1

Overview

1.1 Introduction to the Introduction

The theory of random graphs began in the late 1950s in several papers by Erdös and Rényi. However, the introduction at the end of the twentieth century of the small world model of Watts and Strogatz (1998) and the preferential attachment model of Barabási and Albert (1999) have led to an explosion of research. Querying the Science Citation Index in early July 2005 produced 1154 citations for Watts and Strogatz (1998) and 964 for Barabási and Albert (1999). Survey articles of Albert and Barabási (2002), Dorogovstev and Mendes (2002), and Newman (2003) each have hundreds of references. A book edited by Newman, Barabási, and Watts (2006) contains some of the most important papers. Books by Watts (2003) and Barabási (2002) give popular accounts of the new science of networks, which explains "how everything is connected to everything else and what it means for science, business, and everyday life."[1]

While this literature is extensive, many of the papers are outside the mathematical literature, which makes writing this book a challenge and an opportunity. A number of articles have appeared in *Nature* and *Science*. These journals with their impressive impact factors are, at least in the case of random graphs, the home of *10 second sound bite science*. An example is the claim that "the Internet is robust yet fragile. 95% of the links can be removed and the graph will stay connected. However, targeted removal of 2.3% of the hubs would disconnect the Internet."

These shocking statements grab headlines. Then long after the excitement has subsided, less visible papers show that these results aren't quite correct. When 95% of links are removed the Internet is connected, but the fraction of nodes in the giant component is 5.9×10^{-8}, so if all 6 billion people were connected initially then after the links are removed only 360 people can check their e-mail. The targeted removal result depends heavily on the fact that the degree distribution was assumed to be exactly a power law for all values of k, which forces $p_k \sim 0.832k^{-3}$. However,

[1] This is the subtitle of Barabási's book.

if the graph is generated by the preferential attachment model with $m = 2$ then $p_k \sim 12k^{-3}$ and one must remove 33% of the hubs. See Section 4.7 for more details.

Many of the papers we cover were published in *Physical Review E*. In these we encounter the usual tension when mathematicians and physicists work on the same problems. Feynman once said "if all of mathematics disappeared it would set physics back one week." In the other direction, mathematicians complain when physicists leap over technicalities, such as throwing away terms they don't like in differential equations. They compute critical values for random graphs by asserting that cluster growth is a branching process and then calculating when the mean number of children is > 1. Mathematicians worry about justifying such approximations and spend a lot of effort coping with paranoid delusions, for example, in Section 4.2 that a sequence of numbers all of which lie between 1 and 2 might not converge.

Mathematicians cherish the rare moments where physicists' leaps of faith get them into trouble. In the current setting, physicists use the branching process picture of cluster growth when the cluster is of order n (and the approximation is not valid) to compute the average distance between points on the giant component of the random graph. As we will see, the correct way to estimate the distance from x to y is to grow the clusters until they have size $C\sqrt{n}$ and argue that they will intersect with high probability. In most cases, the two viewpoints give the same answer, but in the case of some power law graphs, the physicists' argument misses a power of 2, see Section 4.5.

While it is fun to point out physicists' errors, it is much more satisfying when we discover something that they don't know. Barbour and Reinert (2001) have shown for the small world and van der Hofstad, Hooghiemstra, and Znamenski (2005) have proved for models with a fixed degree distribution, see Theorems 5.2.1 and 3.4.1, that the fluctuations in the distance between two randomly chosen points are $O(1)$, a result that was not anticipated by simulation. We have been able to compute the critical value of the Ising model on the small world exactly, see Section 5.4, confirming the value physicists found by simulation. A third example is the Kosterlitz–Thouless transition in the CHKNS model. The five authors who introduced this model (only one of whom is a physicist) found the phenomenon by numerically solving a differential equation. Physicists Dorogovstev, Mendes, and Samukhin (2001) demonstrated this by a detailed and semi-rigorous analysis of a generating function. However, the rigorous proof of Bollobás, Janson, and Riordan (2005), which is not difficult and given in full in Section 7.4, helps explain why this is true.

Despite remarks in the last few paragraph, our goal is not to lift ourselves up by putting other people down. As Mark Newman said in an e-mail to me "I think there's room in the world for people who have good ideas but don't have the rigor to pursue them properly – makes more for mathematicians to do." The purpose of this book is to give an exposition of results in this area and to provide proofs for some

facts that had been previously demonstrated by heuristics and simulation, as well as to establish some new results. This task is interesting since it involves a wide variety of mathematics: random walks, large deviations, branching processes, branching random walks, martingales, urn schemes, and the modern theory of Markov chains that emphasizes quantitative estimates of convergence rates.

Much of this book concentrates on geometric properties of the random graphs: primarily emergence of a giant component and its small diameter. However, our main interest here is in processes taking place on these graphs, which is one of the two meanings of our title, *Random Graph Dynamics*. The other meaning is that we will be interested in graphs such as the preferential attachment model and the CHKNS model described in the final section that are grown dynamically rather than statically defined.

1.2 Erdös, Rényi, Molloy, and Reed

In the late 1950s, Erdös and Rényi introduced two random graph models. In each there are n vertices. In the first and less commonly used version, one picks m of the $n(n-1)/2$ possible edges between these vertices at random. Investigation of the properties of this model tells us what a "typical" graph with n vertices and m edges looks like. However, there is a small and annoying amount of dependence caused by picking a fixed number of edges, so here we will follow the more common approach of studying the version in which each of the $n(n-1)/2$ possible edges between these vertices are independently present with probability p. When $p = 2m/n(n-1)$, the second model is closely related to the first.

Erdös and Rényi discovered that there was a sharp threshold for the appearance of many properties. One of the first properties that was studied, and that will be the focus of much or our attention here, is the emergence of a giant component.

- If $p = c/n$ and $c < 1$ then, when n is large, most of the connected components of the graph are small, with the largest having only $O(\log n)$ vertices, where the O symbol means that there is a constant $C < \infty$ so that the probability the largest component is $\le C \log n$ tends to 1 as $n \to \infty$.
- In contrast if $c > 1$ there is a constant $\theta(c) > 0$ so that for large n the largest component has $\sim \theta(c)n$ vertices and the second largest component is $O(\log n)$. Here $X_n \sim b_n$ means that X_n/b_n converges to 1 in probability as $n \to \infty$.

Chapter 2 is devoted to a study of this transition and properties of Erdös–Rényi random graphs below, above, and near the critical value $p = 1/n$. Much of this material is well known and can be found in considerably more detail in Bollobás' (2001) book, but the approach here is more probabilistic than combinatorial, and in any case an understanding of this material is important for tackling the more complicated graphs, we will consider later.

In the theory of random graphs, most of the answers can be guessed using the heuristic that the growth of the cluster is like that of a branching process. In *Physical Review E*, these arguments are enough to establish the result. To explain the branching process approximation for Erdös–Rényi random graphs, suppose we start with a vertex, say 1. It will be connected to a Binomial $(n-1, c/n)$ number of neighbors, which converges to a Poisson distribution with mean c as $n \to \infty$. We consider the neighbors of 1 to be its children, the neighbors of its neighbors to be its grandchildren, and so forth. If we let Z_k be the number of vertices at distance k, then for small k, Z_k behaves like a branching process in which each individual has an independent and mean c number of children.

There are three sources of error: (i) If we have already exposed $Z_0 + \cdots + Z_k = m$ vertices then the members of the kth generation have only $n - m$ new possibilities for connections; (ii) Two or more members of the kth generation can have the same child; and (iii) Members of the branching process that have no counterpart in the growing cluster can have children. In Section 2.2, we will show that when $m = o(\sqrt{n})$, that is, $m/\sqrt{n} \to 0$, the growing cluster is equal to the branching process with high probability, and when $m = O(n^{1-\epsilon})$ with $\epsilon > 0$ the errors are of a smaller order than the size of the cluster.

When $c < 1$ the expected number of children in generation k is c^k which converges to 0 exponentially fast and the largest of the components containing the n vertices will be $O(\log n)$. When $c > 1$ there is a probability $\theta(c) > 0$ that the branching process does not die out. To construct the giant component, we argue that with probability $1 - o(n^{-1})$ two clusters that grow to size $n^{1/2+\epsilon}$ will intersect. The result about the second largest component comes from the fact with probability $1 - o(n^{-1})$ a cluster that reaches size $C \log n$ will grow to size $n^{1/2+\epsilon}$. An error term that is $o(n^{-1})$ guarantees that with high probability all clusters will do what we expect.

When $c > 1$ clusters that don't die out grow like c^k (at least as long as the branching process approximation is valid). Ignoring the parenthetical phrase we can set $c^k = n$ and solve to conclude that the giant component has "diameter" $k = \log n/(\log c)$. For a concrete example suppose $n = 6$ billion people on the planet and the mean number of neighbors $c = np = 42.62$. In this case, $\log n/(\log c) = 6$, or we have six degrees of separation between two randomly chosen individuals. We have placed diameter in quotation marks since it is commonly used in the physics literature for the distance between two randomly chosen points on the giant component. On the Erdös–Renyi random graphs the mathematically defined diameter is $\geq C \log n$ with $C > 1/\log c$, but exact asymptotics are not known, see the discussion after Theorem 2.4.2.

The first four sections of Chapter 2 are the most important for later developments. The next four can be skipped by readers eager to get to recent developments. In Section 2.5, we prove a central limit theorem for the size of the giant component. In Section 2.6, which introduces the combinatorial viewpoint, we show that away

from the critical value, that is, for $p = c/n$ with $c \neq 1$, most components are trees with sizes given by the Borel–Tanner distribution. A few components, $O(1)$, have one cycle, and only the giant component is more complicated.

Section 2.7 is devoted to the critical regime $p = 1/n + \theta/n^{4/3}$, where the largest components are of order $n^{2/3}$ and there can be components more complex than unicyclic. There is a wealth of detailed information about the critical region. The classic paper by Janson, Knuth, Luczak, and Pittel (1993) alone is 126 pages. Being a probabilist, we are content to state David Aldous' (1997) result which shows that in the limit as $n \to \infty$ the growth of large components is a multiplicative coalescent.

In Section 2.8, we investigate the threshold for connectivity, that is, ALL vertices in ONE component. As Theorem 2.8.1 shows and 2.8.3 makes more precise, the Erdös–Rényi random graph becomes connected when isolated vertices disappear, so the threshold $= (\log n)/n + O(1)$. The harder, upper bound, half of this result is used in Section 4.5 for studying the diameter of random graphs with power law degree distributions.

In Chapter 3, we turn our attention to graphs with a fixed degree distribution that has finite second moment. Bollobás (1988) proved results for the interesting special case of a random r-regular graph, but Molloy and Reed (1995) were the first to construct graphs with a general distribution of degrees. Here, we will use the approach of Newman, Strogatz, and Watts (2001, 2002) to define our model. Let $d_1, \ldots d_n$ be independent and have $P(d_i = k) = p_k$. Since we want d_i to be the degree of vertex i, we condition on $E_n = \{d_1 + \cdots + d_n \text{ is even}\}$. To construct the graph now we imagine d_i half-edges attached to i, and then pair the half-edges at random. The resulting graph may have self-loops and multiple edges between points. The number is $O(1)$ so this does not bother me, but if you want a nice clean graph, you can condition on the event A_n that there are no loops or multiple edges, which has $\lim_{n\to\infty} P(A_n) > 0$.

Again, interest focuses first on the existence of a giant component, and the answer can be derived by thinking about a branching process, but the condition is not that the mean $\sum_k kp_k > 1$. If we start with a given vertex x then the number of neighbors (the first generation in the branching process) has distribution p_j. However, this is not true for the second generation. A first generation vertex with degree k is k times as likely to be chosen as one with degree 1, so the distribution of the number of children of a first generation vertex is for $k \geq 1$

$$q_{k-1} = \frac{kp_k}{\mu} \quad \text{where} \quad \mu = \sum_k kp_k$$

The $k - 1$ on the left-hand side comes from the fact that we used up one edge connecting to the vertex. Note that since we have assumed p has finite second moment, q has finite mean $\nu = \sum_k k(k-1)p_k/\mu$.

q gives the distribution of the number of children in the second and all subsequent generations so, as one might guess, the condition for the existence of a

giant component is $\nu > 1$. The number of vertices in the kth generation grows like $\mu\nu^{k-1}$, so using the physicist's heuristic, the average distance between two points on the giant component is $\sim \log n/(\log \nu) = \log_\nu n$. This result is true and there is a remarkable result of van der Hofstad, Hooghiemstra, and Van Mieghem (2004a), see Theorem 3.4.1, which shows that the fluctuations around the mean are $O(1)$. Let H_n be the distance between 1 and 2 in the random graph on n vertices, and let $\bar{H}_n = (H_n | H_n < \infty)$. The Dutch trio showed that $H_n - [\log_\nu n]$ is $O(1)$, that is, the sequence of distributions is tight in the sense of weak convergence, and they proved a very precise result about the limiting behavior of this quantity. As far as I can tell the fact that the fluctuations are $O(1)$ was not guessed on the basis of simulations.

Section 3.3 is devoted to an

Open problem. *What is the size of the largest component when $\nu < 1$?*

The answer, $O(\log n)$, for Erdös–Renyi random graphs is not correct for graphs with a fixed degree distribution. For an example, suppose $p_k \sim Ck^{-\gamma}$ with $\gamma > 3$ so that the variance is finite. The degrees have $P(d_i > k) \sim Ck^{-(\gamma-1)}$ (here and in what follows C is a constant whose value is unimportant and may change from line to line). Setting $P(d_i > k) = 1/n$ and solving, we conclude that the largest of the n degrees is $O(n^{1/(\gamma-1)})$. Trivially, the largest component must be at least this large.

Conjecture. *If $p_k \sim Ck^{-\gamma}$ with $\gamma > 3$ then the largest cluster is $O(n^{1/(\gamma-1)})$.*

One significant problem in proving this is that in the second and subsequent generations the number of children has distribution $q_k \sim Ck^{-(\gamma-2)}$. One might think that this would make the largest of the n degrees $O(n^{1/(\gamma-2)})$, but this is false. The size-biased distribution q can only enhance the probability of degrees that are present in the graph, and the largest degree present is $O(n^{1/(\gamma-1)})$.

In support of the conjecture in the previous paragraph we will now describe a result of Chung and Lu (2002a, 2002b), who have introduced a variant of the Molloy and Reed model that is easier to study. Their model is specified by a collection of weights $w_1, \ldots, w_n$ that represent the expected degree sequence. The probability of an edge between i and j is $w_i w_j / \sum_k w_k$. They allow loops from i to i so that the expected degree at i is

$$\sum_j \frac{w_i w_j}{\sum_k w_k} = w_i$$

Of course, for this to make sense we need $(\max_i w_i)^2 < \sum_k w_k$.

Let $d = (1/n)\sum_k w_k$ be the average degree. As in the Molloy and Reed model, when we move to neighbors of a fixed vertex, vertices are chosen proportional to

their weights, that is, i is chosen with probability $w_i / \sum_k w_k$. Thus the relevant quantity for connectedness of the graph is the second-order average degree $\bar{d} = \sum_i w_i^2 / \sum_k w_k$.

Theorem 3.3.2. *Let $vol(S) = \sum_{i \in S} w_i$. If $\bar{d} < 1$ then all components have volume at most $A\sqrt{n}$ with probability at least*

$$1 - \frac{d\bar{d}^2}{A^2(1-\bar{d})}$$

Note that when $\gamma > 3$, $1/(\gamma - 1) < 1/2$ so this is consistent with the conjecture.

1.3 Six Degrees, Small Worlds

As Duncan Watts (2003) explains in his book *Six Degrees*, the inspiration for his thesis came from his father's remark that he was only six handshakes away from the president of the United States. This remark is a reference to "six degrees of separation," a phrase that you probably recognize, but what does it mean? There are a number of answers.

Answer 1. The most recent comes from the "Kevin Bacon game" that concerns the film actors graph. Two actors are connected by an edge if they appeared in the same movie. The objective is to link one actor to another by a path of the least distance. As three college students who were scheming to get on Jon Stewart's radio talk show observed, this could often be done efficiently by using Kevin Bacon as an intermediate.

This strategy leads to the concept of a Bacon number, that is, the shortest path connecting the actor to Kevin Bacon. For example, Woody Allen has a Bacon number of 2 since he was in *Sweet and Lowdown* with Sean Penn, and Sean Penn was in *Mystic River* with Kevin Bacon. The distribution of Bacon numbers given in the next table shows that most actors have a small Bacon number, with a median value of 3:

0	1	2	3	4	5	6	7	8
1	1673	130,851	349,031	84,615	6,718	788	107	11

The average distance from Kevin Bacon for all actors is 2.94, which says that two randomly chosen actors can be linked by a path through Kevin Bacon in an average of 6 steps. Albert Barabási, who will play a prominent role in the next section, and his collaborators, computed the average distance from each person to all of the others in the film actors graph. They found that Rod Steiger with an average distance of 2.53 was the best choice of intermediate. It took them a long time to find Kevin Bacon on their list, since he was in 876th place.

Erdös Numbers. The collaboration graph of mathematics, in which two individuals are connected by an edge if they have coauthored a paper, is also a small world. The Kevin Bacon of mathematics is Paul Erdös, who wrote more than 1500 papers with more than 500 coauthors. Jerrold Grossman (2000) used 60 years of data from MathSciNet to construct a mathematical collaboration graph with 337,454 vertices (authors) and 496,489 edges. There were 84,115 isolated vertices. Discarding these gives a graph with average degree 3.92, and a giant component with 208,200 vertices with the remaining 45,139 vertices in 16,883 components. The average Erdös number is 4.7 with the largest known finite Erdös number within mathematics being 15. Based on a random sample of 66 pairs, the average distance between two individuals was 7.37. These numbers are likely to change over time. In the 1940s, 91% of mathematics papers had one author, while in the 1990s only 54% did.

Answer 2. The phrase "six degrees of separation" statement is most commonly associated with a 1967 experiment conducted by Stanley Milgram, a Harvard social psychologist, who was interested in the average distance between two people. In his study, which was first published in the popular magazine *Psychology Today* as "The Small World Problem," he gave letters to a few hundred randomly selected people in Omaha, Nebraska. The letters were to be sent toward a target person, a stockbroker in Boston, but recipients could send the letters only to someone they knew on a first-name basis. Thirty-five percent of the letters reached their destination and the median number of steps these letters took was 5.5. Rounding up gives "six degrees of separation."

The neat story in the last paragraph becomes a little more dubious if one looks at the details. One third of the test subjects were from Boston, not Omaha, and one-half of those in Omaha were stockbrokers. A large fraction of the letters never reached their destination and were discarded from the distance computation. Of course, those that reached their destination only provide an upper bound on the distance, since there might have been better routes.

Answer 3. Though it was implicit in his work, Milgram never used the phrase "six degrees of separation." John Guare originated the term in the title of his 1991 play. In the play Ousa, musing about our interconnectedness, tells her daughter, "Everybody on the planet is separated by only six other people. Six degrees of separation. Between us and everybody else on this planet. The president of the United States. A gondolier in Venice . . . It's not just the big names. It's anyone. A native in a rain forest. A Tierra del Fuegan. An Eskimo. I am bound to everyone on this planet by a trail of six people. It is a profound thought."

Answer 4. While the Guare play may be the best known literary work with this phrase, it was not the first. It appeared in Hungarian writer Frigyes Karinthy's story Chains. "To demonstrate that people on Earth today are much closer than ever, a

member of the group suggested a test. He offered a bet that we could name any person among the earth's one and a half billion inhabitants and through at most five acquaintances, one of which he knew personally, he could link to the chosen one."

Answer 5. Our final anecdote is a proof by example. A few years ago, the staff of the German newspaper *Die Zeit* accepted the challenge of trying to connect a Turkish kebab-shop owner to his favorite actor Marlon Brando. After a few months of work, they found that the kebab-shop owner had a friend living in California, who works alongside the boyfriend of a woman, who is the sorority sister of the daughter of the producer of the film *Don Juan de Marco*, in which Brando starred.

In the answers we have just given, it sometimes takes fiddling to make the answer six, but it is clear that the web of human contacts and the mathematical collaboration graph have a much smaller diameter than one would naively expect. Albert, Jeong, and Barabási (1999) and Barabási, Albert, and Jeong (2000) studied the World Wide Web graph whose vertices are documents and whose edges are links. Using complete data on the domain nd.edu at his home institution of Notre Dame, and a random sample generated by a web crawl, they estimated that the average distance between vertices scaled with the size of the graph as $0.35 + 2.06 \log n$. Plugging in their estimate of $n = 8 \times 10^8$ web pages at the time they obtained 18.59. That is, two randomly chosen web pages are on the average 19 clicks from each other. The logarithmic dependence of the distance is comforting, because it implies that "if the web grows by a 1000 per cent, web sites would still only be separated by an average of 21 clicks."

Small World Model. Erdös–Rényi graphs have small diameters, but have very few triangles, while in social networks if A and B are friends and A and C are friends, then it is fairly likely that B and C are also friends. To construct a network with small diameter and a positive density of triangles, Watts and Strogatz started from a ring lattice with n vertices and k edges per vertex, and then rewired each edge with probability p, connecting one end to a vertex chosen at random. This construction interpolates between regularity ($p = 0$) and disorder ($p = 1$).

Let $L(p)$ be the average distance between two randomly chosen vertices and define the clustering coefficient $C(p)$ to be the fraction of connections that exist between the $\binom{k}{2}$ pairs of neighbors of a site. The regular graph has $L(0) \sim n/2k$ and $C(0) \approx 3/4$ if k is large, while the disordered one has $L(1) \sim (\log n)/(\log k)$ and $C(1) \sim k/n$. Watts and Strogatz (1998), showed that $L(p)$ decreases quickly near 0, while $C(p)$ changes slowly so there is a broad interval of p over which $L(p)$ is almost as small as $L(1)$, yet $C(p)$ is far from 0. These results will be discussed in Section 5.1.

Watts and Strogatz (1998) were not the first to notice that random long distance connections could drastically reduce the diameter. Bollobás and Chung (1988)

added a random matching to a ring of n vertices with nearest neighbor connections and showed that the resulting graph had diameter $\sim \log_2 n$. This graph, which we will call the *BC small world*, is not a good model of a social network because every individual has exactly three friends including one long-range acquaintance, however these weaknesses make it easier to study.

The small world is connected by definition, so the first quantity we will investigate is the average distance between two randomly chosen sites in the small world. For this problem and all of the others we will consider below, we will not rewire edges but instead consider Newman and Watts (1999) version of the model in which no edges are removed but one adds a Poisson number of shortcuts with mean $n\rho/2$ and attaches then to randomly chosen pairs of sites. This results in a Poisson mean ρ number of long distance edges per site. We will call this the *NW small world*.

Barbour and Reinert (2001) have done a rigorous analysis of the average distance between points in a continuum model in which there is a circle of circumference L and a Poisson mean $L\rho/2$ number of random chords. The chords are the shortcuts and have length 0. The first step in their analysis is to consider an upper bound model that ignores intersections of growing arcs and that assumes each arc sees independent Poisson processes of shortcut endpoints. Let $S(t)$ be size, that is, the Lebesgue measure, of the set of points within distance t of a chosen point and let $M(t)$ be the number of intervals. Under our assumptions

$$S'(t) = 2M(t)$$

while $M(t)$ is a branching process in which there are no deaths and births occur at rate 2ρ.

$M(t)$ is a Yule process run at rate 2ρ so $EM(t) = e^{2\rho t}$ and $M(t)$ has a geometric distribution

$$P(M(t) = k) = (1 - e^{-2\rho t})^{k-1} e^{-2\rho t}$$

Being a branching process $e^{-2\rho t} M(t) \to W$ almost surely. In the case of the Yule process, it is clear from the distribution of $M(t)$, that W has an exponential distribution with mean 1. Integrating gives

$$ES(t) = \int_0^t 2e^{2\rho s}\, ds = \frac{1}{\rho}(e^{2\rho t} - 1)$$

At time $t = (2\rho)^{-1}(1/2)\log(L\rho)$, $ES(t) = (L/\rho)^{1/2} - 1$. Ignoring the -1 we see that if we have two independent clusters run for this time then the expected number of connections between them is

$$\sqrt{\frac{L}{\rho}} \cdot \rho \cdot \frac{\sqrt{L/\rho}}{L} = 1$$