)$ has at most 1 such component. Suppose that $G(n,p_2)$ has 2 giant components. The probability that, after adding in the edges from $G(n,p_1)$, the two components have not been merged is

$$\begin{aligned}(1-p_1)^{(\varepsilon n)^2} &\sim \left(1-\frac{n^{-1/2}}{n}\right)^{n\cdot\varepsilon^2 n} \\ &\sim e^{-n^{-\frac{1}{2}}\varepsilon^2 n} \\ &= e^{-\varepsilon^2\sqrt{n}} \to 0,\end{aligned}$$

as $n \to \infty$.

This probability decreases exponentially, and there are $\frac{1}{\varepsilon}$ components to merge, so with high probability they all merge in $G(n,p)$ into a single giant component, as claimed.

8. Simplicial Complexes

A generalization of the random graph to higher dimensions brings us to random simplicial complexes. By extending to higher dimensions, simplicial complexes allow us to model not only pairwise interactions through edges, but also network interactions involving multiple objects.

Definition 8.1. A *k-simplex* is the k-dimensional equivalent of an edge, triangle, or tetrahedron, which are the $1, 2$, and 3 dimensional simplices.

Definition 8.2. A *simplicial complex* is a structure described by a set of simplices (vertices, edges, triangles, tetrahedra, etc.). A *simplicial k-complex* is then a simplicial complex where the highest dimension of its simplices is k.

In 2006, Linial and Meshulam began the study of random simplicial complexes with their paper *Homological Connectivity of Random 2-Complexes*. They considered the complete "skeleton" graph on n vertices, then added triangles to the 2-simplicial complex, each with probability p. This leads to the following definition of a random 2-simplicial complex, analogous to that of a random graph.

Definition 8.3. A *random 2-simplicial complex* $Y(n,p)$ is the probability space of 2-dimensional simplicial complexes with n vertices and $\binom{n}{2}$ edges connecting every pair of vertices, such that each of the $\binom{n}{3}$ possible triangles has p probability of being in the complex.

Linial and Meshulam discovered a threshold for the homological connectivity of 2-simplicial complexes analogous to the connectivity of random graphs in Theorem 6.1. In fact, topologically, graph connectivity is homological 0-connectivity. The following theorem appears as stated in their paper.

Theorem 8.4. *Consider a function $\omega(n) \geq 0$ such that $\omega(n)$ tends to ∞ as n tends to ∞. Then, as $n \to \infty$,*

$$\Pr[H^1(Y(n,p), \mathbb{F}_2)] \to \begin{cases} 0 & \text{if } p = \frac{2\log n - \omega(n)}{n}, \\ 1 & \text{if } p = \frac{2\log n + \omega(n)}{n}, \end{cases}$$

so $\frac{2\log n}{n}$ is a threshold for $\mathbb{F}_2$-homological 1-connectivity.

The case of $p = \frac{2\log n - \omega(n)}{n}$ is similar to the corresponding proof for random graphs: we find the expected number of isolated edges, rather than vertices, then apply the second moment method. However, $p = \frac{2\log n + \omega(n)}{n}$ is more complex. The proof of this theorem and other topological thresholds for 2-complexes is an area for further study.

Acknowledgments. The author would like to thank Dr. Simon Rubinstein-Salzedo and Euler Circle for providing the opportunity to write this paper, as well as Nitya Mani for her support and expertise on the topic.

References

[AS90] Noga Alon and Joel Spencer. *The Probabilistic Method*. John Wiley & Sons, Inc., New York, 1990.
[Cse15] Timothy Csernica. Extinction in single and multi-type branching processes. 2015.
[ER60] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. pages 17–61, 1960.
[Gun13] Karen Gunderson. Connected components in erdős-rényi random graphs. 2013.
[Kah13] Matthew K. Kahle. Topology of random simplicial complexes: a survey. 2013.
[Kah16] Matthew K. Kahle. Random simplicial complexes. 2016.
[KF16] Michał Karoński and Alan M. Frieze. *Introduction to Random Graphs*. Cambridge University Press, 2016.
[LM06] Nathan Linial and Roy Meshulam. Homological connectivity of random 2-complexes. 2006.
[Nov13] Artem Novozhilov. Chapter 3: Erdős-rényi random graphs. 2013.
[Roc20] Sebastian Roch. Modern discrete probability: An essential toolkit. 2020.
[Ros14] Sheldon M. Ross. *Introduction to Probability Models*. Academic press, 2014.
[Spi10a] Daniel Spielman. Erdős-rényi random graphs: The giant component. 2010.
[Spi10b] Daniel Spielman. The giant component and diameter. 2010.
[Tes17] Mihai Tesliuc. Random graphs and their applications. 2017.