RANDOM GRAPHS AND THEIR APPLICATIONS

MIHAI TESLIUC

Abstract. We will explore central topics in the field of random graphs, beginning by applying the probabilistic method to prove the existence of certain graph properties, before introducing the Erdös-Rényi and Gilbert models of the random graph. Using these models we will cover theorems and applications regarding the existence of k-cliques, connectivity, and emergence of the giant component. We then conclude by introducing the "small-world" problem with an overview on how random graphs can be used to prove interesting results about this problem.

Contents

1. Introduction

Paul Erdös and Alfred Rényi introduced the concept of a random graph in 1959 as an extension of the probabilistic method to discover the existence of certain graph properties [3]. Random graphs have been used to gain insight on graph behavior and have been applied more broadly to solve combinatorial problems. The topic has been extended to model various random-like networks, such as the unpredictable growth of the Internet's web graph, the spread of human population and social relations, and neural networks. As an example we will show how random graphs are applied to show that the average degree of separation between humans is around six.

The term "random graph" was first mentioned in a paper by Erdös where it was used in a stunning proof of a theorem regarding the existence of certain graphs that demonstrated the power of the probabilistic method.

Theorem 1.1 (Erdös, 1959)**.** *There exist graphs whose girth and chromatic number are both arbitrarily large.*

Date: August 31, 2017.

As a reminder, girth refers to the length of the smallest cycle in the graph, while chromatic number is the least amount of colors necessary to color the vertices such that no two adjacent vertices share the same color. These two concepts are almost opposite to one another, where a graph with high chromatic number would then be very interconnected, and thus have low girth. Hence constructing a graph with both high girth and chromatic number tends to be a difficult problem. The strength of the proof lies in the strategy where instead of finding or describing a particular graph with the desired property, we instead try to read the existence of such a property by analyzing the probabilistic space of random graphs [2].

Proof. Let $k \in \mathbb{N}$ be arbitrary. We want to show there exists a graph G such that $\chi(G) \geq k$ and $g(G) \geq k$.

Proving $\chi(G) \geq k$ is often a hard problem, so instead we will count the independence number $\alpha(G)$, which is the size of the largest set of mutually non-adjacent vertices.

First, we declare a fairly simple relationship between independence number, chromatic number, and the number of vertices, denoted by $|G|$.

$$|G| \leq \chi(G)\alpha(G)$$

Using this inequality, if $\alpha(G) \leq \frac{|G|}{k}$ then it follows that $\chi(G) \geq k$.

Let $\epsilon > 0$ be sufficiently small. Let $p = n^{-(1-\epsilon)}$. Construct a random graph G on n vertices by adding edges independently with probability p.

Formally we are defining a probability space Ω and independent variables I_{vw} for each $vw \in \binom{[n]}{2}$ such that:

$$\mathbb{P}(I_{vw} = 1) = p$$

and defining G to be a graph with vertex set $[n]$ and edge set $\{vw \in \binom{[n]}{2} : I_{vw} = 1\}$.

We claim that with high probability (*w.h.p.*) G has at most $\frac{n}{2}$ cycles of length at most k, and contains no independent set of size $\frac{n}{2k}$. Therefore, if we remove a vertex of each cycle, we will have a graph on $\frac{n}{2}$ vertices with girth at least k, and with no independent set of size $\frac{n}{2k}$, and thus chromatic number at least k. Then we will have found our graph.

We will assign the random variable $X = X(G)$ to count the number of cycles of length l less than or equal to k and the random variable $Y = Y(G)$ to count the number of independent sets of size $\frac{n}{2k}$.

We recall that, if X is a random variable, Markov's inequality states that

$$\mathbb{E}(X) \geq a\mathbb{P}(X < a)$$

Using this we can also bound $\mathbb{E}(X)$, the number of l cycles, which we calculate by summing over u, the set of l vertices:

$$\begin{aligned} \mathbb{E}(X) &= \sum_u \mathbb{E}(\mathbf{1}_{\{u \text{ forms a cycle}\}}) \\ &= \binom{n}{l} p^l \leq n^l p^l \leq n^{\epsilon l} \ll n \end{aligned} \tag{1.2}$$

Similarly,

$$\mathbb{E}(Y) \leq 2^n p^{n^2}/8k \ll 1 \tag{1.3}$$

Therefore,

$$\frac{\mathbb{E}[|\{C_l \subset G : l \leq k\}|]}{n/2} + \mathbb{E}[|\{S : |S| \geq \frac{n}{2k}, e(X) = 0\}|] \longrightarrow 0 \tag{1.4}$$

as $n \longrightarrow \infty$. The left hand side of the sum in (1.4) tends to 0 as the numerator is significantly less than n. The right hand side also tends to 0 as the expected value of Y was significantly less than 1. Thus the entire sum tends to 0 as n increases, Therefore, by Markov, G has the desired properties w.h.p. and so we are done. □

The idea behind this previous proof lies in understanding the space with all possible graphs that are constructed in a random way. This idea naturally motivates the construction of random graphs.

2. The Erdös-Rényi and Gilbert Models

Now we present the formal definitions of random graphs, and the particular models that we will be using in this paper. Note that a random graph is not a graph in its own right, but rather a probability space with graphs as its elements.

Definition 2.1. An Erdös-Rényi random graph $\mathbb{G}(n, m)$ is a probability space of all unlabeled graphs with n vertices and m edges. Each graph in this space has equally likely probability of being chosen, that is

$$\mathbb{P}(G_{n,m} = G_0) = \binom{\binom{n}{2}}{m}^{-1}$$

The formula follows as we are choosing m edges out of $\binom{n}{2}$ possible edges, where each graph occurs with equal probability.

Definition 2.2. A Gilbert random graph $\mathbb{G}(n, p)$ is a probability space on the set of graphs with n vertices where each edge in the graph is added independently with probability p.

Note in the Erdös-Rényi model we fix the number of vertices and edges, while in the Gilbert model we only fix the number of vertices, and the number of edges will vary. The next theorem explains how these two models are interrelated.

Theorem 2.3. *The Erdös-Rényi random graph G_m is precisely a Gilbert graph given that the number of edges of $G_{n,p}$ is m.*

Proof. Fix a graph of n vertices and m edges as G_0. We begin with the probability that a particular Gilbert random graph is G_0, given that the number of edges is m.

$$\begin{aligned}
\mathbb{P}(G_{n,p} = G_0 | e(G) = m) &= \frac{\mathbb{P}(G_{n,p} = G_0)}{\mathbb{P}(e(G_{n,p} = m))} \\
&= \frac{p^m (1-p)^{\binom{n}{2}-m}}{\binom{\binom{n}{2}}{m} p^m (1-p)^{\binom{n}{2}-m}} \\
&= \binom{\binom{n}{2}}{m}^{-1}
\end{aligned}$$

□

Though both models are frequently used, the Gilbert model tends to be more prevalent in random graph theory due to its greater flexibility in choosing the value of p. One more useful concept is that of the graph process.

Definition 2.4. A graph process, denoted by $\tilde{G}$ is a method of constructing a graph by adding an edge to a graph G_0 by choosing one with equal likelihood out of all possible edges. G_t is the graph constructed after t of these steps.

3. Existence of K-cliques

An early theorem proved that when $p \gg \frac{1}{n}$, as n tends to infinity the random graph $G_{n,p}$ is almost certain to contain a triangle, while if $p \ll \frac{1}{n}$, the graph is almost certain not to contain a triangle.

This concept can be expanded to find the values of p such that w.h.p. the graph $G_{n,p}$ is almost certain to either contain or not contain a set of k vertices that are all connected to each other, denoted as a k-clique. This idea is motivated by the fact that as you populate a graph with more edges, you can expect to see more cliques pop up, while sparsely populated graphs will likely not contain large cliques. We are interested in the bounds for p such that these behaviors occur. Clearly if $p = 1$ then we can find any clique which has size less than or equal to n, but we are interested in the minimum value of p where we can still find these cliques.

These concepts can be extended further to find the Gilbert graphs such that for any possible subgraph H, a graph in $G_{n,p}$ will contain a copy of H. However, the theory behind this result is beyond the scope of this paper.

Theorem 3.1. *Let K_r be the complete graph on r vertices, $r \geq 3$*

$$\mathbb{P}(G_{n,p} \supseteq K_r) \to \begin{cases} 0 & \text{if } p \ll n^{-2/(r-1)} \\ 1 & \text{if } p \gg n^{-2/(r-1)} \end{cases}$$

Proof. For the first part of the proof we let $p \ll n^{-2/(r-1)}$. Let X be a random variable which counts the number of complete r-cliques, K_r, in G. Therefore, letting u be a set of vertices that forms a possible r-clique:

$$\begin{aligned} \mathbb{P}(G_{n,p} \supseteq K_r) &= \mathbb{P}(X \geq 1) \leq \mathbb{E}(X) \\ &= \sum_{|u|=r} \mathbb{E}(\mathbf{1}_{u \text{ is complete}}) = \binom{n}{r} p^{\binom{r}{2}} \ll n^r n^{\frac{-2\binom{r}{2}}{r-1}} = 1 \end{aligned} \tag{3.2}$$

The inequality in the first line comes from Markov. To explain the equality of (3.2), we are summing over the number of u, which are sets of r vertices. Our graph has n vertices and therefore there are $\binom{n}{r}$ possible u. For each u to be complete, each of its vertices must be connected with one another, so the probability of this occurring is $p^{\binom{r}{2}}$.

Thus in the first case, as n goes to infinity, the probability $G_{n,p}$ contains a K_r tends to 0.

Now assume $p \gg n^{\frac{-2}{r-1}}$. We want to show that $\mathbb{P}(X = 0) \to 0$. To do so we will use Chebychev's inequality,

$$\mathbb{P}(|X - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}$$

where μ is the mean of X and σ^2 is the variance of X. As a corollary to this, it follows that:

$$\mathbb{P}(X = 0) \leq \frac{\mathrm{Var(X)}}{\mathbb{E}(X)^2}$$

Using the identity that $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, we can then bound the value for the variance of X.

We will be summing over ordered pairs (u, v) where u and v represent potential K_r. We will use the notation $u \wedge v$ to describe the case where both u is an r-clique and v is an r-clique:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 \tag{3.3}$$

$$= \mathbb{E}\left(\sum_{(u,v)} \mathbf{1}[u]\mathbf{1}[v]\right) - \left(\sum_{u} \mathbb{P}(u)\right)^2 \tag{3.4}$$

$$= \sum_{(u,v)} \left(\mathbb{P}(u \wedge v) - \mathbb{P}(u)\mathbb{P}(v)\right) \tag{3.5}$$

$$\leq \sum_{2 \leq |u \cap v| < r} \mathbb{P}(u \wedge v) + \sum_{|u \cap v| = r} \mathbb{P}(u \wedge v) \tag{3.6}$$

$$\leq \sum_{l=2}^{r-1} \left(n^{2r-l} p^{\binom{2r-l}{2}}\right) + n^r p^{\binom{r}{2}} \tag{3.7}$$

Going from (3.6) to (3.7), we are summing over the probabilities that two randomly chosen sets of r vertices will be both complete. This relies on the number of edges their union shares, for if they share all their vertices, the probability that they are both complete would be $p^{\binom{r}{2}}$ as there are only r edges. However, if they only share two vertices, that implies they only share one edge, so there would be $\binom{2r-1}{2}$ edges that need to occur for them both to be complete.

In (3.7) we are summing over all possible configurations of pairs of r vertices where in the first sum, we cover the cases where the two sets would possibly share an edge, while in the second sum, where the two sets are equal, and therefore the second sum is just the expected value of X. Thus:

$$\mathbb{P}(G_{n,p} \not\supseteq K_r) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}(X)^2}$$

$$\leq \left(\sum_{l=2}^{r-1} \left(n^{2r-l} p^{\binom{2r-l}{2}}\right) + n^r p^{\binom{r}{2}}\right) \Big/ n^{2r} p^{2\binom{r}{2}} \tag{3.8}$$

We want to show when $p \gg n^{\frac{-2}{r-1}}$ and n increases, then the whole sum is much less than 1.

To do so we can split the fraction into two parts, and show that they are both much less than 1.

$$\left(\sum_{l=2}^{r-1} \left(n^{2r-l} p^{\binom{2r-l}{2}}\right) + n^r p^{\binom{r}{2}}\right) \Big/ n^{2r} p^{2\binom{r}{2}} = \frac{\sum_{l=2}^{r-1} \left(n^{2r-l} p^{\binom{2r-l}{2}}\right)}{n^{2r} p^{2\binom{r}{2}}} + \frac{n^r p^{\binom{r}{2}}}{n^{2r} p^{2\binom{r}{2}}}$$

$$\frac{n^r p^{\binom{r}{2}}}{n^{2r} p^{2\binom{r}{2}}} = \frac{1}{n^r p^{\binom{r}{2}}} \ll \frac{1}{n^{\frac{-2}{r-1}\binom{r}{2}}} = \frac{1}{n^r n^{-r}} = 1$$

For the second fraction, instead of dealing with a finite sum, we will show that for any value of l, $2 \leq l \leq r-1$, the fraction is much less than 1.

$$\frac{n^{2r-l}p^{\binom{2r-l}{2}}}{n^{2r}p^{2\binom{r}{2}}} = \frac{1}{n^l p^{2\binom{r}{2}-\binom{2r-l}{2}}} \ll 1/n^l n^{\frac{-2}{r-1}\left(r(r-1)-\frac{(2r-l)(2r-l-1)}{2}\right)} \tag{3.9}$$

From where we want to show that the denominator of (3.9) is greater than 1, so the fraction is much less than 1. To do so we will show that the exponents on the variable n in the denominator of (3.9) will be greater than or equal to 0. Therefore our proof follows if we prove that:

$$\frac{-2}{r-1}\left(r(r-1) - \frac{(2r-l)(2r-l-1)}{2}\right) \geq -l$$

This follows as $l \leq r$ and thus $2r - l \geq 2r - r$:

$$-2r + \frac{(2r-l)(2r-l-1)}{r-1} \geq -2r + \frac{(2r-l)(2r-r-1)}{r-1} = -l$$

And so we have proved that $1/n^l n^{\frac{-2}{r-1}\left(r(r-1)-\frac{(2r-l)(2r-l-1)}{2}\right)} \leq 1$. Therefore when $p \gg n^{\frac{-2}{r-1}}$ then:

$$\frac{n^{2r-l}p^{\binom{2r-l}{2}}}{n^{2r}p^{2\binom{r}{2}}} \ll 1$$

And so

$$\left(\sum_{l=2}^{r-1}\left(n^{2r-l}p^{\binom{2r-l}{2}}\right) + n^r p^{\binom{r}{2}}\right)/n^{2r}p^{2\binom{r}{2}} \ll 1$$

And thus to conclude:

$$\mathbb{P}(G_{n,p} \not\supseteq K_r) \leq \frac{\mathrm{Var}(X)}{\mathbb{E}(X)^2} \ll 1$$

□

One of the original proofs of this topic is by Béla Bollobás, who used a purely counting argument. Our proof however gives the notion of a sharp bound on p on which the graph's behavior depends upon.

Knowing which types of graphs contain certain subgraphs is valuable knowledge in graph theory and combinatorics, as well as useful information to be applied. Random graph models that generate large or frequent cliques are well suited for modeling social networks, as humans tend to organize themselves socially within interconnected groups of friends and acquaintances.

4. Connectivity

Now we shift our attention to another graph property, connectedness.

Definition 4.1. We call a graph property Q monotone increasing if for two graphs, H and G such that $H \subseteq G$:

$$H \in Q \implies G \in Q$$

An example of a monotone increasing property is containing a k-cycle, where if a subgraph of G contains a k-cycle, clearly G itself contains a k-cycle. Monotonically increasing properties are some of the most commonly studied, as once they emerge in a graph process they cannot disappear. We introduce this notion because all of our concepts rely on the presence or emergence of these graph properties. An intuitive yet powerful theorem follows:

Theorem 4.2. *For two probabilities p_1, p_2 such that $p_1 \leq p_2$ and for a monotone increasing graph property Q then:*

$$\mathbb{P}(G_{n,p_1} \in Q) \leq \mathbb{P}(G_{n,p_2} \in Q)$$

Proof. Let $p = (p_2 - p_1)/(1 - p_1)$. Choose independently $G_1 \in \mathbb{G}(n, p_1)$ and $G \in \mathbb{G}(n, p)$, and let $G_2 = G_1 \cup G$.
The edges of G_2 have thus been selected independently with probability $p_1 + p - p_1 p = p_2$, so G_2 is an element of $G(n, p_2)$. However, because Q is monotonically increasing then if G_1 has Q than G_2 must have Q. Therefore $\mathbb{P}(G_{n,p_1} \in Q) \leq \mathbb{P}(G_{n,p_1} \in Q)$. □

Using similar logic, the opposite statement holds true for monotonically decreasing properties. This theorem gives us the notion that for an increasing graph property, a random graph with a higher edge-probability will be more likely to contain that property. Equivalently, adding edges to a graph is more likely to result in a graph having the desired increasing property. One of these properties that we are interested in is the connectivity of a graph. Similar to our proof of the existence of k-cliques, we can define a value of p such that w.h.p. the graph $G_{n,p}$ will be connected or not:

Theorem 4.3. *For all $\epsilon > 0$*

$$\mathbb{P}(G_{n,p} \text{ is connected}) \to \begin{cases} 0 & \text{if } p \leq \frac{(1-\epsilon)\log(n)}{n} \\ \\ 1 & \text{if } p \geq \frac{(1-\epsilon)\log(n)}{n} \end{cases}$$

We will prove the theorem below, but first it is interesting to note that compared to the bound of p in the previous section, where p had to be asymptotically less than or greater than a function of n^{-x}, now we are working with a discrete epsilon bound of a function of $\frac{\log(n)}{n}$, suggesting that the nature for a random graph to be connected is much more sensitive to the value of p than that of containing a subgraph.

Proof. If a graph has any isolated vertices, then it is not connected. So to prove the first part we will define the random variable X as the number of isolated vertices in our graph, and let $p = \frac{(1-\epsilon)\log(n)}{n}$. Therefore, for sufficiently large n:

$$\mathbb{E}(X(G_{n,p})) = n(1-p)^{n-1} > ne^{-pn-p^2 n} \gg n^{\epsilon/2} \tag{4.4}$$

The first equality in this equation comes from the fact that if a node is isolated, it must not be connected to any other node, the probability of which is $(1-p)^{n-1}$, and we are counting over n possible isolated nodes. We then used the inequality where for sufficiently small $x > 0$,

$$(1 - x) > e^{-x-x^2}$$

Additionally we know the variance of X is bounded, using a similar method as the proof of [3.1]:

$$\begin{aligned} \text{Var}(X(G_{n,p})) &\leq \sum_{(u,v)} \left(\mathbb{P}(u \wedge v) - \mathbb{P}(u)\mathbb{P}(v)\right) \\ &\leq \mathbb{E}(X) + n^2\left((1-p)^{2n-3} - (1-p)^{2n-2}\right) \qquad (4.5) \\ &= \mathbb{E}(X) + \frac{p}{1-p}\mathbb{E}(X)^2 \qquad (4.6) \end{aligned}$$

Therefore, $\text{Var}(X) \ll \mathbb{E}(X)^2$, so Chebychev implies that as $n \to \infty$, the probability that $G_{n,p}$ is connected goes to 0.

For the second part of the proof, let the random variable Y_k count the number of components of size k in G. We then let $p = \frac{(1+\epsilon)\log(n)}{n}$.
If $k \leq \epsilon n$ then we use the following inequality:

$$\binom{n}{k} \leq (\frac{en}{k})^k$$

Hence,

$$\mathbb{E}(Y_k(G_{n,p})) = \binom{n}{k}(1-p)^{k(n-k)} \leq \left(\frac{en}{k}e^{-(1-\epsilon)pn}\right)^k$$

Additionally we note that if G is not connected, then there must exist a component of size at most $\frac{n}{2}$. Therefore in the case where $k \geq \epsilon n$,

$$\mathbb{E}(Y_k(G_{n,p})) = \binom{n}{k}(1-p)^{k(n-k)} \leq 2^n e^{-\epsilon(1+\epsilon)\log(n)\frac{n}{2}} \tag{4.7}$$

We then apply Markov and thus:

$$\begin{aligned}
\mathbb{P}(G_{n,p} \text{ is not connected}) &\leq \sum_{k=1}^{n/2} \mathbb{E}(Y_k(G_{n,p})) \\
&= \sum_{k=1}^{\epsilon n} \left(\frac{en}{k}e^{-(1-\epsilon)pn}\right)^k + \sum_{k=\epsilon n}^{n/2} 2^n e^{-\epsilon(1+\epsilon)\log(n)\frac{n}{2}} \\
&\leq \sum_{k=1}^{\infty} \left(\frac{e}{k}e^{(-1+\epsilon)(1+\epsilon)\log n}\right)^k + \sum_{k=\epsilon n}^{n/2} 2^n e^{-\epsilon(1+\epsilon)\log(n)\frac{n}{2}} \\
&\leq \sum_{k=1}^{\infty} \left(\frac{e}{k}e^{-\epsilon\log n}\right)^k + ce^{-\frac{\epsilon n\log(n)}{2}} \\
&\leq \sum_{k=1}^{\infty} \left(\frac{e}{k}n^{-\epsilon}\right)^k + ce^{-\frac{\epsilon n\log(n)}{2}} \to 0 \qquad (4.8)
\end{aligned}$$

Expression (4.8) tends to 0 as $n \to \infty$, where c is a constant. □

In this proof we have defined p as a discrete function of ϵ and n, yet for p less than or equal to the values we have given, the proof holds due to Theorem 4.2.
It is significant that the bound of p for which a graph is connected is so sharp, and in fact this sudden transition can be expanded upon further. A later result in Bollobás seminal text on random graphs finds the precise time t (after t graph processes) where a random graph becomes connected.

5. Emergence of the Giant Component

One of Erdös and Rényi's most striking discoveries was that prior to a graph becoming connected, random graphs tended to follow a similar pattern, where there were few components of small size, and one of large size. This largest component was dubbed "the giant component" and is defined as the component whose size is greater than $n^{2/3}$.

As a graph process continues, after a certain point one component tends to dominate in size while the other components grow at a much lower order. The

logic of this can be illustrated when thinking about an example, the popular Erdös number. In the universe of mathematical papers, Erdös and his collaborators have a large interconnected web of collaborations. There may be other separate families of mathematicians who collaborate with one another yet are independent from the Erdös collection, yet due to the large size of this component, it is extremely probable that a mathematician would collaborate with another mathematician who has some sort of link to Paul Erdös, and suddenly that separate family of research becomes connected to the Erdös web.

Additionally the growth of this giant component is particularly important to mathematicians. It turns out when a random graph is fairly young, precisely when the average degree of the graph is less than 1, all components of a random graph grow with logarithmic order $O(\log n)$. However as the average degree passes through one, suddenly there emerges a component which begins to grow with linear order. We will be presenting one of Bollobás first results on the topic, which is proving that as the time t of a graph process is around $n/2$, then almost every graph process will see a gap between the size of its components emerge [1]. Precisely we have that the graph G_t has no component whose orders are between $n^{2/3}/2$ and $n^{2/3}$. Formally defining the term "almost every", if Q is any property, then almost every graph of a certain distribution has property Q if $\mathbb{P}(Q) \to 1$ as n approaches infinity [1].

Theorem 5.1 (Bollobás 1984)**.** *Let $s_0 = (2\log n)^{1/2}n^{2/3}$. For a variable s, $s \neq 0, s_0 \leq |s| \leq \frac{n}{2}$, let $k_0(s) = \lfloor (3\log s - \log n - \log\log n)n^2/s^2 \rfloor$.*
Then almost every graph process $\tilde{G} = (G_t)_0^{\binom{n}{2}}$ is such that if $t = \frac{n}{2} + s$ and G_t has a component of order k, then either $k < k_0(s)$ or $k > n^{2/3}$.

Proof. We refer to a component of our graph as a (k, d) component if it has k vertices and $k + d$ edges. We will count the number of these components with the random variable $X(k, d)$.
Without loss of generality let $s > 0$, where from the statement of the theorem s can be thought of as the additional time after $t = n/2$ steps of a graph process. The proof for $s < 0$ follows similarly. Therefore $t > \frac{n}{2}$. Additionally we are only interested in counting those components whose order are at most $n^{2/3}$, so we suppose that $k_0(s) \leq k \leq n^{2/3}$.

Let $\mathbb{E}_t(k)$ be the expected number of components of order k in G_t. Then:

$$\mathbb{E}_t(k) = \sum [\mathbb{E}_t\{X(k, d)\} : -1 \leq d \leq \binom{k}{2} - k]$$

The expected value of the number of components of order k is the sum of all (k, d) components. Let $k_1(s) = \min\{\lfloor n^{2/3} \rfloor, 3k_0(s)\}$. Therefore $k_0(s_0) < \frac{1}{2}n^{2/3}$ and the functions $k_0(s), k_1(s)$ are monotonically decreasing in $s_0 \leq s < n/2$ and $k_0(s + 1) < 2k_1(s)$.

Therefore it suffices to show that almost every graph process is such that for $s_0 \leq s < n/2$ and $t = n/2 + s$, the graph G_t has no component whose order is at least $k_0(s)$ and at most $k_1(s)$.

Suppose $\tilde{G}$ is such a graph process, let $t < n-1$, and assume G_t has no component whose order is between $k_0(s)$ and $n^{2/3}$.

Since $k_1(s + 1) < 2k_0(s)$, the $(t + 1)$st edge of the graph process cannot merge two components of G_t with fewer than $k_0(s)$ vertices to form a component of order at least $k_1(s + 1)$ in G_{t+1}.

Therefore G_{t+1} has no component whose order is between $k_0(s+1)$ and $k_1(s+1)$, so G_{t+1} has no components of order between $k_0(s+1)$ and $n^{2/3}$.

It then remains to show that:

$$\sum_{t=t_0}^{n} \sum_{k=k_0(s)}^{k_1(s)} \mathbb{E}_t(k) = o(1) \tag{5.2}$$

If we show this then we show that the limit of the expected value of the number of components of order k that are between $k_0(s)$ and $k_1(s)$ tends to 0. Therefore with large enough n and high probability then the graph process G_t will not have components within this range and we will be done.

This is where the heavy computation and manipulation in the proof lies, yet the theory used is beyond the scope of this paper, so instead we will cover the morality of the argument. Bollobás bounds $\mathbb{E}_t(k)$ as

$$\mathbb{E}_t(k) \leq 4n^{1/2}\mathbb{E}_p(k)$$

where $\mathbb{E}_p(k)$ is defined the same as $\mathbb{E}_t(k)$, yet as the expected number of k-order components in a Gilbert graph rather than a graph process. From there he proves that the sum of these components equals $o(n^{-1/2})$, which in turn proves 5.2. The details of the proof can be found in chapter 6 of Bollobás text, yet the main idea is that by bounding the expected value of k-order components by the expected value of random variables that are easier to mention, we can perform manipulations to reach the desired conclusion. □

This giant component and its extremely fast growth rate is one of the most applied behaviors of random graph in a modeling capacity. In the case of genealogy and ancestry trees it is possible to predict which branches of a family tree or genus will most likely die out while which branch will prosper. The sharp change in order of the giant component has been used to model critical phase transitions of matter going through changes of state. The giant component has even been proven to entirely dictate the expected cost of a graph process relative to the Weighted Quick-Find algorithm, which will be presented below and can be explored in further detail in chapter 6 of Bollobás' text [1]:

Problem 5.3. Weighted Quick-Find Algorithm

The Union-Find problem proposed by Aho, Hopcroft, and Ullman deals with a set S of items which are separated into partitions and attempts to design an efficient algorithm to connect items into the same partition. For example, this problem can be used in computer science graph theory to efficiently connect two vertices of a graph, where the set S would correspond to the set of vertices, and the partitions of S would correspond to connected sets of vertices. Regardless of the data structure, the computer numbers the items of S numerically and stores them in an array.

Union-Find algorithms want to perform two operations, "Find" and "Union". Find determines if two items are connected by determining if the two numbers belong in the same set. Union connects two items by creating a new partition which is the union of the partitions containing both respectively. To illustrate, let S be an array of numbers and let curly brackets denote partitions of connected items, [**1 2 3 4** {**5 6 7**} {**8 9**}]. In this example, Find[**5**, **6**] would return true, as **5** and **6** are in the same partition. Union[**5**, **8**] would create a new partition to connect **5** and **8**, and our data set S would then look like [**1 2 3 4** {**5 6 7 8 9**}]. The goal of the

problem is to find data structures that allow these two operations to be conducted most efficiently.

The simplest algorithm uses an array "Name" which stores the name for the partition containing certain elements. Therefore, if you had the previous example:

$$S = [\mathbf{1\,2\,3\,4}\,\{\mathbf{5\,6\,7}\}\,\{\mathbf{8\,9}\}]$$
$$\text{Name} = [\mathbf{1\,2\,3\,4}\quad \mathbf{5\,5\,5}\quad \mathbf{6\,6}]$$

In our data set, the items $\mathbf{5}, \mathbf{6}, \mathbf{7}$ are connected as they are in the same partition. The Name array makes note of this by assigning each of those items the same number in the array, in this case the number 5.

In this basic algorithm, Find[$\mathbf{x}, \mathbf{y}$] merely checks the values of Name and determines if the two names are equal. Find[$\mathbf{7}, \mathbf{8}$] would look up the 7th and 8th positions in the Name array, which are 5 and 6 respectively, and determine if they are equal. In this case Find[$\mathbf{7}, \mathbf{8}$] would then return False.

Let the partition containing an element $\mathbf{x}$ be denoted as C_x, while $\mathbf{y}$ is in the partition of C_y. As stated before, the result of Union[$\mathbf{x}, \mathbf{y}$] a new partition $C_x \cup C_y$.

Suppose we implement our simple algorithm and find that the items $\mathbf{5}$ and $\mathbf{8}$ are not in the same partition, and we wish to apply Union[$\mathbf{5}, \mathbf{8}$]. Our information about which item belongs to which partition is stored in the Name array. Therefore, to union their respective partitions, the computer must change all the values of items belonging to C_5 in the Name array to the same number as the one that corresponds to C_8. Union[$\mathbf{5}, \mathbf{8}$] would then update the Name array to read as:

$$\text{Name} = [\mathbf{1\,2\,3\,4}\quad \mathbf{6\,6\,6}\quad \mathbf{6\,6}]$$

The new data set looks like:

$$S = [\mathbf{1\,2\,3\,4}\,\{\mathbf{5\,6\,7\,8\,9}\}]$$

Formally, to implement Union[$\mathbf{x}, \mathbf{y}$] in this simple algorithm where Name[$\mathbf{x}$] $\neq$ Name[$\mathbf{y}$] one must first set Name[$\mathbf{u}$] to equal Name[$\mathbf{y}$] for each $\mathbf{u}$ in C_x and then finally append C_x to C_y. This is exactly what was done in the example, where the names of the elements $\mathbf{5}, \mathbf{6}, \mathbf{7}$ were changed to equal the names of $\mathbf{8}$ and $\mathbf{9}$.

Therefore, the cost of the operation Union is proportional to the cardinality of C_x. Note that we can make this operation faster if we change the names of the items belonging to the smaller partition to that of the greater partition. By maintaining a separate array N[$\mathbf{s}$] which stores the cardinality of the partition containing $\mathbf{s}$, and changing the name of the element which belongs to the smaller of the two partitions, the cost of Union[$\mathbf{x}, \mathbf{y}$] reduces to $\min\{|C_x|, |C_y|\}$. This in turn is the **Weighted Quick-Find algorithm**.

When running a graph process on a computer, adding in another random edge to $\tilde{G} = (G_t)_0^n$ can be interpreted by the computer through the lens of this Union-Find problem as a union between two vertices. One can consider the t^{th} graph process operation as Union$[x, y]$, where the t^{th} edge added into the graph is xy if $G_t = G_{t-1} + xy$. Therefore running a graph process on a computer efficiently is a form of the Union-Find problem. We can then say that the cost or weight of this tth operation is $w(x, y) = \min\{|C_x|, |C_y|\}$, where C_x and C_y are the vertex sets containing x and y respectively. The final cost of the graph $W(\tilde{G})$ is the sum of the costs of its edges. A result by Bollobás and Simon concerning the expectation of $W(\tilde{G})$ follows:

Theorem 5.4 (1993). *The expected cost of a random graph process $\tilde{G}$ relative to the Weighted Quick-Find algorithm is*

$$\mathbb{E}(W(\tilde{G})) = c_0 n + O(n/\log n)$$

This theorem gives us a notion of the resource cost of running a graph process on a computer using the Weighted Quick-Find algorithm.

6. The "Small-World" Problem

The Small-World problem is a particularly famous problem stating that most people in the world are seperated by around 6 or less degrees of seperation. The problem has been explored experimentally, but through random graphs we can create close models to human social networks and attempt to prove the concept through these abstract models.

The Small-World problem is originally found in the field of psychology and anthropology. The problem got its name from a paper written by American social psychologist Stanley Milgram published in 1967 [6]. Using empirical and experimental data, Milgram observed that as certain social systems grow, they become connected such that two nodes in a model have at most 6 degrees of separation between them.

Random graphs were first used in measuring this observation in the paper "Collective dynamics of 'small-world' networks"[4] by sociologist Duncan Watts and mathematician Steven Strogatz in 1998. They developed what is know referred to as the Watts and Strogatz model of random graphs, which they exhibited to possess the key features of social networks that our familiar models of random graphs could not exhibit.

A typical Erdös-Rényi graph tends to have a low clustering coefficient, denoted as $C(p)$, which measures the clique-forming property of the graph and is defined as follows:

Definition 6.1. We denote c_i to be the number of neighbors of vertex i. This is predetermined by the model, and refers to the x nodes that are geometrically closest to vertex i. There are $c_i(c_i - 1)/2$ possible edges between these neighbors. Denote C_i as the fraction of these edges that are actually present in the graph. $C(p)$ is then the average over all C_i

In social networks, humans tend to live in cities or areas with denser populations and be involved in tight friend groups, as opposed to being evenly spread out around the globe. The Watts and Strogatz graph is generated by an algorithm which produces a graph with high clustering coefficient.

To generate these types of graphs, we perform the following algorithm:

Definition 6.2. A Watts-Strogatz graph is a random graph generated by the following algorithm:

(1) Construct a regular ring lattice, a graph with n nodes where each node is connected to its k geometrically closest neighbors.
(2) For each node n_i, take every edge $n_i n_j$ and reconnect to another node with probability p. That is, with probability p, edge $n_i n_j$ becomes replaced by $n_i n_k$ where n_k is randomly chosen with equal probability out of all nodes that are not n_i or n_j.

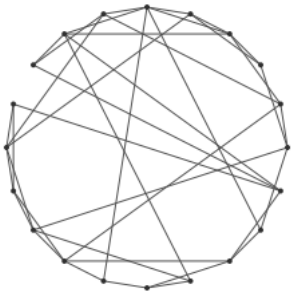

FIGURE 1. Graph generated by W-S model

The following example depicts a graph generated in this manner. Notably, this graph is bound to have high clustering as the ring lattice it starts from is a x-regular graph, thus its vertices are highly clustered around their neighborhood. The random reconnecting of edges in turn separates clusters from one another and also helps greatly reduce path lengths.

Why is this graph model then suitable for modeling social networks? Some blaring simplifications are obvious, most notably that the Watts-Strogatz model produces an x-regular graph. In a social network that would imply each node is connected to a constant amount, which is not true. However, when dealing with large enough graphs, this assumption becomes less and less impactful as it seems almost feasible that most humans have a similar number of friends or aquiantances. What is interesting however is the clustering coefficient of these graphs. Using experimental data, the graphs produced from the models had similar clustering coefficients to various human networks.

From here Watts and Strogatz then decided to test properties of their graphs, and found that with high probability a pair of two randomly chosen points would have a path length of around 6. By finding a plausible model, Watts and Strogatz were able to test observations made by sociologists decades earlier.

This is just a broad view of how random graphs are applied to existing problems, for a more in depth explanation of the process I would recommend the papers published by Watts and Strogatz and the paper "On the properties of small-world network models" written by A. Barret and M. Weigt [5].

Hopefully this problem serves as an illustration as to how random graphs not only reveal or explain certain phenomena, but they dictate the very nuanced properties that are inherent to any network or system.

Acknowledgments. It is a pleasure to thank my mentor, Anthony Santiago Chaves Aguilar, for the support and extensive knowledge into this topic that made this paper possible.

REFERENCES

[1] Bela Bollobás. Random Graphs Second Edition. Print.
[2] Robert Morris, Roberto Imbuzeiro Oliviera. Extremal and Probabilistic Combinatorics. https://impa.br/wp-content/uploads/2017/04/28CBM_04.pdf

[3] Mihyun Kang, Zdeněk Petrášek. Random Graphs: Theory and Applications from Nature to Society to the Brain. https://www.math.tugraz.at/ kang/papers/imn-randomgraphs.pdf
[4] Duncan Watts, Steven Strogatz. Collective dynamics of 'small-world' networks. http://worrydream.com/refs/Watts-CollectiveDynamicsOfSmallWorldNetworks.pdf
[5] Alain Barrat, Martin Weigt. On the properties of small-world network models. http://www.cpt.univ-mrs.fr/ barrat/epjb13.pdf
[6] Stanley Milgram, Jeffrey Travers. An Experimental Study of the Small World Problem. https://snap.stanford.edu/class/cs224w-readings/travers69smallworld.pdf
[7] Steven Kleiman, Glenn Tesler. Writing a Math Phase Two Paper. http://web.mit.edu/18.704/www/piiUJM2.pdf
[8] Tobias Oekiter, Hubert Partl, Irene Hyna and Elisabeth Schlegl. The Not So Short Introduction to LaTeX 2_ε. http://tobi.oetiker.ch/lshort/lshort.pdf.