Chapter 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 20^{th} 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 36 people can check their email. 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 if the graph is generated by the preferential attachment model with

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

$m = 2$ 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, e.g., 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, Hooghiemestra, and Znamenski (2005a) 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 (2004), 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 email 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 which 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 1950's 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 $\leq C \log n$ 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, etc. 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. (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 $m = o(\sqrt{n})$, i.e., $m/\sqrt{n} \to 0$, 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, i.e., 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, i.e., 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 (1980) 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 $d_1, \dots d_n$ 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 (2004), 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 $O(1)$, i.e., 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 (2002), 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, i.e., 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 explains in his (2003) 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, i.e., 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 876^{th} 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 co-authors. 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. 35% 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, i.e., 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

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

since the middle factor gives the expected number of shortcuts per unit distance and the last one the probability a short cut will hit the second cluster. The precise result is:

Theorem 5.2.1. *Suppose $L\rho \to \infty$. Let O be a fixed point of the circle, choose P at random, and let D be the distance from O to P. Then*

$$P\left[D > \frac{1}{\rho}\left(\frac{1}{2}\log(L\rho) + x\right)\right] \to \int_0^\infty \frac{e^{-y}}{1 + 2e^{2x}y}\,dy$$

Note that the fluctuations in the distance are of order 1.

Sections 5.3, 5.4, and 5.5 are devoted to a discussion of processes taking place on the small world. We will delay discussion of these results until after we have introduced our next family of examples.

1.4 Power laws, preferential attachment

One of my favorite quotes is from the 13 April 2002 issue of *The Scientist*

> "What do the proteins in our bodies, the Internet, a cool collection of atoms and sexual networks have in common? One man thinks he has the answer and it is going to transform the way we view the world."

Albert-László Barabási (the man in the quote above) and Reka Albert (1999) noticed that the actor collaboration graph and the world wide web had degree distributions that were power laws $p_k \sim Ck^{-\gamma}$ as $k \to \infty$. Follow up work has identified a large number of examples with power law degree distributions, which are also called *scale-free random graphs*. When no reference is given, the information can be found in the survey article by Dorogovstev and Mendes (2002). We omit biological networks (food webs, metabolic networks, and protein interaction networks) since they are much smaller and less well characterized compared to the other examples.

- By the world wide web, we mean the collection of web pages and the oriented links between them. Barabási and Albert (1999) found that the in-degree and out-degrees of web pages follow power laws with $\gamma_{\text{in}} = 2.1$, $\gamma_{\text{out}} = 2.7$.

- By the Internet, we mean the physically connected network of routers that move email and files around the Internet. Routers are united into domains. On the interdomain level the Internet is a small network. In April 1998, when Faloutos, Faloutos, and Faloutos (1999) did their study, there were 3,530 vertices, 6,432 edges and the maximum degree was 745, producing a power law with $\gamma = 2.16$. In 2000 there were about 150,000 routers connected by 200,000 links and a degree distribution that could be fit by a power law with $\gamma = 2.3$.

- The movie actor network in which two actors are connected by an edge if they have appeared in a film together has a power law degree distribution with $\gamma = 2.3$.

- The collaboration graph in a subject is a graph with an edge connecting two people if they have written a paper together. Barabási et al. (2002) studied papers in mathematics and neuroscience published in 1991–1998. The two databases that they used contained 70,901 papers with 70,975 authors, and 210,750 papers with 209,293 authors, respectively. The fitted power laws had $\gamma_{\text{M}} = 2.4$ and $\gamma_{\text{NS}} = 2.1$.

- Newman (2001a,b) studied the collaboration network in four parts of what was then called the Los Alamos preprint archive (and is now called the arXiv). He found that the number of collaborators was better fit by a power law with an exponential cutoff $p_k = Ck^{-\tau} \exp(-k/k_c)$

- The citation network is a directed graph with an edge from i to j if paper i cites paper j. Redner (1998) studied 783,339 papers published in 1981 in journals cataloged by the ISI and 24,296 papers published in volumes 11–50 of Physical Review D. The first graph had 6,716,198 links, maximum degree 8,904, and $\gamma_{\text{in}} = 2.9$. The second had 351,872 links, maximum degree 2,026, and $\gamma_{\text{in}} = 2.6$. In both cases the out degree had an exponentially decaying tail. One reason for the rapid decay of the out-degree is that many journals have a limit on the number of references.

- Liljeros et al. (2001) analyzed data gathered in a study of sexual behavior of 4,781 Swedes, and found that the number of partners per year had $\gamma_{\text{male}} = 3.3$ and $\gamma_{\text{female}} = 3.5$.

- Ebel, Mielsch, and Bornholdt (2002) studied email network of Kiel University, recording the source and destination of every email to or from a student account for 112 days. They found a power law for the degree distribution with $\gamma = 1.81$ and an exponential cutoff at about 100. Recently the Federal Energy Regulatory Commission has made a large email data set available posting 517,341 emails from 151 users at Enron.

To give a mechanistic explanation for power laws Barabási and Albert (1999) introduced the preferential attachment model. For a mental image you can think of a growing world wide web in which new pages are constantly added and they link to existing pages with a probability proportional to their popularity. Suppose, for concreteness, that the process starts at time 1 with two vertices linked by m parallel edges. (We do this so that the total degree at any time t is $2mt$.) At every time $t \geq 2$, we add a new vertex with m edges that link the new vertex to m vertices already present in the system. To incorporate preferential attachment, we assume that the probability π_i that a new vertex will be connected to a vertex i depends on the connectivity of that vertex, so that $\pi_i = k_i / \sum_j k_j$. To be precise, when we add a new vertex we will add edges one a time, with the second and subsequent edges doing preferential attachment using the updated degrees. This scheme has the desirable property that a graph of size n for a general m can be obtained by running the $m = 1$ model for nm steps and then collapsing vertices $km, km-1, \ldots (k-1)m+1$ to make vertex k.

The first thing to be proved for this model, see Theorem 4.1.4, is that the fraction vertices of degree k converges to:

$$p_k \to \frac{2m(m+1)}{k(k+1)(k+2)} \quad \text{as } n \to \infty$$

This distribution $\sim Ck^{-3}$ as $k \to \infty$, so for any value of m we always get a power of 3. Krapivsky, Redner, and Leyvrasz (2000) showed that we can get other behavior for p_k by generalizing the model so that vertices of degree k are chosen with probability proportional to $f(k)$.

- if $f(k) = k^\alpha$ with $\alpha < 1$ then $p_k \approx \mu k^{-\alpha} \exp(-ck^{1-\alpha})$

- if $f(k) = k^\alpha$ with $\alpha > 1$ then the model breaks down: there is one vertex with degree of order $\sim t$ and the other vertices have degree $O(1)$
- if $f(k) = a + k$ and $a > -1$, $p_k \sim Ck^{-(3+a)}$

In the last case we can achieve any power in $(2, \infty)$. However, there are many other means to achieving this end. Cooper and Frieze (2003) describe a very general model in which: old nodes sometimes generate new edges, and choices are sometimes made uniformly instead of by preferential attachment. See Section 4.2 for more details and other references.

Does preferential attachment actually happen in growing networks? Liljeros is quoted in the April 2002 of *The Scientist* as saying "Maybe people become more attractive the more partners they get." Liljeros, Edling, and Amaral (2003) are convinced of the relevance of this mechanism for sexual networks, but Jones and Handcock (2003) and others are skeptical. Jeong, Néda, and Barabási (2001) and Newman (2001c) have used collaboration databases to study the growth of degrees with time. The first paper found support for $\alpha \approx 0.8$ in the actor and neuroscience collaboration networks, while Newman finds $\alpha = 1.04 \pm 0.04$ for Medline and $\alpha = 0.89 \pm 0.09$ for the arXiv, which he argues are roughly compatible with linear preferential attachment. However, as Newman observes, a sublinear power would predict a stretched exponential distribution, which is consistent with his data.

These models and results may look new, but in reality they are quite old. If we think of a vertex i with degree $d(i)$ as $d(i)$ balls of color i, then the $m = 1$ version of the preferential attachment model is just a Polya urn scheme. We pick a ball at random from the urn, return it and a ball of the same color to the urn and add a ball of a new color. For more the connection between preferential attachment and urn schemes, see the second half of Section 4.3.

Yule (1925) used a closely related branching process model for the number of species of a given genus, which produces limiting frequencies $p_k \sim Ck^{-\gamma}$ for any $\gamma > 1$. Simon (1955) introduced a model of word usage in books, where the $(n+1)$th word is new with probability α or is otherwise chosen at random from the previous n words, and hence proportional to their usage. The limiting frequency of words used k times $p_k \sim Ck^{-1+1/(1-\alpha)}$. Again this allows any power in $(2, \infty)$. For more details, see Section 4.2.

The first two sections of Chapter 4 concentrate on the fraction of vertices with a fixed degree k. In Section 4.3 we shift our attention to the other end of the spectrum and look at the growth of the degrees of a fixed vertex j. Mori (2005) has used martigales to study the case $f(k) = k + \beta$ and to show that if M_n is the maximum degree when there are n vertices.

Theorem 4.3.2. *With probability one,* $n^{-1/(2+\beta)} M_n \to \mu$.

Since $\sum_{k=K}^{\infty} p_k \sim CK^{-(2+\beta)}$ this is the behavior we should expect by analogy with maxima of i.i.d. random variables.

Having analyzed the limiting degree distribution in preferential attachment models, we turn our attention now to the distance between two randomly chosen vertices. For simplicity, we consider the fixed degree formulation in which the graph is created in one step rather than grown. When $2 < \gamma < 3$ the size biased distribution $q_k \sim Ck^{-(\gamma-1)}$ so the mean is infinite. Let $\alpha = \gamma - 2$. In the branching process cartoon of cluster growth, the number of vertices at distance m, Z_m grows doubly exponentially fast:

Theorem 4.5.1. $\alpha^m(\log(Z_m + 1)) \to W$.

Intuitively, the limit theorem says $\log(Z_m+1) \approx \alpha^{-m} W$, so replacing Z_m+1 by n, discarding the W and solving gives $m \sim (\log\log n)/(\log(1/\alpha)$. However the right result which van der Hofstadt, Hooghiemestra, and Znamenski (2005a) have proved for the fixed degrees model is that the average distance

$$\sim 2 \cdot \frac{\log\log n}{\log(1/\alpha)}$$

To see the reason for the 2, notice that if we grow clusters from x and y until they have $\sqrt{n}$ members then each process takes time $(\log\log n)/(\log(1/\alpha)$ to reach that size. In Theorem 4.5.2, we prove the upper bound for the corresponding Chung and Lu model.

In the borderline case $\gamma = 3$, Bollobás and Riordan (2004b) have shown for the preferential attachment model, see Theorem 4.6.1, that the diameter $\sim \log n/(\log\log n)$. Chung and Lu have shown for the corresponding case of their model that the distance between two randomly chosen vertices $O(\log n/(\log\log n))$, while the diameter due to dangling ends is $O(\log n)$. To foreshadow later developments, we note that if we want degree distribution function $F(x) = P(d_i \leq x)$ then we can choose the weights in Chung and Lu's model to be

$$w_i = (1 - F)^{-1}(i/n)$$

If $1 - F(x) = Bx^{-\gamma+1}$ solving gives $w_i = (nB/i)^{1/(\gamma-1)}$. Recalling that the probability of an edge from i to j in the Chung and Lu model is $p_{i,j} = w_i w_j / \sum w_k$ and $\sum w_k \sim \mu n$ since the degree distribution has finite mean μ, we see that when $\gamma = 3$

$$p_{i,j} = c/\sqrt{ij}$$

Bollobás and Riordan (2004b) prove their result by relating edge probabilities for the preferential attachment graph to this nonhomogeneous percolation problem. This process will also make an appearance in Section 7.4 in the proof of the Kosterlitz-Thouless transition for the CHKNS model, which has connectivity probabilities $c/(i \wedge j)$. A recent paper of Bollobás, Janson, and Riordan (2006), which is roughly half the length of this book, investigates inhomogenenous random graphs in great detail.

1.5 Epidemics and percolation

The spread of epidemics on random graphs has been studied extensively. There are two extremes: in the first all individuals are susceptible and there is a probability p that an infected individual will transmit the infection to a neighbor, in the second only a fraction p of individuals are susceptible, but the disease is so contagious that if an individual gets infected all of their susceptible neighbors will become infected.

In percolation terms, the first model is bond percolation, where edges are retained with probability p and deleted with probability $1 - p$. The second is site percolation, where the randomness is applied to the sites instead of the edges. Percolation is easy to study on a random graph, since the result of retaining a fraction p of the edges or sites is another random graph. Using the branching process heuristic, percolation occurs (there will be a giant component) if and only if the mean of the associated branching process is > 1. This observation is well known in the epidemic literature, where it is phrased "the epidemic will spread if the number of secondary infections caused by an infected individual is > 1."

When the degree distribution has finite variance, the condition for a supercritical bond percolation epidemic is $E(\hat{D}(\hat{D}-1))/E(\hat{D}) > 1$ where $\hat{D}$ is the number of edges along which the disease will be transmitted, see Section 3.5. Newman (2002) was the first to do this calculation for random graphs with a fixed degree distribution, but incorrectly assumed that the transmission events for different edges are independent, which is false when the duration of the infectious period is random. While the random graph setting for epidemics is new, the associated supercriticality condition is not. May and Anderson (1988) showed that for the transmission of AIDS and other diseases where there is great heterogeneity in the number of secondary infections, k, the basic reproductive number $R_0 = \rho_0(1 + C_V^2)$ where $\rho_0 =< k >$ is the average number of secondary infections, $C_V = (< k^2 > / < k >^2) - 1$ is the coefficient of variation of the connectivity distribution, and $< X >$ is physicist's notation for expected value EX.

As noted in the previous section, many networks have power law degree distributions with power $2 < \gamma < 3$. In this case the sized biased distribution q has infinite mean. Thus for any $p > 0$ the mean number of secondary contacts is > 1 and the critical value for percolation $p_c = 0$. This "surprising result" has generated a lot of press since it implies that "within the observed topology of the Internet and the www, viruses can spread even when the infection probabilities are vanishingly small."

This quote is from Lloyd and May's (2001) discussion in *Science* of Pastor-Satorras and Vespignani (2001). This dire prediction applies not only to computers but also to sexually transmitted diseases. "Sexual partnership networks are often extremely heterogeneous because a few individuals (such as prostitutes) have very high numbers of partners. Pastor-Satorras and Vespignani's results may be of relevance in this context. This study highlights the potential importance of studies on communication and other networks, especially those with scale-free and small world properties, for those seeking to manage epidemics within human and animal populations." Fortunately for the people of Sweden, $\gamma_{\text{male}} = 3.3$ and $\gamma_{\text{female}} = 3.5$, so sexually transmitted diseases have a positive epidemic threshold.

Dezsö and Barabási (2002) continue this theme in their work: "From a theoretical perspective viruses spreading on a scale free network appear unstoppable. The question is, can we take advantage of the increased knowledge accumulated in the past few years about the network topology to understand the conditions in which one can successfully eradicate the viruses?" The solution they propose is obvious. The vanishing threshold is a consequence of the nodes of high degree, so curing the "hubs" is a cost-effective method for combating the epidemic. As Liljeros et al. (2001) say in the subhead of their paper: "promiscuous individuals are the vulnerable nodes to target in safe-sex campaigns." For more on virus control strategies for technological networks, see Balthrop, Forrest, Newman, and Williamson (2004). The SARS epidemic with its superspreaders is another situation where the highly variable number of transmissions per individual calls for us to rethink our approaches to preventing the spread of disease, see e.g., Lloyd-Smith, Schreiber, Kopp and Getz (2005).

One of the most cited properties of scale-free networks, which is related to our discussion of epidemics, is that they are "robust to random damage but vulnerable to malicious attack." Albert, Jeong, and Barabási (2000) performed simulation studies on the result of attacks on a map of the Internet consisting of 6,209 vertices and 24,401 links. Their simulations and some approximate calculations suggested that 95% of the links can be removed and the graph will stay connected. Callaway, Newman, Strogatz, and Watts (2000) modeled intentional damage as removal of the vertices with degrees $k > k_0$, where k_0 is chosen so that the desired fraction of vertices f is eliminated. They computed threshold values for the distribution $p_k = k^{-\gamma}/\zeta(\gamma)$ when $\gamma = 2.4, 2.7, 3.0$. Here ζ is Riemann's function, which in this context plays the mundane role of giving the correct normalization to produce a probability distribution. The values of f_c in the three cases are 0.023, 0.010, and 0.002, so using the first figure for the Internet, the targeted destruction of 2.3% of the hubs would disconnect the Internet.

The results in the last paragraph are shocking, which is why they attracted headlines. However, as we mentioned earlier, one must be cautious in interpreting them. Bollobás and Riordan (2004c) have done a rigorous analysis of percolation on the Barbási-Albert preferential attachment graph, which has $\beta = 3$. In the case $m = 1$ the world is a tree and destroying any positive fraction of the edges disconnects it.

Theorem 4.7.3. *Let $m \geq 2$ be fixed. For $0 < p \leq 1$ there is a function*

$$\exp(-C/p^2) \leq \lambda(p) \leq \exp(-c/p)$$

so that with probability $1 - o(1)$ the size of largest component is $(\lambda(p) + o(1))n$ and the second largest is $o(n)$.

Heuristic results presented in Section 4.7 suggest that the upper bound is the right answer and for the concrete case considered above $c = 1/\zeta(3)$. In words, if p is small then the giant component is tiny, and it is unlikely you will be able to access the Internet from your house. Using $\zeta(3) = 1.202057$ and setting $p = 0.05$ gives the result quoted in the first section of this introduction that the fraction of nodes in the giant component in this situation is 5.9×10^{-8}

Bollobás and Riordan (2004c) have also done a rigorous analysis of intentional damage for the preferential attachment model, which they define as removal of the first nf nodes, which are the ones likely to have the largest degrees.

Theorem 4.7.4. *Let $m \geq 2$ and $0 < f < 1$ be constant. If $f \geq (m-1)/(m+1)$ then with probability $1 - o(1)$ the largest component is $o(n)$. If $f < (m-1)/(m+1)$ then there is a constant $\theta(f)$ so that with probability $1 - o(1)$ the largest component is $\sim \theta(f)n$, and the second largest is $o(n)$.*

It is difficult to compare this with the conclusions of Callaway, Newman, Strogatz, and Watts (2000) since for any m in the preferential attachment model we have $p_k \sim 2m(m+1)k^{-3}$ as $k \to \infty$. However, the reader should note that even when $m = 2$, one can remove 1/3 of the nodes.

One of the reasons why CNSW get such small number is that, as Aiello, Chung, and Lu (2000,2001) have shown, graphs with degree distribution $p_k = k^{-\gamma}/\zeta(\gamma)$ have no giant component for $\gamma > 3.479$. Thus the fragility is an artifact of assuming that there is an exact power law, while in reality the actual answer for graphs with $p_k \sim Ck^{-\gamma}$ depends on the value of C as well. This is just one of many criticisms of the claim that the Internet is “robust yet fragile.” Doyle et al. (2005) examine in detail how the scale free depiction of compares with the real Internet.

Percolation on the small world is studied in Section 5.3. Those results are the key to the ones in the next section.

1.6 Potts models and the contact process

In the Potts model, each vertex is assigned a spin σ_x which may take one of q values. Given a finite graph G with vertices V and edges E, e.g., the small world, the energy of a configuration is

$$H(\sigma) = 2 \sum_{x,y \in V, x \sim y} 1\{\sigma(x) \neq \sigma(y)\}$$

where $x \sim y$ means x is adjacent to y. Configurations are assigned probabilities $\exp(-\beta H(\sigma))$, where β is a variable inversely proportional to temperature. We define a probability measure on $\{1, 2, \dots q\}^V$ by

$$\nu(\sigma) = Z^{-1} \exp(-\beta H(\sigma))$$

where Z is a normalizing constant that makes the $\nu(\sigma)$ sum to 1. When $q = 2$ this is the Ising model, though in that case it is customary to replace $\{1, 2\}$ by $\{-1, 1\}$, and write the energy as

$$H_2(\sigma) = - \sum_{x,y \in V, x \sim y} \sigma(x)\sigma(y)$$

This leads to the same definition of ν since every pair with $\sigma(x) \neq \sigma(y)$ increases H_2 by 2 from its minimum value in which all the spins are equal, so $H - H_2$ is constant and after normalization the measures are equal.

To study the Potts model on the small world, we will use the random-cluster model of Fortuin and Kastelyn. This is a $\{0, 1\}$-valued process η on the edges E of the graph:

$$\mu(\eta) = Z^{-1} \left\{ \prod_{e \in E} p^{\eta(e)} (1-p)^{1-\eta(e)} \right\} q^{\chi(\eta)}$$

where $\chi(\eta)$ is the number of connected components of η when we interpret 1-bonds as occupied and 0-bonds as vacant and Z is another normalizing constant.

Having introduced the model with a general q, we will now restrict our attention to the Ising model with $q = 2$. By using some comparison arguments we are able to show that on a general graph the Ising model has long range order for $\beta > \beta_I$ where $\tanh(\beta_I) = p_c$, the threshold for percolation in the model with independent bonds. See Theorem 5.4.5. To explain the significance of this equation, consider the Ising model on a tree with forward branching number $b \geq 2$. The critical value for the onset of "spontaneous magnetization" has $\tanh(\beta_c) = 1/b$. This means that when $\beta > \beta_c$ if we impose $+1$ boundary conditions at sites a distance n from the root and let $n \to \infty$ then in the resulting limit spins $\sigma(x)$ have positive expected value. When $0 \leq \beta \leq \beta_c$ there is a unique limiting Gibbs state independent of the boundary conditions. See e.g., Preston (1974).

This connection allows us to show in Section 5.4 that for BC small world, which looks locally like a tree of degree 3, the Ising model critical value is

$$\beta_I = \tanh^{-1}(1/2) = 0.5493$$

$1/\beta_I = 1.820$ agrees with the critical value of the temperature found from simulations of Hong, Kim, and Choi (2002), but physicists seem unaware of this simple exact result. Using results of Lyons (1989, 1990) who defined a branching number for trees that are not regular, we are able to extend the argument to the nearest neighbor NW small world, which locally like a two type branching process.

In making the connection with percolation, we have implicitly been considering the SIR (susceptible-infected-removed) epidemic model in which sites, after being infected, become removed from further possible infection. This is the situation for many diseases, such as measles, and would seem to be reasonable for computers whose anti-virus software has been updated to recognize the virus. However, Pastor-Satorras and Vespignani (2001a, 2001b, 2002) and others have also considered the SIS (susceptible-infected-susceptible) in which sites that have been cured of the infection are susceptible to reinfection. We formulate the model in continuous time with infected sites becoming healthy (and again susceptible) at rate 1, while an infected site infects each of its susceptible neighbors at rate λ. In the probability literature, this SIS model is called Harris' (1974) contact process. There it usually takes place on a regular lattice like $\mathbb{Z}^2$ and is more often thought of as a model for the spread of a plant species.

The possibility of reinfection in the SIS model allows for an endemic equilibrium in which the disease persists infecting a positive fraction of the population. Since the graph is finite the infection will eventually die out, but as we will see later, there is a critical value λ_c of the infection rate the disease persists for an extremely long time. Pastor-Satorras and Vespigniani have made an extensive study of this model using mean-field methods. To explain what this means, let $\rho_k(t)$ denote the fraction of vertices of degree k that are infected at time t, and $\theta(\lambda)$ be the probability that a given link points to an infected site. If we make the mean-field assumption that there are no correlations then

$$\frac{d}{dt}\rho_k(t) = -\rho_k(t) + \lambda k[1 - \rho_k(t)]\theta(\lambda)$$

Analysis of this equation suggests the following conjectures about the SIS model on power law graph with degree distribution $p_k \sim Ck^{-\gamma}$.

- If $\gamma \leq 3$ then $\lambda_c = 0$.
- If $3 < \gamma < 4$, $\lambda_c > 0$ but $\theta(\lambda) \sim C(\lambda - \lambda_c)^{1/(\gamma-3)}$ as $\lambda \downarrow \lambda_c$.
- If $\beta > 4$ then $\lambda_c > 0$ and $\theta(\lambda) \sim C(\lambda - \lambda_c)$ as $\lambda \downarrow \lambda_c$.

The second and third claims are interesting open problems. Berger, Borgs, Chayes, and Saberi (2004) have considered the contact process on the Barbási-Albert preferential attachment graph. They have shown that $\lambda_c = 0$ and proved some interesting results about the probability the process will survive from a randomly chosen site. The proof of $\lambda_c = 0$ is very easy and is based on the following

Lemma 4.8.2. *Let G be a star graph with center 0 and leaves $1, 2, \dots k$. Let A_t be the set of vertices infected in the contact process at time t when $A_0 = \{0\}$. If $k\lambda^2 \to \infty$ then*

$$P(A_{\exp(k\lambda^2/10)} \neq \emptyset) \to 1$$

The largest degree in the preferential attachment graph is $O(n^{1/2})$ so if $\lambda > 0$ is fixed the process will survive for time at least $\exp(cn^{1/2})$.

At this point we have considered the Ising model on the small world, and the contact process on graphs with a power law degree distribution. The other two combinations have also been studied. Durrett and Jung (2005) have considered the contact process on a generalization of the BC small world and showed that like the contact process on trees the system has two phase transitions, see Section 5.5 for details.

Dorogovstev, Goltsev, and Mendes (2002) have studied the Ising model on power law graphs. Their calculations suggest that $\beta_c = 0$ for $\gamma \leq 3$, the spontaneous magnetization $M(\beta) \sim (\beta - \beta_c)^{1/(\gamma-3)}$ for $3 < \gamma < 5$ while for $\gamma > 5$, $M(\beta) \sim (\beta - \beta_c)^{1/2}$. A rigorous proof of the results for critical exponents seems difficult, but can one use the connection between the Ising model and percolation to show $\beta_c = 0$ for $\gamma_c \leq 3$?

1.7 Random walks and voter models

There have been quite a few papers written about the properties of random walks on small world networks studying the probability the walker is back where it started after n steps, the average number of sites visited, etc. See for example, Monasson (1999), Jespersen, Sokolov, and Blumen (2000), Lahtinen, Kertesz, and Kaski (2001), Pandit and Amritkar (2001), Almaas, Kulkarni, and Stroud (2003), and Noh and Reiger (2004). In most cases the authors have concentrated on the situation in which the density of shortcuts p is small, and shown that for small times $t \ll \xi^2$ with $\xi = 1/p$ the behavior is like a random walk in one dimension, at intermediate times the behavior is like a random walk on a tree, and at large times the walker realizes it is on a finite set.

Here, we will concentrate instead on the rate of convergence to equilibrium for random walks. Let $K(x, y)$ be the transition kernel of the lazy walk that stays put with probability 1/2 and otherwise jumps to a randomly chosen neighbor. The laziness gets rid of problems with periodicity and negative eigenvalues. Let $\pi(x)$ be its stationary distribution, and define $Q(x, y) = \pi(x)K(x, y)$ to be the flow from x to y in equilibrium. Our walks satisfy the detailed balance condition, i.e., $Q(x, y) = Q(y, x)$.

In most cases we will bound the rate of convergence to equilibrium by considering the conductance

$$h = \min_{\pi(S) \le 1/2} \frac{Q(S, S^c)}{\pi(S)}$$

where $Q(S, S^c) = \sum_{x \in S, y \in S^c} Q(x, y)$. If all of the vertices have the same degree d and we let $e(S, S^c)$ be the number of edges between S and S^c, $h = \iota/2d$ where

$$\iota = \min_{|S| \le n/2} \frac{e(S, S^c)}{|S|}$$

is the *edge isoperimetric constant*.

Cheeger's inequality and standard results about Markov chains, see Sections 6.1 and 6.2, imply that if h is bounded away from 0 as the size of the graph n tends to ∞, then convergence to equilibrium takes time that is $O(\log n)$. This result takes care of most of our examples. Bollobás (1988) estimated the isoperimetric constant for random regular graphs, the special case of a fixed degree distribution with $p_r = 1$ for some $r \ge 3$. In Section 6.3, we will prove a more general result with a worse constant due to Gkantsidis, Mihail, and Saberi (2003)

Theorem 6.3.2. *Consider a random graph with a fixed degree distribution in which the minimum degree is $r \ge 3$. There is a constant $\alpha_0 > 0$ so that $h \ge \alpha_0$.*

In Sections 6.4 and 6.5 we will show that the same conclusion holds for Barabási and Albert's preferential attachment graph, a result of Mihail, Papadimitrou, and Saberi (2004), and for connected Erdös-Rényi random graphs $ER(n, (c \log n)/n)$ with $c > 1$, a result of Cooper and Frieze (2003).

It is easy to show that the random walk on the BC small world in which each vertex has degree 3, mixes in time $O(\log n)$. In contrast the random walk on the NW small world mixes in time at least $O(\log^2 n)$ and at most $O(\log^3 n)$. The lower bound is easy to see and applies to any graph with a fixed degree distribution with $p_2 > 0$. There are paths of length $O(\log n)$ in which each vertex has degree 2. The time to escape from this interval starting from the middle is $O(\log^2 n)$ which gives a lower bound of $O(\log^2 n)$ on the mixing time. The upper bound comes from showing that the conductance $h \geq C/\log n$, which translates into a bound of order $\log^3 n$. We believe that the lower bound is the right order of magnitude. In Section 6.7 we prove this is correct for a graph with a fixed degree distribution in which $p_2 + p_3 = 1$.

The voter model is a very simple model for the spread of an opinion. On any of our random graphs it can be defined as follows. Each site x has an opinion $\xi_t(x)$ and at the times of a rate 1 Poisson process decides to change its opinion. To do this it picks a neighbor at random and adopts the opinion of that neighbor. If you don't like this simple minded sociological interpretation, you can think instead of this as a spatial version of the Moran model of population genetics.

To analyze the voter model we use a "dual process" $\zeta_s^{x,t}$ that works backwards in time to determine the source of the opinion at x at time t and jumps if voter at $\zeta_s^{x,t}$ at time $t - s$ imitated one of its neighbors. The genealogy of one opinion is a random walk. If we consider several at once we get a coalescing random walk since $\zeta_s^{x,t} = \zeta_s^{t,y}$ implies that the two processes will agree at all later times.

If we pick the starting points x and y according to the stationary distribution π for the random walk and let T_A be the time at which they first hit, Proposition 23 of Aldous and Fill (2002) implies

$$\sup_t |P_\pi(T_A > t) - \exp(-t/E_\pi T_A)| \leq \tau_2/E_\pi T_A$$

where τ_2 is the relaxation time, which they define (see p. 19) to be 1 over the spectral gap. In many of our examples $\tau_2 \leq C \log^2 n$ and as we will see $E_\pi T_A \sim cn$ so the hitting time is approximately exponential. To be precise, we will show this in Section 6.9 for the BC and NW small worlds, fixed degree distributions with finite variance, and connected Erdös-Rényi random graphs.

Holley and Liggett (1975) showed that on the d-dimensional lattice $\mathbb{Z}^d$, if $d \leq 2$ the voter model approaches complete consensus, i.e., $P(\xi_t(x) = \xi_t(y)) \to 1$, while if $d \geq 3$ and we start from product measure with density p (i.e., we assign opinions 1 and 0 independently to sites with probabilities p and $1 - p$) then as $t \to \infty$, ξ_t^p converges in distribution to ξ_∞^p, a one parameter family of stationary distributions.

On a finite set the voter model will eventually reach an absorbing state in which all voters have the same opinion. Cox (1989) studied the voter model on a finite torus $(Z \bmod N)^d$ and showed that if $p \in (0,1)$ then the time to reach consensus τ_N satisfies $\tau_N = O(s_N)$ where $s_N = N^2$ in $d = 1$, $s_N = N^2 \log N$ in $d = 2$, and $s_N = N^d$ in $d \geq 3$. Our results for the voter model on the BC or NW small worlds show that while the world starts out one dimensional, the long range connections make the behavior like that of a voter model in $d \geq 3$, where the time to reach the absorbing state is proportional to the volume of the system. Before reaching the absorbing state the voter model settles into a quasi-stationary distribution, which is like the equilibrium state in dimensions $d \geq 3$. Castelano, Vilone, and Vespigiani (2003) arrived at these conclusions on the basis of simulation.

While the voter model we have studied is natural, there is another one with nicer properties. Consider the voter model defined by picking an edge at random from the graph, flipping a coin to decide on an orientation (x, y), and then telling the voter at y to imitate the voter at x. The random walk in this version of the voter model has a uniform stationary distribution and in the words of Suchecki, Eguuíluz and Miguel (2004): "conservation of the global magnetization." In terms more familiar to probabilists, the number of voters with a given opinion is a time change of simple random walk and hence is a martingale. If we consider the biased voter model in which changes from 0 to 1 are always accepted but changes from 1 to 0 occur with probability $\lambda < 1$, then the last argument shows that the fixation probability for a single 1 introduced in a sea of 0's does not depend on the structure of the graph, the small world version of a result of Maruyama (1970) and Slatkin (1981). Because of this property, Lieberman, Hauert, and Nowak (2005), who studied evolutionary dynamics on general graphs, call the random walk *isothermal*.

1.8 CHKNS model

Inspired by Barabási and Albert (1999), Callaway, Hopcroft, Kleinberg, Newman, and Strogatz (2001) introduced the following simple version of a randomly grown graph. Start with $G_1 = \{1\}$ with no edges. At each time $n \geq 2$, we add one vertex and with probability δ add one edge between two randomly chosen vertices. Note that the newly added vertex is not necessarily an endpoint of the added edge and when n is large, it is likely not to be.

In the original CHKNS model, which we will call model #0, the number of edges was 1 with probability δ, and 0 otherwise. To obtain a model that we can analyze rigorously, we will study the situation in which a Poisson mean δ number of vertices are added at each step. We prefer this version since, in the Poisson case, if we let $A_{i,j,k}$ be the event no (i,j) edge is added at time k then $P(A_{i,j,k}) = \exp\left(-\delta/\binom{k}{2}\right)$ for $i < j \leq k$ and these events are independent.

$$
\begin{aligned}
P(\cap_{k=j}^{n} A_{i,j,k}) &= \prod_{k=j}^{n} \exp\left(-\frac{2\delta}{k(k-1)}\right) \\
&= \exp\left(-2\delta\left(\frac{1}{j-1} - \frac{1}{n}\right)\right) \geq 1 - 2\delta\left(\frac{1}{j-1} - \frac{1}{n}\right) \qquad \#1
\end{aligned}
$$

The last formula is somewhat ugly, so we will also consider two approximations

$$
\begin{aligned}
&\approx 1 - 2\delta\left(\frac{1}{j} - \frac{1}{n}\right) \qquad \#2 \\
&\approx 1 - \frac{2\delta}{j} \qquad \#3
\end{aligned}
$$

The approximation that leads to #3 is not as innocent as it looks. If we let $\mathcal{E}_n$ be the number of edges then using the definition of the model $E\mathcal{E}_n \sim \delta n$ in models #1 and #2 but $E\mathcal{E}_n \sim 2\delta n$ in model #3. Despite this, it turns out that models #1, #2, and #3 have the same qualitative behavior, so in the long run we will concentrate on #3.

The first task is to calculate the critical value for the existence of a giant component. CHKNS showed that the generating function $g(x)$ of the size of the component containing a randomly chosen site satisfied

$$g'(x) = \frac{1}{2\delta x} \cdot \frac{x - g(x)}{1 - g(x)}$$

and used this to conclude that if $g(1) = 1$ then the mean cluster size $g'(1) = (1 - \sqrt{1 - 8\delta})/4\delta$. Since this quantity becomes complex for $\delta > 1/8$ they concluded $\delta_c = 1/8$. See Section 7.1 for a more complete description of their argument. One may quibble with the proof but the answer is right. As we will prove in Section 7.2, in models #1, #2, or #3 the critical value $\delta_c = 1/8$.

In contrast to the situation with ordinary percolation on the square lattice where Kesten (1980) proved the physicists' answer was correct nearly twenty year after they had guessed it, this time the rigorous answer predates the question by more than ten years. We begin by describing earlier work on the random graph model on $\{1, 2, 3, \ldots\}$ with $p_{i,j} = \lambda/(i \vee j)$. Kalikow and Weiss (1988) showed that the probability G is connected (ALL vertices in ONE component) is either 0 or 1, and that $1/4 \le \lambda_c \le 1$. They conjectured $\lambda_c = 1$ but Shepp (1989) proved $\lambda_c = 1/4$. To connect with the answer $\delta_c = 1/8$, note that $\lambda = 2\delta$. Durrett and Kesten (1990) proved a result for a general class of $p_{i,j} = h(i, j)$ that are homogeneous of degree -1, i.e., $h(ci, cj) = c^{-1}h(i, j)$. It is their methods that we will use to prove the result.

To investigate the size of the giant component, CHKNS integrated the differential equation for the generating function g near $\delta = 1/8$. Letting $S(\delta) = 1 - g(1)$ the fraction of vertices in the infinite component they plotted $\log(-\log S))$ vs $\log(\delta - 1/8)$ and concluded that

$$S(\delta) \sim \exp(-\alpha(\delta - 1/8)^{-\beta})$$

where $\alpha = 1.132 \pm 0.008$ and $\beta = 0.499 \pm 0.001$. Based on this they conjectured that $\beta = 1/2$. Note that, in contrast to the many examples we have seen previously where $S(\delta) \sim C(\delta - \delta_c)$ as $\delta \downarrow \delta_c$, the size of the giant component is infinitely differentiable at the critical value. In the language of physics we have a Kosterlitz-Thouless transition. If you are Russian you add Berezinskii's name at the beginning of the name of the transition.

Inspired by CHKNS' conjecture Dorogovstev, Mendes, and Samukhin (2001) computed that as $\delta \downarrow 1/8$,

$$S \equiv 1 - g(1) \approx c \exp(-\pi/\sqrt{8\delta - 1})$$

To compare with the numerical result we note that $\pi/\sqrt{8} = 1.1107$. To derive their formula DMS change variables $u(\xi) = 1 - g(1 - \xi)$ in the differential equation for g to get

$$u'(\xi) = \frac{1}{2\delta(1-\xi)} \cdot \frac{u(\xi) - \xi}{u(\xi)}$$

They discard the $1 - \xi$ in the denominator (without any justification or apparent guilt at doing so), solve the differential equation explicitly and then do some asymptotic analysis of the generating function, which one can probably make rigorous. The real mystery is why can you drop the $1 - \xi$?

Again one may not believe the proof, but the result is correct. Bollobás, Janson, and Riordan (2005) have shown, see Theorem 7.4.1, that if $\eta > 0$ then

$$S(\delta) \le \exp(-(1 - \eta)/\sqrt{8\delta - 1})$$

when $\delta - \delta_c > 0$ is small, and they have proved a similar lower bound. Their proof relates the percolation process in the random graph in which i is connected to j with probability $c/(i \vee j)$ to the one in which the probability is $c/\sqrt{ij}$. The latter process played a role in the analysis of the diameter of the preferential attachment model in Section 4.6.

In addition to the striking behavior of the size of the giant component, the behavior of the cluster size distribution is interesting at the critical point and in the subcritical regime.

As we show in Section 7.3 for models #0 or #1, at $\delta = 1/8$ the probability a randomly chosen site belongs to a cluster of size k, $b_k \sim 1/(k \log k)^2$, when $\delta < 1/8$

$$b_k \sim C_\delta k^{-2/(1-\sqrt{1-8\delta})}$$

Our results on the probability of $i \to j$, i.e., a path from i to j, are not as good. We are able to show, see Section 7.5 for model #3, that for $\delta = 1/8$, and $1 \le i < j \le n$,

$$P(i \to j) \le \frac{3}{8}\Gamma^n_{i,j} \quad \text{where} \quad \Gamma^n_{i,j} = \frac{(\log i + 2)(\log n - \log j + 2)}{(\log n + 4)}.$$

and that this implies

$$\frac{1}{n}\sum_{i=1}^n E|\mathcal{C}_i| \le 6$$

so the mean cluster size is finite at the critical value. However, we are not able to prove that $P(i \to j) \ge c\Gamma_{i,j}$ when i is close to 1 and j is close to n. In the subcritical regime, we prove in Section 7.3 for model #3 that if $i < j$ then

$$P(i \to j) \le \frac{c}{2ri^{1/2-r}j^{1/2+r}}$$

where $r = \sqrt{1-8\delta}/2$. This upper bound is an important ingredient in the proof of the Bollobás, Janson, and Riordan (2005) result in Section 7.4, but in view of our difficulties when $\delta = 1/8$, we have not investigated lower bounds.

.