The phase transition in the uniformly grown random graph has infinite order*

Béla Bollobás[†‡§] Svante Janson[¶] Oliver Riordan[†‖]

September 30, 2003; revised July 28, 2004

Abstract

The aim of this paper is to study the emergence of the giant component in the uniformly grown random graph $G_n(c)$, $0 < c < 1$, the graph on the set $[n] = \{1, 2, \ldots, n\}$ in which each possible edge ij is present with probability $c/\max\{i, j\}$, independently of all other edges. Equivalently, we may start with the random graph $G_n(1)$ with vertex set $[n]$, where each vertex j is joined to each 'earlier' vertex $i < j$ with probability $1/j$, independently of all other choices. The graph $G_n(c)$ is formed by the open bonds in the bond percolation on $G_n(1)$ in which a bond is open with probability c.

The model $G_n(c)$ is the finite version of a model proposed by Dubins in 1984, and is also closely related to a random graph process defined by Callaway, Hopcroft, Kleinberg, Newman and Strogatz [8].

Results of Kalikow and Weiss [15] and Shepp [19] imply that the percolation threshold is at $c = 1/4$. The main result of this paper is that for $c = 1/4 + \varepsilon$, $\varepsilon > 0$, the giant component in $G_n(c)$ has order

$$\exp\left(-\Theta(1/\sqrt{\varepsilon})\right) n.$$

In particular, the phase transition in the bond percolation on $G_n(1)$ has infinite order. Using non-rigorous methods, Dorogovtsev, Mendes and Samukhin [9] showed that an even more precise result is likely to hold.

1 Introduction

The emergence of a giant component is one of the most frequently studied phenomena in the theory of random graphs. Much of the interest is due to the fact

*This is a preprint of an article accepted for publication in Random Structure & Algoritms

†Department of Mathematical Sciences, University of Memphis, Memphis TN 38152, USA

‡Trinity College, Cambridge CB2 1TQ, UK

§Research supported by NSF grant ITR 0225610 and DARPA grant F33615-01-C-1900

¶Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden

‖Royal Society Research Fellow, Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, UK

that a giant component in a finite graph corresponds to an infinite component, or 'infinite cluster', in percolation on an infinite graph. In fact, it can be argued that it is more important and more difficult to study detailed properties of the emergence of the giant component than to study the corresponding infinite percolation near the critical probability.

The quintessential example of the emergence of a giant component is in the classical random graph model $G_{n,p}$, the graph with vertex set $\{1, 2, \ldots, n\}$ in which each pair of vertices is joined with probability p, independently of all other pairs. Let us say that an event holds *with high probability* (**whp**), if it holds with probability tending to 1 as $n \to \infty$. In 1960, Erdős and Rényi [12, 13] showed that the critical probability for $G_{n,p}$ is $1/n$: if $c < 1$ is a constant then **whp** the largest component of $G_{n,c/n}$ has $O(\log n)$ vertices, while there is a function $\theta(c) > 0$ such that for constant $c > 1$, **whp** $G_{n,c/n}$ has a component of order $(\theta(c)+o(1))n$, and no other component of order larger than $O(\log n)$. The proper 'window' of the phase transition was found much later by Bollobás [3] and Łuczak [17]; for numerous other detailed and deep related results see [16, 14, 4]. In $G_{n,c/n}$ the giant component emerges rather rapidly: the right-derivative of $\theta(c)$ at $c = 1$ is 2; this makes the study of the phenomenon manageable. For a discussion of this see [2].

Our task in this paper is considerably harder, since in the model we shall study the giant component emerges much more slowly. Our model, $G_n(c)$, is the finite version of a model first proposed by Dubins in 1984 (see [15, 19]), see below for details: it is parametrized by n, the number of vertices, and a constant $c > 0$ to which edge probabilities are proportional, just as for $G_{n,c/n}$. It can be read out of results of Kalikow and Weiss [15] and Shepp [19] that there is a critical value $c = 1/4$ above which a giant component is present. In $G_n(c)$, the transition from having no giant component to having a giant component is rather tantalizing, since it is very slow indeed. As we shall see, for any c less than $1/4$, **whp** the largest component of $G_n(c)$ already contains $n^{\Theta(1)}$ vertices, which is *much* larger than the $O(\log n)$ we have in $G_{n,a/n}$, $a < 1$. For $c > 1/4$, **whp** there is a giant component of order proportional to n, and the other components are small. In fact, it is more than likely that there is a function $\phi(c)$, equal to 0 for $c \leq 1/4$ but positive for $c > 1/4$, such that **whp** the largest component of $G_n(c)$ has order $(\phi(c) + o(1))n$. However, rather than having positive right-derivative at the critical point, in this case every derivative of $\phi(c)$ at $c = 1/4$ is zero. This phenomenon is often called a phase transition of *infinite order*. Somewhat surprisingly, in spite of this extremely gentle growth of the giant component, we shall be able to give good bounds on $\phi(c)$ from above and below, showing, in particular, that $\phi^{(k)}(1/4) = 0$ for every k.

A somewhat similar, although less surprising, phenomenon was studied in [6], where for a different model it was shown that for *every* positive value of the appropriate parameter c there is a giant component, but its normalized size has all derivatives zero at $c = 0$. Nevertheless, a gentle increase at the very beginning is considerably less difficult to handle than a 'sudden' gentle increase in a function which is zero up to some positive value.

Turning to the model, in [8], Callaway, Hopcroft, Kleinberg, Newman and

Strogatz introduced a simple new model (which we shall call the CHKNS model) for random graphs growing in time. They gave heuristic arguments to find the critical point for the percolation phase transition in this graph, and numerical results (from integrating an equation, rather than just simulating the graph) to suggest that this transition has infinite order. Durrett [10] has recently proved, among other results, that the critical probability they obtain is correct (we return to this later). Non-rigorous arguments for an infinite order phase transition in this and other models have been given by Dorogovtsev, Mendes and Samukhin [9].

Here we consider an even simpler and more natural model, the *uniformly grown random graph*, or '$1/j$*-graph*'. This is the finite version of a model proposed by Dubins in 1984. We define the $1/j$-graph $G_n^{1/j}$ as the random graph on $\{1, 2, \ldots, n\}$ in which each pair $i < j$ of vertices is joined independently with probability $1/j$. We may think of $G_n = G_n^{1/j}$ as a graph growing in time, where each vertex joins to a set of earlier vertices chosen uniformly at random, the set itself having a random size, which is essentially Poisson with mean 1. We study the random subgraph $G_n(c)$ of G_n obtained by selecting edges independently with probability $c < 1$. Of course, $G_n(c)$ can be defined directly by specifying that each pair $i < j$ is joined independently with probability c/j. With this definition, values of c greater than one make sense, provided we replace c/j by $\min\{c/j, 1\}$. Since we are interested in behaviour below, at, or just above the critical value $c = 1/4$, we shall not consider $c > 1$ here.

Kalikow and Weiss [15] showed that for $c < 1/4$ the infinite version $G_\infty(c)$ of $G_n(c)$ is disconnected with probability one. It is implicit in their work that **whp** the largest component in the finite graph $G_n(c)$, $c < 1/4$, has order $o(n)$. In the other direction, Shepp [19] showed that for $c > 1/4$, $G_\infty(c)$ is connected with probability 1; his proof involved showing that $G_n(c)$ has a component of order $\Theta(n)$ with probability bounded away from zero. Hence, the threshold for the emergence of a giant component in $G_n(c)$ is at $c = 1/4$. A similar result for a more general model was proved by Durrett and Kesten [11].

Here we study the size of the giant component above the threshold, showing that the giant component emerges very slowly, having size $\exp(-\Theta(\varepsilon^{-1/2}))n$ as $n \to \infty$ with $c = 1/4 + \varepsilon$. Thus the phase transition is of 'infinite order'. Our method involves counting paths in the *subcritical* graph (with $c < 1/4$). We shall comment on related results of Shepp [19], Durrett and Kesten [11], Zhang [24] and Durrett [10] in the body of the paper.

Although we work throughout in terms of the $1/j$-graph, as it has a simpler and more natural static description, we shall show at the end that all our results carry over to the CHKNS model. This is also true of the earlier threshold results, which predate the CHKNS model by more than 10 years!

The threshold for emergence, and even the size, of the giant component has been studied for various inhomogeneous random graphs by several people. For example, Söderberg [20, 21, 22] considers a general model with finitely many types of vertices, giving formulae for the critical point and giant component size without proof. This model does not cover cases such as that studied here.

Other specific models are the LCD graph of [5], or the Barabási-Albert model of [1]; see [6] for results about these models as well as further references. Yet another growing graph model was introduced by Turova [23]. Although we shall not consider this model here, it may well be that our methods are applicable to it.

1.1 Relationship to earlier rigorous work

We shall make frequent reference to the work of Durrett [10] on the CHKNS model. The relationship of our work to his is as follows. The existence of Durrett's work was public before our work started (see [18]). Durrett's paper [10] was published, and we became aware of his work, after the first draft of this paper was completed.

Using non-rigorous methods, Dorogovtsev, Mendes and Samukhin [9] obtained an extremely precise formula for the size of the giant component in the CHKNS model, which would imply our main results, namely Theorems 3 and 4 (see Section 11). However, as rigorous mathematical theorems, these results are new. Certain of our intermediate results were proved in different forms in [10], and some in even earlier work. In some cases the results in [10] are stronger, in some cases our results are stronger. For the most important case, equation (5) of Theorem 7, the bound given by Durrett [10] is *exactly* the same, although his methods are entirely different from ours.

For the value of the critical probability, Durrett [10] gave the first proof for the CHKNS model. As we have pointed out, the result for the $1/j$-graph, our main focus, follows immediately from results of Kalikow and Weiss [15] and Shepp [19]. One can easily translate this result to the CHKNS model. Durrett argues directly in the CHKNS model using the method of Durrett and Kesten [11], which generalizes [19]. In any case, as stated in [10], the value of the critical probability for the CHKNS model was (essentially) determined rigorously at least 10 years before the introduction of the model.

2 Results

For a graph G we write $C_1(G)$ for the order of the largest component of G. We say that an event holds *with high probability* (**whp**) if it holds with probability tending to 1 as $n \to \infty$. Although our focus is the supercritical graph $G_n(c)$, $c > 1/4$, we shall prove our results by counting paths in the subcritical graph $G_n(c)$, $c < 1/4$. In addition to proving the new bound we shall actually need, our path counting methods give us as a side effect different proofs of some known results. In particular, we shall re-prove the following result, an immediate consequence of results of Zhang [24].

Theorem 1 (Implicit in Zhang [24]). *If* $\omega(n) \to \infty$ *then* **whp** *we have* $C_1(G_n(1/4)) \leq \omega\sqrt{n \log n}$.

In other words, $C_1(G_n(1/4)) = O_p(\sqrt{n \log n})$. In particular, at, or below, $c = 1/4$ there is no giant component in $G_n(c)$. A stronger bound, removing the log factor, follows from results of Durrett [10], who also gives a bound $O(\sqrt{n}/\log n)$ on the size of the component containing vertex 1.

Below the critical point, $G_n(c)$ and the classical model $G_{n,a/n}$ behave very differently. For $0 < a < 1$, the largest component of $G_{n,a/n}$ has order $O(\log n)$ **whp** as $n \to \infty$. In contrast, for any $0 < c < 1/4$, the graph $G_n(c)$ has **whp** a component of order some power of n. In the following result, $\beta(c) = 1/2 - \sqrt{1/4 - c}$, a function that will play an important role in all our proofs.

Theorem 2. *For any $c < 1/4$ the expected size of the component of $G_n(c)$ containing the vertex 1 is $O(n^{\beta(c)})$. Also, for any $0 < c' < c < 1/4$,* **whp** *the component of $G_n(c)$ containing the vertex 1 has size at least $n^{c'}$.*

The upper bound in Theorem 2 has been obtained independently by Durrett [10]; see Section 5.1 for details. The rather crude lower bound establishes that for any positive c the largest component of $G_n(c)$ has order $n^{\Theta(1)}$.

We now turn to our main results, giving upper and lower bounds on the size of the giant component for $c > 1/4$. Perhaps the most important is the upper bound, which shows that the phase transition has infinite order.

Theorem 3. *For any $\eta > 0$ there is an $\varepsilon(\eta) > 0$ such that if $0 < \varepsilon < \varepsilon(\eta)$ then*

$$C_1(G_n(1/4 + \varepsilon)) \geq \exp\left(-\frac{\pi + \eta}{2\sqrt{\varepsilon}}\right) n$$

holds **whp** *as $n \to \infty$.*

Theorem 4. *For any $\eta > 0$ there is an $\varepsilon(\eta) > 0$ such that if $0 < \varepsilon < \varepsilon(\eta)$ then*

$$C_1(G_n(1/4 + \varepsilon)) \leq \exp\left(-\frac{1 - \eta}{2\sqrt{\varepsilon}}\right) n$$

holds **whp** *as $n \to \infty$.*

Taken together, Theorems 3 and 4 show that as $n \to \infty$ with ε fixed, the largest component of $G_n(1/4 + \varepsilon)$ has order $\exp(-\Theta(1/\sqrt{\varepsilon}))n$, where the $\Theta(.)$ refers to $\varepsilon \to 0$.

For the closely related CHKNS model, Dorogovtsev, Mendes and Samukhin [9] use very different methods to obtain a very precise formula for the size of the giant component (although presumably one should interpret = in their equation (C10) as $\sim$). Their argument is not rigorous (see Section 11, or [10]); however, it may well be possible to make it rigorous. In this case, it would certainly follow that

$$C_1(G_n(1/4 + \varepsilon)) = \exp\left(-\frac{\pi + o(1)}{2\sqrt{\varepsilon}}\right) n.$$

The formula in [9] for the CHKNS model gives even the limiting value as $\varepsilon \to 0$ of the constant in front of the exponential factor, but there would be some loss of accuracy involved when translating this result to $G_n(c)$ by a straightforward comparison argument.

A much weaker lower bound than Theorem 3 is implicit in the work of Shepp [19] who used it to show that the infinite graph is connected with probability 1. Using his method, or that of Durrett and Kesten [11], one obtains a bound with C/ε in the exponent, rather than $C/\sqrt{\varepsilon}$. For the CHKNS model, a lower bound with C/ε in the exponent has been obtained in this way by Durrett [10]. Our lower bound is obtained using a general result bounding from below the size of the giant component in a certain inhomogeneous graph, in terms of an eigenfunction of the associated continuous operator. This result, given in Section 8, is not in general best possible but gives a very good bound here. Indeed, the arguments of [9] suggest that the constant $\pi/2$ in the exponent in Theorem 3 is best possible.

As mentioned in the introduction, above the critical probability, the second largest component is rather small.

Theorem 5. *Let $c > 1/4$ be fixed. Then* **whp** *the second largest component of $G_n(c)$ has order at most $(\log n)^4$.*

Most likely, the real size of the second largest component is $O(\log n)$ above the critical probability, as in $G(n, c/n)$. This is what one would expect from the exponential decay of exponent sizes described in [9].

So far we have considered the graph analogue of bond percolation. Instead one could consider the analogue of site percolation: starting with a graph $G = G_n$, instead of selecting edges independently with probability c to obtain a subgraph $G(c)$, we could select vertices independently with probability c to obtain an induced subgraph $G'(c)$. Both types of deletion are commonly considered in percolation questions: in some contexts, for example looking at the robustness of the internet graph, vertex deletion is more natural. Here there is no difference.

Remark. All the results of this section hold with the vertex deleted graph $G'_n(c)$ in place of $G_n(c)$.

This is easy to see from the nature of the arguments we use. In particular, we prove all our upper bounds on component sizes by counting expected numbers of paths. For any G, any given path is less likely to be present in $G'(c)$ than in $G(c)$, in fact by exactly a factor c. So all bounds on expected numbers of paths proved for $G_n(c)$ are valid for $G'_n(c)$. For the lower bound, Theorem 3, we use a neighbourhood expansion argument. Usually such arguments have an extra factor c for vertex deletion (which would not affect the result as stated); in fact, looking at the details of the non-standard argument in Section 8, the proof goes through essentially unchanged for $G'_n(c)$, with no extra factor.

In the proofs that follow we shall omit $\lfloor . \rfloor$ and $\lceil . \rceil$ signs where their omission makes no difference to the argument. Surprisingly, in one of our arguments we cannot simply replace x by $\lfloor x \rfloor$, even when x is large; see the end of Section 6.

3 Exponents

To prove Theorems 1 and 4 we use path counting arguments, bounding the expected number of paths between certain vertices in $G_n(c)$. Much of the time, it will be more convenient to work in the infinite graph $G_\infty(c)$ with vertex set $\{1, 2, \ldots, \}$, in which each pair $i < j$ is joined independently with probability c/j. We shall write K_∞ for the complete graph on $\{1, 2, \ldots, \}$, so $G_\infty(c) \subset K_\infty$.

The arguments in the next two sections will use a 'renormalization' method, in which paths in $G_\infty(c)$ are replaced by paths in a similarly defined graph with different edge probabilities, of the form $\alpha(j/i)^\beta/j$. Repeated application of this argument leads to a series of exponents α_k, β_k whose limiting behaviour we shall need to analyze. Throughout this paper c will be the edge-probability parameter used to define $G_n(c)$ or $G_\infty(c)$. For this and the next two sections, we restrict our attention to $c \leq 1/4$.

Let $(\alpha_0, \beta_0) = (c, 0)$, and for $k \geq 0$ let

$$(\alpha_{k+1}, \beta_{k+1}) = \left(\frac{\alpha_k^2}{1 - 2(\alpha_k + \beta_k)}, \alpha_k + \beta_k \right), \tag{1}$$

assuming for the moment that $\alpha_k + \beta_k < 1/2$ for every k, so there is no problem with the division, and α_{k+1} is positive. Note that $\beta_k = \sum_{i<k} \alpha_i$, so β_k is an increasing sequence. Let $\beta = \beta(c) = \lim_{k\to\infty} \beta_k$.

If $c = 1/4$, then it is easy to check that $\alpha_k = 2^{-(k+2)}$, $\beta_k = 1/2 - 2^{-(k+1)}$, and $\beta = 1/2$. Now, for each $k > 0$, α_k and β_k are increasing functions of α_{k-1}, β_{k-1}, and hence, inductively, of c. Thus, for any $c \leq 1/4$, α_k and β_k are bounded above by the values for $c = 1/4$. In particular, the assumption $\alpha_k + \beta_k < 1/2$ does indeed hold for every k. Furthermore, $\alpha_k \to 0$, and the limit β_k exists and is at most $1/2$.

For each k, let $\theta_k = (1 - 2(\alpha_k + \beta_k))^2 - 4\alpha_k^2$. For $k \geq 0$ we have

$$\begin{aligned} \theta_{k+1} &= (1 - 2\beta_{k+1} - 2\alpha_{k+1})^2 - 4\alpha_{k+1}^2 \\ &= (1 - 2\beta_{k+1})^2 - 4(1 - 2\beta_{k+1})\alpha_{k+1} \\ &= (1 - 2(\alpha_k + \beta_k))^2 - 4\alpha_k^2 = \theta_k. \end{aligned}$$

Thus for every k we have $\theta_k = \theta_0 = (1 - 2c)^2 - 4c^2 = 1 - 4c$. Since $\alpha_k \to 0$ it follows that $\beta_k \to 1/2 - \sqrt{1/4 - c}$. In other words, for any $c \leq 1/4$ we have

$$\beta(c) = 1/2 - \sqrt{1/4 - c}.$$

4 Paths between 1 and 2

This is a warm-up for the next section, but we shall essentially re-use the proof, so there isn't much harm in writing it out separately.

Theorem 6. *For $c \leq 1/4$ the expected number of paths in $G_\infty(c)$ between vertices 1 and 2 is at most $\sum_{k\geq 0} \alpha_k(c) 2^{-\beta_k(c)} \leq \beta(c)$.*

Hence, at $c = 1/4$, the probability that 1 and 2 lie in the same component is at most $\beta(1/4) = 1/2$.

Note that Shepp [19] used a very different method to bound the expected number of 1-2 paths in $G_\infty(c)$, obtaining

$$\sum_{k\geq 0} c^{k+1}\binom{2k}{k}\frac{1}{k+1}$$

as an upper bound. This sum is essentially the generating function for the Catalan numbers, and is $1/2 - \sqrt{1/4-c} = \beta(c)$ exactly!

Proof of Theorem 6. Monotone paths (of the form $x_0x_1\cdots x_t$ with $x_0 < x_1 < \cdots < x_t$) are easy to deal with, so we think of a general path as a zigzag of monotone paths. In fact, we'll sum over the tops of the zigzags, and deal with *hooks*. By an *x-y hook* we shall mean an x-y path which consists of a monotone increasing x-z path followed by a monotone decreasing z-y path, where z is strictly greater than $\max\{x, y\}$.

Now any 1-2 path in K_∞ is either a single edge, or a series of hooks stuck together. More precisely, let P be a 1-2 path with more than 1 edge, and let $x_0, x_1, \ldots, x_r$ be the local minima of P in the order they appear along P. Since 1 and 2 are the two minimum vertices, the endpoints of P are local minima, so $x_0 = 1$ and $x_r = 2$. Hence, as any path zigzags between local minima and local maxima, P consists of an x_0-x_1 hook followed by an x_1-x_2 hook, and so on. Note that $x_0x_1\cdots x_r$ is a 1-2 path strictly shorter than P.

Given any path P', to *expand* P' will mean to replace every edge of P' by a hook. Thus, the 1-2 path P is obtained by expanding a shorter 1-2 path, and any 1-2 path is obtained by expanding the single edge 1-2 path some non-negative number of times.

By a *k-fold expansion* of a path P we mean any path obtained by expanding P k times. Note that an l-fold expansion of a k-fold expansion of P is a $(k+l)$-fold expansion of P. Let us call a k-fold expansion of a single edge ab a *level-k* path, so any 1-2 path is a level-k path for some k. Note that any level-k a-b path is a $(k-1)$-fold expansion of an expansion of $P = ab$, i.e., a $(k-1)$-fold expansion of an a-b hook. Since expansion operates separately on the edges in a path, it follows that any level-k path can be obtained from a hook by replacing each edge by a level-$(k-1)$ path; this is how we calculate.

Claim. Let $\alpha_k = \alpha_k(c)$ and $\beta_k = \beta_k(c)$ be the quantities defined in Section 3. Then for $k \geq 0$ and $a < b$ the expected number $N_k(a, b)$ of level-k a-b paths is at most

$$\alpha_k a^{-\beta_k} b^{-1+\beta_k}.$$

We prove the claim by induction on k; for $k = 0$ there is nothing to prove, as $\alpha_0 = c$ and $\beta_0 = 0$. To prove the induction step, fix $a < b$ and recall that any level-$(k+1)$ a-b path can be obtained from a hook $a = x_0 < x_1 < \cdots < x_r = z = y_s > \cdots > y_1 > y_0 = b$ by replacing the edges by edge-disjoint level-k

paths. (In fact the paths are internally vertex disjoint, but we shall not use this.) Since edge-disjoint paths are present independently in $G_\infty(c)$, the expectation $N_{k+1}(a,b)$ is at most

$$\sum_{r,s\geq 1}\ \sum_{a=x_0<x_1<\cdots<x_r=z=y_s>\cdots>y_1>y_0=b}\ \prod_{i=0}^{r-1} N_k(x_i,x_{i+1}) \prod_{i=0}^{s-1} N_k(y_i,y_{i+1}), \qquad (2)$$

where the sums simply run over all a-b hooks, and, as usual, an empty product has value 1.

[It is not important, but to make the derivation of (2) absolutely clear, let us note that we are not using the (non-trivial) van den Berg-Kesten inequality, but rather the following trivial special case: Suppose that $\mathcal{A}_j$, $1\leq j\leq J$, are families of sets of edges of K_∞, G_∞ is a random subgraph of K_∞ obtained by selecting edges independently with certain probabilities, $X_j = |\{A\in\mathcal{A}_j : A\subset E(G_\infty)\}|$, $\mathcal{A}$ is the set of *disjoint* J-tuples $(A_1,\ldots,A_J)$ with $A_j\in\mathcal{A}_j$, and X is the number of $(A_1,\ldots,A_J)\in\mathcal{A}$ with $\bigcup A_j\subset E(G_\infty)$. Then $\mathbf{E}(X)\leq\prod_{j=1}^J\mathbf{E}(X_j)$. Indeed,

$$\mathbf{E}(X) = \sum_{(A_1,\ldots,A_J)\in\mathcal{A}}\ \prod_j \Pr(A_j\subset E(G_\infty)) \leq \prod \mathbf{E}(X_j),$$

where the equality follows from the disjointness of the A_j, and the inequality from the fact that the final product is equal to the sum of the same summand over the larger set of all tuples $(A_1,\ldots,A_J)$ with $A_j\in\mathcal{A}_j$.]

Using the induction hypothesis, the first product in (2) can be bounded as follows:

$$\prod_{i=0}^{r-1} N_k(x_i,x_{i+1}) \leq \alpha_k^r \prod_{i=0}^{r-1} x_i^{-\beta_k} x_{i+1}^{-1+\beta_k} = \alpha_k^r a^{-\beta_k} z^{-1+\beta_k} \prod_{i=1}^{r-1} x_i^{-1}, \qquad (3)$$

as $x_0=a$, $x_r=z$. Fixing z and summing over the internal vertices in the a-z path, we obtain

$$\begin{aligned}
\sum_{r\geq 1}\ \sum_{a=x_0<x_1<\cdots<x_r=z}\ &\prod_{i=0}^{r-1} N_k(x_i,x_{i+1}) \\
&\leq \alpha_k a^{-\beta_k} z^{-1+\beta_k} \sum_{r\geq 1}\ \sum_{a<x_1<\cdots<x_{r-1}<z}\ \prod_{i=1}^{r-1}\frac{\alpha_k}{x_i} \\
&= \alpha_k a^{-\beta_k} z^{-1+\beta_k} \prod_{j=a+1}^{z-1}(1+\alpha_k/j),
\end{aligned}$$

where the first step follows from applying (3) to each summand, and the last step simply by expanding the product in the final line. Let us write $S(a,z)$ for $\prod_{j=a+1}^{z-1}(1+\alpha_k/j)$. Then, bounding the sum over $y_1<\cdots<y_{s-1}$ of the second

product in (2) in the same way, we see that $N_{k+1}(a,b)$ is at most

$$\begin{aligned} N_{k+1}(a,b) &\leq \sum_{z=b+1}^{\infty} \left(\alpha_k a^{-\beta_k} z^{-1+\beta_k} S(a,z)\right) \left(\alpha_k b^{-\beta_k} z^{-1+\beta_k} S(b,z)\right) \\ &= \alpha_k^2 a^{-\beta_k} b^{-\beta_k} \sum_{z>b} z^{-2+2\beta_k} S(a,z) S(b,z). \end{aligned}$$

Now $\log(1+x) \leq x$, so

$$\log S(a,z) \leq \sum_{j=a+1}^{z-1} \frac{\alpha_k}{j} < \int_{x=a}^{z-1} \frac{\alpha_k}{x} \mathrm{d}x = \alpha_k \log((z-1)/a) < \alpha_k \log(z/a). \quad (4)$$

Hence $S(a,z) < (z/a)^{\alpha_k}$, and

$$\begin{aligned} N_{k+1}(a,b) &\leq \alpha_k^2 a^{-\beta_k} b^{-\beta_k} \sum_{z>b} z^{-2+2\beta_k} z^{2\alpha_k} a^{-\alpha_k} b^{-\alpha_k} \\ &= \alpha_k^2 a^{-\beta_{k+1}} b^{-\beta_{k+1}} \sum_{z>b} z^{-2+2\beta_{k+1}} \\ &< \alpha_k^2 a^{-\beta_{k+1}} b^{-\beta_{k+1}} \int_{z=b}^{\infty} z^{-2+2\beta_{k+1}} \mathrm{d}z \\ &= \frac{\alpha_k^2}{1-2\beta_{k+1}} a^{-\beta_{k+1}} b^{-1+\beta_{k+1}}, \end{aligned}$$

completing the proof of the induction step, and hence of the claim. Note that we used $\beta_{k+1} < \beta(c) \leq 1/2$ here to ensure convergence of the final integral.

To complete the proof of the theorem, note that the expected number of 1-2 paths is at most $\sum_{k=0}^{\infty} N_k(1,2) \leq \sum_k \alpha_k 2^{-1+\beta_k} \leq \sum_k \alpha_k = \beta(c)$, as claimed. □

5 Below the critical probability

In this section we give an upper bound on the expected number of paths between two given vertices in the subcritical graph $G_n(c)$, $c < 1/4$. We shall use this bound in the next section to establish slow emergence of the giant component above $c = 1/4$.

The counting is a little more complicated than in the previous section, as the endpoints of a general path need not be minima. We fix $s < t$ and aim to bound the expected number of s-t paths. As before, we write $\beta = \beta(c) = 1/2 - \sqrt{1/4 - c}$ for the function defined in Section 3, and used repeatedly in Section 4.

Theorem 7. *For $s < t$ and $c < 1/4$ the expected number of s-t paths in $G_\infty(c)$ is at most*

$$\left(\beta + \beta^2/(1-2\beta)\right) s^{-\beta} t^{-1+\beta}. \quad (5)$$

For $s < t$ the expected number of s-t paths in $G_\infty(1/4)$ is at most

$$\frac{3+\log s}{4\sqrt{st}}. \tag{6}$$

We shall not in fact use the second statement, which is essentially the same as the result of Zhang [24], who obtained a bound $c_1 \log s/\sqrt{st}$ (the upper bound in his equation (2)) by a very different method. We state (6) because we obtain it from our proof of the result we shall need, namely (5).

In [10] Durrett sketches a proof of *exactly* the same bound (5), with the constant written in the simpler form $2\delta/\sqrt{1-8\delta}$, where 2δ corresponds to c. His proof uses very different methods from ours. He also sketches a proof of a stronger form of Zhang's bound applicable to $G_n(1/4)$ rather than $G_\infty(1/4)$.

Before turning to the proof of Theorem 7 let us note the following corollary, concerning the total number of paths in G. Here, and throughout, we do not count a single vertex as a path.

Corollary 8. *For $c < 1/4$ the expected number of paths in $G_n(c)$ is at most $f(c)n$, where $f(1/4-\varepsilon) = \Theta(1/\sqrt{\varepsilon})$. For $n \geq e^3$ the expected number of paths in $G_n(1/4)$ is at most $n \log n$.*

Durrett's stronger bound [10] removes the log factor when $c = 1/4$.

Proof. Suppose first that $c < 1/4$. Fix $n \geq 2$ and $1 \leq s < t \leq n$. By Theorem 7 the expected number of paths from s to t in $G_\infty(c)$ is at most

$$\big(\beta + \beta^2/(1-2\beta)\big)\, s^{-\beta} t^{-1+\beta}.$$

Since we may consider $G_n(c)$ as a subgraph of $G_\infty(c)$, this is also an upper bound for the expected number of s-t paths in $G_n(c)$. Summing over $s < t$ we see that the total expected number of paths in $G_n(c)$ is at most

$$\sum_{1\leq s<t\leq n} \big(\beta + \beta^2/(1-2\beta)\big)\, s^{-\beta} t^{-1+\beta}.$$

With t fixed we have

$$\sum_{s<t} s^{-\beta} < \int_0^t s^{-\beta}\mathrm{d}s = t^{1-\beta}/(1-\beta),$$

so the expected number of paths is at most $f(c)n$, where

$$f(c) = (\beta + \beta^2/(1-2\beta))/(1-\beta),$$

which is $\Theta(1/\sqrt{1/4-c})$ as $c \to 1/4$.

Suppose now that $c = 1/4$. This time we use the bound (6) from Theorem 7. Crudely, we have $3 + \log s \leq 3 + \log n \leq 2\log n$, for $n \geq e^3$. Thus the expected number of paths in $G_n(1/4)$ is at most

$$\sum_{1\leq s<t\leq n} \frac{\log n}{2\sqrt{st}} \leq \frac{\log n}{2} \int_{0\leq x\leq y\leq n} \frac{1}{\sqrt{xy}}\mathrm{d}x\mathrm{d}y = n\log n.$$

□

As an immediate consequence we have the following result concerning the largest component $C_1(G)$ of G, which is just Theorem 1 of Section 2, and is implicit in the work of Zhang [24]. A stronger result, removing the $\log n$ factor, has been given by Durrett [10].

Corollary 9. *Let $\omega(n) \to \infty$. Then the probability that $C_1(G_n(1/4)) \geq \omega\sqrt{n \log n}$ tends to zero as $n \to \infty$.*

Proof. The number of paths in any graph G is at least $\binom{C_1(G)}{2}$, so the result follows from Corollary 8 by Markov's inequality. □

Thus, although this is not how we shall use it, Theorem 7 incidentally gives the known lower bound on the critical probability for $G_n(c)$.

Proof of Theorem 7. Let $c \leq 1/4$ and $1 \leq s < t$ be fixed throughout. We shall work in $G_\infty(c)$.

Given an s-t path P, let the *simplification* $S(P)$ of P be the s-t path whose internal vertices are those internal vertices of P that are local minima in P. Note that $S(P)$ is strictly shorter than P unless all internal vertices are local minima, i.e., $P = st$ or $P = sat$ for some $a < s$. Let us call paths of these two types *basic*. Thus, for any s-t path P, repeatedly applying the simplification operation $S(.)$ we eventually reach a basic path, which is unchanged by further applications of $S(.)$.

Given an s-t path P', which paths P have $S(P) = P'$? Internal edges in P' correspond to paths in P between consecutive minima, i.e., hooks, but external edges may correspond either to hooks or to monotone paths in P.

Let us call the edge ab, $a < b$, of K_∞ (or of an s-t path) *special* if $b = s$ or $b = t$, and *normal* otherwise. Note that the classification of an edge depends on its upper endpoint (i.e., $\max\{i, j\}$ if the edge is ij), so we shall often write a particular edge that we are considering as ij, $i < j$. For an s-t path P, every normal edge in $S(P)$ comes from a hook, while special edges may come from a hook or from a monotone path. We classify edges of K_∞ of the form sb, $s < b$, $b \neq t$, as normal since, if such an edge arises in the simplification $P' = S(P)$ of an s-t path, then b is a local minimum in P; more precisely, b is the first internal local minimum in P. Since $b > s$ it follows that the part of P from s to b must be a hook. The same argument applies to edges tb, $b > t$.

For a normal edge ab, $a < b$, let us define the *expansions* of ab to be the a-b hooks not meeting $\{s, t\}$ except, possibly, at a. Note that the expansion of a normal edge is a path consisting of normal edges. For a special edge ab, $a < b$, the expansions of ab are a-b hooks together with monotone a-b paths, in both cases subject to the restriction of not meeting $\{s, t\}$ except at b and, possibly, at a. An expansion of a path is a path obtained by expanding each edge. We have chosen the definitions so that any s-t path P is an expansion of $S(P)$. Hence, any s-t path P is, for all sufficiently large k, obtainable from some basic path by k-fold expansion.

As before, for $a < b$ we inductively bound $E_k(a, b)$, the expected number k-fold expansions of the path $P = ab$ which are present in $G_\infty(c)$, again in terms of the quantities α_k, β_k defined in Section 3. We use different notation here and in Section 4 (E_k rather than N_k) as the notions of expansion are different. Note that from the remarks above the expected number of s-t paths is at most

$$\sup_k E_k(s,t) + \sum_{a<s} \sup_k E_k(a,s)E_k(a,t), \tag{7}$$

recalling that for a special edge the expectation $E_k(a,b)$ is non-decreasing in k, as we may expand the edge to itself if we wish, and that any s-t path is a k-fold expansion of a basic path for some k, and hence for all sufficiently large k.

Now for a normal edge ab, $a < b$, expansion replaces ab by a hook of normal edges, so, using results from Section 4, we have

$$E_k(a,b) \leq N_k(a,b) \leq \alpha_k a^{-\beta_k} b^{-1+\beta_k}. \tag{8}$$

(The first inequality would be equality were it not for the technicality that we have excluded s and t.)

For a special edge ab, $a < b$, $b \in \{s,t\}$, we claim that

$$E_k(a,b) \leq \beta_{k+1} a^{-\beta_k} b^{-1+\beta_k}. \tag{9}$$

Again the proof is by induction on k, and as $\beta_1 = \alpha_0 = c$ the base case $k = 0$ is trivial. To prove the induction step, fix $a < b$, and consider any $(k+1)$-fold expansion of the special edge ab. As in Section 4, this expansion can be viewed as a k-fold expansion of an expansion P of ab. If the first expansion from ab to P replaces ab by a hook, this is a hook of normal edges, so, as before, the expected number of paths obtained in this way present in $G_\infty(c)$ is at most

$$X_1 = \alpha_{k+1} a^{-\beta_{k+1}} b^{-1+\beta_{k+1}}. \tag{10}$$

On the other hand, if the first step is to expand to a monotone path $a = x_0 < x_1 < \cdots < x_r = b$, then the edge $x_{r-1}b$ is special while the other edges are normal. Using (8) and the induction hypothesis (9), the expected number of paths obtained this way is at most

$$X_2 = \sum_{r\geq 1} \sum_{a=x_0<x_1<\cdots<x_r=b} \beta_{k+1} \alpha_k^{r-1} a^{-\beta_k} b^{-1+\beta_k} \prod_{i=1}^{r-1} x_i^{-1}.$$

(The product collapses exactly as the second product in (3); the only difference for the special edge is a factor of β_{k+1} instead of α_k.) Arguing as before we may bound X_2 by $\beta_{k+1} a^{-\beta_k} b^{-1+\beta_k} S(a,b)$, where, as before, $S(a,b) = \prod_{j=a+1}^{b-1}(1 + \alpha_k/j)$. Using our bound (4) on $\log S(a,b)$, we obtain

$$X_2 \leq \beta_{k+1} a^{-\beta_k} b^{-1+\beta_k} (b/a)^{\alpha_k} = \beta_{k+1} a^{-\beta_{k+1}} b^{-1+\beta_{k+1}}.$$

Putting the bounds above together, we have

$$E_{k+1}(a,b) \leq X_1 + X_2 \leq (\alpha_{k+1} + \beta_{k+1}) a^{-\beta_{k+1}} b^{-1+\beta_{k+1}}.$$

Since $\alpha_{k+1} + \beta_{k+1} = \beta_{k+2}$, this proves the induction step, and so completes the proof of (9).

Writing, as before, β for $\beta_\infty = \lim_{k\to\infty} \beta_k$, we see that as k increases, for a special edge ab our upper bound (9) on $E_k(a,b)$ increases to $\beta a^{-\beta} b^{-1+\beta}$. Hence, from (7), the expected number $\mathbf{E}P(s,t)$ of s-t paths is bounded by

$$\beta s^{-\beta} t^{-1+\beta} + \sum_{a=1}^{s-1} \beta^2 a^{-2\beta} s^{-1+\beta} t^{-1+\beta}.$$

Suppose first that $c < 1/4$. Then $\beta < 1/2$, so

$$\sum_{a=1}^{s-1} a^{-2\beta} < \int_{x=0}^{s} x^{-2\beta} \mathrm{d}x = \frac{s^{1-2\beta}}{1-2\beta},$$

implying that

$$\begin{aligned} \mathbf{E}P(s,t) &\leq \beta s^{-\beta} t^{-1+\beta} + \frac{s^{1-2\beta}}{1-2\beta} \beta^2 s^{-1+\beta} t^{-1+\beta} \\ &= \big(\beta + \beta^2/(1-2\beta)\big)\, s^{-\beta} t^{-1+\beta}. \end{aligned}$$

For $c = 1/4$, when $\beta = 1/2$, we have $\sum_{a=1}^{s-1} a^{-2\beta} \leq 1 + \log s$, so

$$\mathbf{E}P(s,t) \leq \frac{3 + \log s}{4\sqrt{st}},$$

completing the proof. □

5.1 The component containing the first vertex

This subsection is a slight diversion from our main course, but again uses path counting arguments. We aim to prove Theorem 2, giving upper and lower bounds on the size of the component containing vertex 1 in $G_n(c)$, $c < 1/4$.

Proof of Theorem 2. Fix $c < 1/4$. The upper bound follows from Theorem 7 in a similar way to Corollary 8: from (5), the expected number of paths in $G_n(c)$ from vertex 1 to vertex t is at most

$$\big(\beta + \beta^2/(1-2\beta)\big)\, t^{-1+\beta}.$$

Summing over t, the expected size of the component containing 1 is at most

$$1 + \big(\beta + \beta^2/(1-2\beta)\big) \sum_{t=2}^{n} t^{-1+\beta} = \Theta(n^\beta),$$

where we once again use comparison with an integral to bound the sum.

For the lower bound we argue directly, and rather crudely, considering only monotone paths. Given $0 < c' < c$, choose any c'' strictly between c' and c, and let ε be a very small constant (depending on c', c'' and c) to be chosen later.

For $0 \le r \le T$, set $t_r = n^\varepsilon(1+\varepsilon)^r$, where T is chosen so that $(1-\varepsilon)n \le t_T \le n$. We ignore rounding and treat t_r as an integer; this makes no difference to the argument. For $0 \le r \le T$, set $s_r = (\varepsilon^2 \log n)(1+\varepsilon)^{c''r}$, again ignoring rounding. Note that $s_0 = \Theta(\log n)$, while $s_T = (\varepsilon^2 \log n)(t_T/t_0)^{c''}$, which is at least $n^{c'}$ if ε is chosen small enough and n is large.

We shall show that with very high probability the subgraph $G_{t_r}(c)$ of $G_n(c)$ induced by the first t_r vertices is such that the component $C^1(t_r)$ containing vertex 1 has order at least s_r.

For $r = 0$ this is easy to check: the expected degree of vertex 1 in $G_{t_0}(c)$ is

$$\sum_{j=2}^{t_0} \frac{c}{j} \sim c \log t_0 \sim \varepsilon c \log n,$$

which, for ε small, is much larger than s_0. Since vertices are joined to 1 independently, it is easy to check (for example by computing the variance and using Chebyshev's inequality) that with high probability the degree of 1 in $G_{t_0}(c)$ is at least half its expectation, say, and hence at least s_0.

Suppose now that the component $C^1(t_r)$ of $G_{t_r}(c)$ containing vertex 1 has order at least s_r, and let us condition on $G_{t_r}(c)$. For each j, $t_r < j \le t_{r+1}$, the conditional probability that j is joined to $C^1(t_r)$ is at least

$$(1-\varepsilon)\frac{cs_r}{t_{r+1}}.$$

Indeed the probability that j is joined directly to any given earlier vertex is $c/j \ge c/t_{r+1}$, so the expected number of edges from j to $C^1(t_r)$ is at least $\mu = cs_r/t_{r+1}$. Since each possible edge from j to $C^1(t_r)$ is present independently, and $\mu \le cs_0/t_1 = o(1)$, the probability that at least one edge from j to $C^1(t_r)$ is present is at least $(1-\varepsilon)\mu$.

It follows that the number $X(r)$ of vertices j with $t_r < j \le t_{r+1}$ joined to $C^1(t_r)$ stochastically dominates a binomial distribution with mean

$$(1-\varepsilon)(t_{r+1}-t_r)\frac{cs_r}{t_{r+1}} \ge \varepsilon c(1-\varepsilon)^2 s_r.$$

Now as s_r is reasonably large, namely at least $s_0 = \Theta(\log n)$, it follows from the Chernoff bounds that there is a $\gamma > 0$ such that with probability at least $1-n^{-\gamma}$ we have $X(r) \ge \varepsilon c(1-\varepsilon)^3 s_r$. But choosing ε small enough, $c(1-\varepsilon)^4 > c''$, so with high probability we have $X(r) \ge \frac{\varepsilon}{1-\varepsilon}c'' s_r > s_{r+1} - s_r$, and hence $|C^1(t_{r+1})| \ge s_{r+1}$.

In summary, given that $|C^1(t_r)| \ge s_r$, with probability at least $1 - n^{-\gamma}$ we have $|C^1(t_{r+1})| \ge s_{r+1}$. As $|C^1(t_0)| \ge s_0$ **whp** and $T = O(\log n) = o(n^\gamma)$, it follows that **whp** we have $|C^1(t_T)| \ge s_T \ge n^{c'}$, completing the proof. $\square$

The argument for the lower bound is very simple, relying only on monotone paths, and the bound obtained is likely to be far from the truth. In fact, we believe that there isn't much overcounting in the upper bound, and that there is a great deal of independence. Hence, we believe that **whp** the component containing vertex 1 has order $n^{\beta(c)-o(1)}$ as $n \to \infty$ with $c < 1/4$ fixed. It might be possible to show this using second moment or Janson inequality methods, but this would involve dealing with arbitrary paths: we cannot restrict to monotone paths, or level-k paths for any fixed k, and hope to obtain the correct exponent. Thus the calculations are likely to be rather involved, as are the lower bounds given by Zhang [24] and Durrett [10] on the probability that two given vertices lie in the same component.

The upper bound in Theorem 2 has been obtained independently by Durrett [10], who also raises the question as to whether the size of the component containing 1 is close to its expectation. He also says 'it seems likely that the component containing 1 will with high probability be the largest component'. This is easily seen to be false; as shown in Section 4, or from Shepp [19], the expected number of paths between vertices 1 and 2 in $G_n(c)$, $c \le 1/4$, is at most 1/2. Thus with probability at least 1/2 the vertices 1 and 2 lie in different components. Now the first two vertices are not in fact distinguished by the model, so with probability at least 1/4 the component containing vertex 2 is different from, and at least as large as, the component containing 1. Actually, it is easy to see that for $c \le 1/4$ fixed and for any fixed k, there is a positive probability that the first $k+1$ vertices of $G_n(c)$ lie in different components, and hence a positive probability that the component containing vertex 1 is not one of the k largest. If one writes $C_k = C_k(G_n(c))$ for the order of the kth largest component, for $c < 1/4$ we expect the vector $(C_1 n^{-\beta(c)}, \ldots, C_k n^{-\beta(c)})$ to converge in distribution as $n \to \infty$, to some non-trivial random vector, each of whose marginals will have positive density on all of $(0, \infty)$.

6 An upper bound on the giant component

Although Theorem 7 only applies when $c \le 1/4$, we can nonetheless use it to obtain results about $G_n(c)$ for $c = 1/4+\varepsilon$, by considering the graph as the union of two graphs $G_n(c_1)$, $G_n(c_2)$ with c_1, $c_2 < 1/4$. We end up with a comparison to a graph in which each edge ij is present independently with probability $p/\sqrt{ij}$. This graph is much easier to handle than $G_n(c)$, and has a giant component for any $p > 0$, but it is easy to see that the giant component is very small: a corresponding result for a much more complicated model is given in [6].

Proof of Theorem 4. Let us recall the statement of the theorem: $\eta > 0$ is given, and we must show that there is an $\varepsilon(\eta) > 0$ such that if $0 < \varepsilon < \varepsilon(\eta)$ then we have

$$C_1(G_n(1/4+\varepsilon)) \le \exp\left(-\frac{1-\eta}{2\sqrt{\varepsilon}}\right) n$$

holding **whp** as $n \to \infty$.

We shall view $G_n(1/4+\varepsilon)$ as the union of $G_n(1/4-\varepsilon)$ and $G_n(2\varepsilon)$, and use Theorem 7. Recall that $\beta(c) = 1/2 - \sqrt{1/4-c}$. Thus $\beta(1/4-\varepsilon) = 1/2 - \sqrt{\varepsilon}$, and $\beta(\varepsilon) = \varepsilon + O(\varepsilon^2)$. To avoid square roots, it will be natural to work in terms of $\delta = \sqrt{\varepsilon}$, considering $G_n(1/4 - \delta^2)$ and $G_n(2\delta^2)$.

To be precise, let $\delta > 0$ be fixed throughout. We shall assume, as we may, that δ is smaller than some very small constant. In various places this assumption will be needed for our estimates to hold; we shall use it without comment. Set $c_1 = 1/4 - \delta^2$, $c_2 = 2\delta^2$ and $c = 1/4 + \delta^2 = c_1 + c_2$. We construct simultaneously three graphs $G = G_n(c)$, G_1 and G_2 on the vertex set $\{1, 2, \dots, n\}$, so that G_r has the distribution of $G_n(c_r)$, and $G = G_1 \cup G_2$. To do this, first construct $G = G_n(c)$ in the usual way, selecting each edge ij, $i < j$ independently with probability c/j. Then for each edge e present in G, toss a biased coin, putting the edge into G_1 with probability c_1/c and into G_2 with probability $c_2/c = 1 - c_1/c$. We do this independently for each edge. Note the following key facts. Firstly, G is exactly the edge-disjoint union of the graphs G_1 and G_2. Secondly, G_r does indeed have the distribution of $G_n(c_r)$. Finally, while G_1 and G_2 are not independent, from the independence in the definition of G and in the partitioning of G, for $\{i,j\} \neq \{i',j'\}$, the presence of the edge ij in either of G_1, G_2 is independent of the presence of $i'j'$ in either graph.

As $c_1 = 1/4 - \delta^2$ we have $\beta(c_1) = 1/2 - \delta$. Thus, from Theorem 7, for $1 \le s < t \le n$ the expected number $N_1(s,t)$ of s-t paths in G_1 satisfies

$$N_1(s,t) \le (1/8 + o(1))\delta^{-1} s^{-\frac{1}{2}+\delta} t^{-\frac{1}{2}-\delta}. \tag{11}$$

Similarly, the expected number $N_2(s,t)$ of s-t paths in G_2 satisfies

$$N_2(s,t) \le (2 + o(1))\delta^2 s^{-\varepsilon'} t^{-1+\varepsilon'}, \tag{12}$$

where $\varepsilon' = \beta(2\delta^2) = 1/2 - \sqrt{1/4 - 2\delta^2} = (2 + o(1))\delta^2$. The implicit functions in the $o(.)$ terms depend on δ only, not n, s or t.

For $s < t$ let us bound $N_{12}(s,t)$, the expected number of s-t paths in $G_n(c) = G_1 \cup G_2$ consisting of an s-u path in G_1 followed by a u-t path in G_2, where $u \notin \{s,t\}$. From (11), (12) and independence, we have

$$N_{12}(s,t) \le (1/4 + o(1))\delta \left(\sum_{0<u<s} u^{-\frac{1}{2}+\delta} s^{-\frac{1}{2}-\delta} u^{-\varepsilon'} t^{-1+\varepsilon'} + \sum_{s<u<t} s^{-\frac{1}{2}+\delta} u^{-\frac{1}{2}-\delta} u^{-\varepsilon'} t^{-1+\varepsilon'} + \sum_{u>t} s^{-\frac{1}{2}+\delta} u^{-\frac{1}{2}-\delta} t^{-\varepsilon'} u^{-1+\varepsilon'} \right).$$

Now each sum over u can be bounded by an integral with the same limits, because of the strict inequalities in the limits of the sums. Thus

$$N_{12}(s,t) \le (1/4 + o(1))\delta \left(s^{-\frac{1}{2}-\delta} t^{-1+\varepsilon'} \int_{u=0}^{s} u^{-\frac{1}{2}+\delta-\varepsilon'} \mathrm{d}u + s^{-\frac{1}{2}+\delta} t^{-1+\varepsilon'} \int_{u=s}^{t} u^{-\frac{1}{2}-\delta-\varepsilon'} \mathrm{d}u + s^{-\frac{1}{2}+\delta} t^{-\varepsilon'} \int_{u=t}^{\infty} u^{-\frac{3}{2}-\delta+\varepsilon'} \mathrm{d}u \right).$$

The integrals are convergent, so

$$
\begin{aligned}
N_{12}(s,t) &\leq (1/4+o(1))\delta \Bigg(\frac{s^{-\varepsilon'}t^{-1+\varepsilon'}}{\frac{1}{2}+\delta-\varepsilon'} \\
&\qquad + \frac{-s^{-\varepsilon'}t^{-1+\varepsilon'} + s^{-\frac{1}{2}+\delta}t^{-\frac{1}{2}-\delta}}{\frac{1}{2}-\delta-\varepsilon'} + \frac{s^{-\frac{1}{2}+\delta}t^{-\frac{1}{2}-\delta}}{\frac{1}{2}+\delta-\varepsilon'} \Bigg) \\
&= (1+o(1))\delta s^{-\frac{1}{2}+\delta}t^{-\frac{1}{2}-\delta} - (2+o(1))\delta^2 s^{-\varepsilon'}t^{-1+\varepsilon'} \\
&\leq (1+o(1))\delta s^{-\frac{1}{2}+\delta}t^{-\frac{1}{2}-\delta} \leq \frac{(1+o(1))\delta}{\sqrt{st}},
\end{aligned}
$$

where the last step is rather crude, and just uses $s < t$. We also need to bound $N_{21}(s,t)$, the expected number of s-t paths in G consisting of an s-u path in G_2 followed by a u-t path in G_1, where $u \notin \{s,t\}$. (There is no obvious symmetry, as we used $s < t$, and the formulae for individual edge probabilities are not symmetric in the endpoints.) As before, for $s < t$, from (11), (12) and independence, we have

$$
\begin{aligned}
N_{21}(s,t) \leq (1/4+o(1))\delta &\Bigg(\sum_{0<u<s} u^{-\varepsilon'}s^{-1+\varepsilon'}u^{-\frac{1}{2}+\delta}t^{-\frac{1}{2}-\delta} \\
&+ \sum_{s<u<t} s^{-\varepsilon'}u^{-1+\varepsilon'}u^{-\frac{1}{2}+\delta}t^{-\frac{1}{2}-\delta} + \sum_{u>t} s^{-\varepsilon'}u^{-1+\varepsilon'}t^{-\frac{1}{2}+\delta}u^{-\frac{1}{2}-\delta} \Bigg).
\end{aligned}
$$

Bounding the sums by integrals with the same limits, as before, after straightforward calculation one obtains $N_{21}(s,t) \leq (1+o(1))\delta/\sqrt{st}$. In fact, the upper bound on $N_{21}(s,t)$ obtained is *exactly* the same as our upper bound on $N_{12}(s,t)$: a different set of integrals sum to the same bound.

Let us remark that our final bounds are symmetric in s and t; this will again be crucial, allowing us to define $N_{ij}(s,t) = N_{ij}(t,s)$ for $s > t$ and still to use the same bound,

$$
N_{ij}(s,t) \leq \frac{(1+o(1))\delta}{\sqrt{st}}, \tag{13}
$$

for $\{i,j\} = \{1,2\}$. Similarly, defining $N_i(s,t) = N_i(t,s)$ for $s > t$, it will be useful to note the crude bound

$$
N_1(s,t) + N_2(s,t) \leq \frac{1}{4\delta\sqrt{st}}
$$

which follows very crudely from (11) and (12), using $\varepsilon' < 1/2$, and assuming δ sufficiently small to bound the $o(1)$ terms.

The key idea is to use (13) to make an appropriate comparison with a graph in which each edge ij is present independently with probability $p/\sqrt{ij}$, where $p = (1+o(1))\delta$. The crucial fact is that while this graph contains a giant component, vertices with large indices ('late' vertices) are very unlikely to lie in this component, as it is only the 'early' vertices that have large degrees, and each late vertex is highly unlikely to be joined to an early vertex.

Recall that $\eta > 0$ is given. Let us set $\rho = \exp(-(1-\eta/2)/\delta)$. Thus when $\delta = \sqrt{\varepsilon}$ is fixed, ρ is a positive constant, but, as $\varepsilon \to 0$, ρ tends to zero very quickly. We shall consider vertices with indices at most ρn to be *early* vertices, and all other vertices to be *late* vertices. Let $f(\varepsilon) > 0$ be arbitrarily small, but independent of n. Then it is easy to see that G is unlikely to contain a component of order at least $f(\varepsilon)n$ that does not contain any early vertices: otherwise, there is already a component C of order at least $f(\varepsilon)n$ in the subgraph of G induced by the late vertices. Putting the early vertices back in, each has probability at least $cf(\varepsilon)/2$ of being joined directly to a vertex of C. As there are many early vertices, **whp** at least one is joined to C. Hence, the probability that G has a giant component containing no early vertex is small, and to show that **whp** any giant component of G is small, it suffices to show that **whp** very few vertices of G are joined to early vertices.

Let us fix a late vertex a, and estimate the probability that a is joined to some early vertex in G. If it is, then there is a path $P = x_0x_1\cdots x_{r-1}x_r$ in G, where $x_0 = a$, $b = x_r$ is an early vertex, and all of $x_0, \ldots, x_{r-1}$ are late vertices. Now any path in G consists of an alternating sequence of paths in G_1 and G_2. Thus there is some sub-sequence $a = y_0y_1\cdots y_s = b$ of $x_0x_1\cdots x_r$ such that G_1 contains a y_{2i}-y_{2i+1} path for each i, and G_2 a y_{2i+1}-y_{2i+2} path, or vice versa. Let us write $\sigma(i)$ for the *parity* of i, namely 1 if i is odd and 2 if i is even. Given $y_0\cdots y_s$, the expected number of corresponding paths P present in G is at most

$$\prod_{i=0}^{s-1} N_{\sigma(i+1)}(y_i, y_{i+1}) + \prod_{i=0}^{s-1} N_{\sigma(i)}(y_i, y_{i+1}). \tag{14}$$

Here we are using the fact that distinct edges are present in G_1 and G_2 independently, and the fact that, in G_i, disjoint paths are present independently. (The bound then follows using the same reasoning used for (2).) We must now sum the expression (14) over $s \geq 1$ and over all intermediate late vertices $y_1, \ldots, y_{s-1}$.

Given a, b and s, let us write $S(s)$ for the sum of the quantity (14) over all late intermediate vertices $y_1, \ldots, y_{s-1}$, taking $y_0 = a$, $y_s = b$. Suppose first that s is even. Summing over all possibilities for the y_i, i odd, we can bound $S(s)$ as follows:

$$
\begin{aligned}
S(s) &\leq \sum_{\rho n < y_2, y_4, \ldots, y_{s-2} \leq n} \left(\prod_{i=0}^{s/2-1} N_{12}(y_{2i}, y_{2i+2}) + \prod_{i=0}^{s/2-1} N_{21}(y_{2i}, y_{2i+2}) \right) \\
&\leq 2 \sum_{\rho n < y_2, y_4, \ldots, y_{s-2} \leq n} \prod_{i=0}^{s/2-1} \frac{(1+o(1))\delta}{\sqrt{y_{2i} y_{2i+2}}} \\
&= \frac{2}{\sqrt{ab}} \{(1+o(1))\delta\}^{s/2} \sum_{\rho n < y_2, y_4, \ldots, y_{s-2} \leq n} \prod_{i=1}^{s/2-1} y_{2i}^{-1} \\
&= \frac{2}{\sqrt{ab}} \{(1+o(1))\delta\}^{s/2} \left(\sum_{\rho n < y \leq n} y^{-1} \right)^{s/2-1} \\
&\leq \frac{(2+o(1))\delta}{\sqrt{ab}} \{(1+o(1))\delta \log(1/\rho)\}^{s/2-1}.
\end{aligned}
$$

Similarly, for s odd we obtain

$$
\begin{aligned}
S(s) &\leq \sum_{\rho n < y_2, y_4, \ldots, y_{s-1} \leq n} \left(\prod_{i=0}^{(s-3)/2} N_{12}(y_{2i}, y_{2i+2}) N_1(y_{s-1}, y_s) \right. \\
&\qquad \left. + \prod_{i=0}^{(s-3)/2} N_{21}(y_{2i}, y_{2i+2}) N_2(y_{s-1}, y_s) \right) \\
&\leq \frac{1}{4\delta\sqrt{ab}} \{(1+o(1))\delta\}^{(s-1)/2} \left(\sum_{\rho n < y \leq n} y^{-1} \right)^{(s-1)/2} \\
&\leq \frac{1}{4\delta\sqrt{ab}} \{(1+o(1))\delta \log(1/\rho)\}^{(s-1)/2}.
\end{aligned}
$$

Now we have chosen ρ so that $(1+o(1))\delta \log(1/\rho) = (1+o(1))(1-\eta/2) < 1-\eta/3$. Hence, when we sum $S(s)$ over s, we obtain

$$
\sum_s S(s) \leq \frac{(2+o(1))\delta}{\sqrt{ab}} 3\eta^{-1} + \frac{1}{4\delta\sqrt{ab}} 3\eta^{-1} \leq \frac{1}{\delta\eta\sqrt{ab}}.
$$

We have argued above that this quantity is an upper bound on the probability that a given late vertex a is connected to a given early vertex b by a path involving no intermediate early vertices. Summing over b, the probability that a given late vertex a is connected by a path to *some* early vertex is at most

$$
\sum_{b=1}^{\rho n} \frac{1}{\delta\eta\sqrt{ab}} \leq \frac{1}{\delta\eta\sqrt{a}} \int_{b=0}^{\rho n} b^{-1/2} \mathrm{d}b = \frac{2}{\delta\eta} \sqrt{\frac{\rho n}{a}}.
$$

Finally, summing over a, the expected number of late vertices joined to early vertices is at most

$$\frac{4\rho^{1/2}}{\delta\eta}n. \tag{15}$$

Using the observation that any giant component **whp** contains some early vertex, we obtain an upper bound on the expected size of the largest component:

$$\mathbf{E}(C_1(G)) \le \frac{4\rho^{1/2}}{\delta\eta}n + \rho n + o(n) \le \frac{5}{\delta\eta}\rho^{1/2}n = \frac{5}{\delta\eta}\exp\left(-\frac{1-\eta/2}{2\delta}\right)n,$$

if δ is small enough and n is large. For η fixed we may absorb the prefactor into the exponential, obtaining that for δ small enough,

$$\mathbf{E}(C_1(G)) \le \exp\left(-\frac{1-\eta}{2\delta}\right)n = \exp\left(-\frac{1-\eta}{2\sqrt{\varepsilon}}\right)n,$$

recalling that $\delta = \sqrt{\varepsilon}$.

So far, we have obtained a bound on the expected size of the giant component. Our aim is to show that

$$C_1(G) \le \frac{4}{\delta\eta}\rho^{1/2}n + o(n) + \rho n \le \exp\left(-\frac{1-\eta}{2\sqrt{\varepsilon}}\right)n \tag{16}$$

holds **whp**; to deduce this we shall have to work a little.

Let us define a *late-early* path to be a path $x_0x_1\cdots x_i$ in G where $x_0,\dots,x_{i-1}$ are late vertices and x_i is an early vertex. Let X_i be the number of late-early paths in G having length i. Note that since, **whp**, any giant component of G contains some early vertex,

$$C_1(G) \le \rho n + \sum_{i\ge 1} X_i \tag{17}$$

holds **whp**. For each fixed i, the expectation of X_i is given by a multiple sum:

$$\mathbf{E}(X_i) \le \sum_{\rho n < x_0 \le n} \cdots \sum_{\rho n < x_{i-1} \le n} \sum_{1 \le x_i \le \rho n} \prod_{j=0}^{i-1} \frac{c}{\max\{x_j, x_{j+1}\}}.$$

In fact, the inequality above is an equality if the sum is restricted to distinct $x_0,\dots,x_i$. It is easy to see that this sum is very well approximated by an integral, and hence that

$$\mathbf{E}(X_i) = (\theta_i + o(1))n, \tag{18}$$

where

$$\theta_i = \int_{x_0=\rho}^{1} \cdots \int_{x_{i-1}=\rho}^{1} \int_{x_i=0}^{\rho} \prod_{j=0}^{i-1} \frac{c}{\max\{x_i, x_{i+1}\}} \mathrm{d}x_i \mathrm{d}x_{i-1} \cdots \mathrm{d}x_0.$$

Also, if ρn is an integer, then the multiple sum is strictly bounded by n times the integral, so

$$\mathbf{E}(X_i) \le \theta_i n. \tag{19}$$

Note that unlike relation (18), the bound (19) contains no $o(1)$ error term. In the argument that follows we shall need the exact bound (19). Thus we cannot simply replace ρn by $\lfloor \rho n \rfloor$; instead we must ensure that ρn is an integer. We do this as follows: modifying the definition of ρ by $o(1)$, we shall take ρ to be rational, with denominator d, say. We shall prove (16) for all sufficiently large values of n that are multiples of d. For intermediate values $n = (k-1)d + 1, \ldots, kd-1$, let us define early vertices to be those with index at most $\rho k d$, noting that there are $(\rho + o(1))n$ such vertices. As n increases from $(k-1)d+1$ to kd, new late vertices are added, and the number of late-early paths increases. Thus the bound (16) for general n follows from the bound for multiples of d.

In the argument that follows the form of θ_i will be irrelevant; all that matters is that θ_i is a constant, independent of n.

Let us write θ for $\frac{4\rho^{1/2}}{\delta\eta}$. From (15), the total expected number of late-early paths is at most θn. It follows using (18) that for any fixed L, we have $\sum_{i=1}^{L} \theta_i \le \theta$. Hence $\sum_{i=1}^{\infty} \theta_i \le \theta$, and, for any $L = L(n) \to \infty$, $\sum_{i>L} \theta_i = o(1)$. Thus, from (19), $\mathbf{E}(\sum_{i>L} X_i) = o(n)$. Hence, **whp** there are $o(n)$ late-early paths of length longer than $L(n)$.

To complete our proof of (16), it suffices to show that for any *fixed* i, $X_i \le (\theta_i + o(1))n$ holds **whp**. This implies that there is *some* function $L = L(n) \to \infty$ such that $\sum_{i\le L} X_i \le (\sum_{i \le L} \theta_i)n + o(n)$ holds **whp**. Since $\sum_{i>L} \theta_i = o(1)$ and **whp** we have $\sum_{i>L} X_i = o(n)$, it follows that $\sum_i X_i \le (\theta + o(1))n$ holds **whp**. Using (17), (16) follows.

From now on, we consider a *fixed* $i \ge 1$. To establish the required concentration of X_i we use a simple trick. Note that the expected degree of any late vertex in G is $O(1)$, and that, rather crudely, **whp** every late vertex has degree at most $(\log n)^2$. Let us write $Y_i = Y_i(G)$ for the maximum number y such that there are y late-early paths $P_1, \ldots, P_y$ of length i in G whose union forms a graph in which every late vertex has degree at most $(\log n)^2$. Then **whp** we have $X_i = Y_i$, so it suffices to establish concentration of Y_i. But this follows by standard martingale methods. Starting with the subgraph of G induced by the early vertices (which is irrelevant for calculating Y_i), we uncover G in $n - \rho n$ stages, at each stage deciding to which set of earlier vertices a certain late vertex j is joined.

More formally, let $G[t]$ be the subgraph of G induced by the first t vertices, and let

$$Z_t = Z_{t,i} = \mathbf{E}(Y_i \mid G[t]).$$

Then the random variable $Z_{\rho n}$ is constant (as Y_i does not depend on edges between early vertices), and hence equal to $\mathbf{E}(Y_i)$. We claim that the martingale differences $Z_{t+1} - Z_t$, $\rho n \le t \le n-1$, are bounded by $\Delta = i(\log n)^{2i}$. Indeed with $G[t]$ fixed, for any two possible values $G_1[t+1]$, $G_2[t+1]$ of $G[t+1]$, the values of Z_{t+1} conditioned on $G[t+1] = G_j[t+1]$ differ by at most Δ: we can construct coupled random graphs G_1, G_2 on $[n]$ extending $G_j[t+1]$ so that G_1, G_2 differ only in the edges from vertex $t+1$ to earlier vertices. In the collection $P_1, \ldots, P_y$ of paths defining $Y_i(G_j)$, at most Δ paths can pass through any given late vertex, so the reduction in $Y_i(G_j)$ associated with deleting all edges in G_j from vertex $t+1$ is at most Δ, and $|Y_i(G_1) - Y_i(G_2)| \leq \Delta$. Hence the values of the conditional expectation Z_{t+1} on $\{G : G[t+1] = G_j[t+1]\}$ differ by at most Δ. Since $Z_t = \mathbf{E}(Z_{t+1} \mid G[t])$, the claim follows.

Now i is fixed, so

$$\sum_{j=\rho n}^{n-1} \left(i(\log n)^{2i}\right)^2 = o(n^2),$$

and it follows from the Hoeffding-Azuma inequality that $Z_n = Z_{\rho n} + o(n)$ holds **whp**, i.e., that $Y_i = \mathbf{E}(Y_i) + o(n)$ **whp**. As X_i and Y_i are bounded by n, $X_i = Y_i$ **whp**, and $\mathbf{E}(X_i) \sim \theta_i n$, it follows that $X_i = (\theta_i + o(1))n$ holds **whp**. As noted above, this completes the proof of (16), and hence of Theorem 4. □

We have proved that an upper bound of the form $\exp(-\Theta(1/\sqrt{\varepsilon}))n$ on the size of the largest component of $G_n(1/4+\varepsilon)$ holds **whp**. Although the form of the bound is correct, the constant very likely is not. In fact it seems likely, from the work of Dorogovtsev, Mendes and Samukhin [9], that the constant $\pi/2$ in Theorem 3 is best possible.

7 A continuous problem

In the next two sections we prove Theorem 3, giving a lower bound on the rate of emergence of the giant component above $c = 1/4$.

The idea here (as in [6]) is to consider the neighbourhood expansion process in $G_n(c)$. At the critical probability, we expect this to be a critical branching process. If there were only finitely many types of vertices, with the same number of each, and vertices were joined independently with probability depending on their types, we could keep track of the number of neighbours of each type, obtaining a finite type branching process. This would be critical when the corresponding probability matrix had maximum eigenvalue one. Here we use a continuous version of this, but there is an annoying singularity at zero. To deal with this we truncate, ignoring early vertices. All rigorous comparisons with $G_n(c)$ come in the next section - this paragraph merely motivates studying the following operator.

For $u < v$ let $C[u, v]$ be the Banach space of continuous functions on $[u, v]$ with the supremum norm. Let $0 < \rho < 1$ be fixed, and consider the operator $T_\rho : C[\rho, 1] \to C[\rho, 1]$ given by

$$T(f)(x) = \int_{y=\rho}^{x} \frac{f(y)}{x} \mathrm{d}y + \int_{y=x}^{1} \frac{f(y)}{y} \mathrm{d}y.$$

We aim to show that T has an eigenvalue that approaches 4 from below as $\rho \to 0$. Kalikow and Weiss [15] showed that the corresponding $n \times n$ matrix A, with entries $a_{ij} = 1/\max\{i,j\}$, has maximum eigenvalue at most 4. Here we shall need the lower bound. This was given for a more general form of the matrix A by Durrett and Kesten [11], but *without truncation*; the matrix corresponds to our operator for $\rho = 0$. Actually, Durrett and Kesten do consider truncation, showing continuity as $\rho \to 0$, but giving no result about the rate of convergence. Here, the rate of convergence is crucial. As we shall need to truncate anyway, it turns out to be best to analyze exactly the $\rho > 0$ case. If one follows through the methods of [11] to obtain a lower bound on the giant component in $G_n(1/4+\varepsilon)$, the wrong power of ε is obtained in the exponent.

Throughout we work with the reciprocal c of the eigenvalue, which will correspond to the critical c in a truncated form of $G_n(c)$. Thus we consider the equation

$$f = cT(f). \tag{20}$$

Writing $g(x) = xf(x)$, equation (20) becomes

$$g(x) = c\int_{y=\rho}^{x} \frac{g(y)}{y}\mathrm{d}y + cx\int_{y=x}^{1} \frac{g(y)}{y^2}\mathrm{d}y.$$

Differentiating, we obtain

$$g'(x) = c\frac{g(x)}{x} + c\int_{y=x}^{1} \frac{g(y)}{y^2}\mathrm{d}y - cx\frac{g(x)}{x^2} = c\int_{y=x}^{1} \frac{g(y)}{y^2}\mathrm{d}y. \tag{21}$$

Differentiating again:

$$g''(x) = -c\frac{g(x)}{x^2}.$$

Setting $c = 1/4 + \delta^2$, the general solution to this equation is

$$g(x) = x^{1/2}\left(A\cos(\delta\log x) + B\sin(\delta\log x)\right).$$

From (21) we have $g'(1) = 0$, so, normalizing, we may take $A = 2\delta$, $B = -1$, obtaining

$$f(x) = x^{-1/2}\left(2\delta\cos(\delta\log x) - \sin(\delta\log x)\right) \tag{22}$$

as a potential eigenfunction; there is still an unchecked boundary condition. To enforce this condition, and as we shall need the integral anyway, let us evaluate $\int_{x=\rho}^{1} f(x)\mathrm{d}x$. Substituting $x = e^{-t}$, so that $\mathrm{d}x = -e^{-t}\mathrm{d}t$, this becomes

$$\begin{aligned}\int_{t=0}^{\log(1/\rho)} f(e^{-t})e^{-t}\mathrm{d}t &= \int_{t=0}^{\log(1/\rho)} e^{-t/2}\left(2\delta\cos(\delta t) + \sin(\delta t)\right)\mathrm{d}t \\ &= \left[\frac{e^{-t/2}}{1/4+\delta^2}\left(-2\delta\cos(\delta t) + (2\delta^2 - 1/2)\sin(\delta t)\right)\right]_{t=0}^{t=\log(1/\rho)}. \end{aligned} \tag{23}$$

From (20) we have $f(x) = cT(f)(x)$ for every $x \in [\rho, 1]$. Taking $x = 1$ we obtain

$$2\delta = (1/4 + \delta^2)\int_{y=\rho}^{1} f(y)\mathrm{d}y \tag{24}$$

as our second boundary condition. To solve this, let us write $R = \log(1/\rho)$, and note that the $t = 0$ term from the final bracket in (23) gives exactly $2\delta/(1/4+\delta^2)$. Thus we must solve

$$-2\delta\cos(\delta R) + (2\delta^2 - 1/2)\sin(\delta R) = 0,$$

giving R as the (minimal) positive solution to

$$\tan(\delta R) = \frac{2\delta}{2\delta^2 - 1/2},$$

or, using $\tan(2\theta) = 2\tan(\theta)/(1 - \tan(\theta)^2)$,

$$R = 2\delta^{-1}\tan^{-1}\left(\frac{1}{2\delta}\right). \tag{25}$$

We started with $\rho = \exp(-R)$ fixed, in which case one can view this equation as defining δ. Actually, it will be more convenient to think of δ as given, and R and ρ as functions of δ. Note that $R = 2\delta^{-1}(\pi/2 - \tan^{-1}(2\delta)) = \pi/\delta + O(1)$ as $\delta \to 0$, and $\rho = \exp(-(\pi + o(1))/\delta)$.

It turns out that for our purposes we can ignore any questions of uniqueness of solutions, or maximality of eigenvalues. All we shall use is that, taking R as above and $\rho = \exp(-R)$, the function $f(x)$ given by (22) is strictly positive, and solves (20) with $c = 1/4 + \delta^2$. We have in fact shown that $f(x)$ satisfies (20) for this value of c; in any case, this is easy to verify by integration using the substitution $y = e^{-s}$. For positivity, note that

$$f(x) = x^{-1/2}\left(2\delta\cos(\delta\log(1/x)) + \sin(\delta\log(1/x))\right),$$

and $\delta\log(1/x)$ ranges from 0 to $\pi - 2\tan^{-1}(2\delta)$, whereas the expression above first crosses zero when $\delta\log(1/x)$ reaches $\pi - \tan^{-1}(2\delta)$.

To collect the results we need, let $\phi = ((1/4+\delta^2)/(2\delta))f$ be the normalized version of f. Then ϕ is strictly positive on $[\rho, 1]$ and satisfies

$$\phi = (1/4 + \delta^2)T(\phi).$$

Also, from (24) we have

$$\int_{x=\rho}^{1} \phi(x)\mathrm{d}x = 1.$$

Finally, from the form of T we see that ϕ is decreasing, so

$$\sup_x \phi(x) = \phi(\rho) = \frac{1/4+\delta^2}{2\delta}\rho^{-1/2}\left(2\delta\cos(\delta R) + \sin(\delta R)\right).$$

A little calculation shows that the bracket above is exactly 2δ, so

$$\sup_x \phi(x) = (1/4+\delta^2)\rho^{-1/2} = \exp\left(-(\pi/2 + o(1))/\delta\right),$$

where the $o(1)$ term refers to $\delta \to 0$.

8 A general lower bound for giant components

For this section we consider the giant component in a random graph model that generalizes (a truncated form of) $G_n(c)$.

Let $\kappa(x,y)$ be a positive continuous function on $[0,1]^2$, with $\kappa(x,y)=\kappa(y,x)$. Let G_n^κ be the random graph on $\{1,2,\ldots,n\}$ in which each pair i,j of vertices is joined independently, with probability $\kappa(i/n,j/n)/n$. Note that the scaling is such that the vertex degrees are of order 1 as $n\to\infty$, and the total number of edges is of order n. Indeed, from continuity, the expected degree of vertex i is

$$\int_{y=0}^{1}\kappa(x,y)\mathrm{d}y+o(1),$$

as $i,n\to\infty$ with $i/n\to x$. Similarly, the expected number of edges in the whole graph is

$$\frac{n}{2}\int_{x=0}^{1}\int_{y=0}^{1}\kappa(x,y)\mathrm{d}y\mathrm{d}x+o(n).$$

Given G_n^κ, and $c<1$, we form $G_n^\kappa(c)$ by selecting edges of G_n^κ independently with probability c. Of course, G_n^κ has the same distribution as $G_n^{c\kappa}$, but separating out the factor c will be convenient, as we wish to vary this and study the change in the size of any giant component.

The related problem where κ is defined on $(0,1]^2$ and is homogeneous of degree -1 was introduced by Durrett and Kesten [11], who found the critical probability for the emergence of a giant component. A result for a discrete version of this problem giving the size of the giant component was stated by Söderberg [20] without proof, as 'it follows by analogy to the corresponding result for the classical model'. Söderberg also describes but does not analyze a continuous version.

Note that the graph $G_n(c)$ fits into Durrett and Kesten's framework, and almost fits into ours, but not quite: the corresponding kernel $\kappa=1/\max\{x,y\}$ is not continuous (or defined) when $x=0$ or $y=0$. However, we shall be able to obtain results for $G_n(c)$ by truncating to avoid this problem. Note also that the classical random graph $G_{n,C/n}$ is just G_n^κ with $\kappa(x,y)=C$.

Our aim is to prove the following general result, described in terms of the integral operator $T_\kappa: C[0,1]\to C[0,1]$ with kernel κ, given by

$$T_\kappa(f)(x)=\int_{y=0}^{1}\kappa(x,y)f(y)\mathrm{d}y.$$

Here, as before, $C[a,b]$ is the space of continuous functions on the interval $[a,b]$. Note that as κ is continuous on a compact set, it is uniformly continuous. It follows that for $f\in C[0,1]$, $T_\kappa(f)$ is defined and is continuous.

Theorem 10. *Let $\kappa(x,y)$ be any strictly positive symmetric continuous function on $[0,1]^2$, and let $c>0$ be a constant. Suppose that $\phi\in C[0,1]$ is a strictly positive function such that $cT_\kappa(\phi)(x)\geq\phi(x)$ holds for all $x\in[0,1]$. Suppose* *that* $0 < \gamma < 1$ *and that* $c' = (1+\gamma)c \le 1$. *Then, as* $n \to \infty$, **whp** *the graph* $G_n^\kappa(c')$ *contains a component of order at least* $Cn - o(n)$, *where*

$$C = \frac{\gamma}{1+\gamma} \int_0^1 \phi(x)\mathrm{d}x \Big/ \sup_{0 \le x \le 1} \phi(x).$$

Remark 1. It is most natural to apply the theorem with ϕ an eigenfunction of T_κ, and c the reciprocal of the corresponding eigenvalue.

Remark 2. We have stated our result for the subgraph $G(c)$ obtained from a certain graph G by keeping *edges* independently with probability c. In the proof that follows it will make essentially no difference if *vertices* are kept independently with probability c, giving an induced subgraph of G. The final result will be exactly the same, bearing in mind that n remains the number of vertices in the original graph.

Proof. Throughout the proof κ, ϕ, c and γ will be fixed. Without loss of generality we may normalize so that $\int_{x=0}^1 \phi(x)\mathrm{d}x = 1$.

Our plan is to discretize, run an expanding neighbourhood argument, and then compare with a suitable finite type branching process.

We shall discretize by dividing $[0,1]$ into L parts, where L is a constant chosen large enough, depending on κ, ϕ, c and γ. We define a discrete version of T as follows: for $1 \le i,j \le L$ let

$$K_{ij}^{(L)} = \inf\{\kappa(x,y) : (i-1)/L \le x \le i/L,\ (j-1)/L \le y \le j/L\}$$

be the minimum value of κ on a small square, and let $T^{(L)}$ be the n by n matrix with entries $K_{ij}^{(L)}$. As κ is a continuous function on a compact set, it is uniformly continuous. Hence, as $L \to \infty$, the step function $\kappa^{(L)}$ taking the value $K_{ij}^{(L)}$ on the corresponding square converges uniformly to κ. Similarly, we define a discrete version of ϕ by $\phi_i^{(L)} = \phi(i/L)$, and note that the corresponding step function $\phi^{(L)}$ converges uniformly to ϕ. Hence, $cT_{\kappa^{(L)}}(\phi^{(L)})$ converges uniformly to $cT_\kappa(\phi)$. Since ϕ is bounded below away from zero, it follows that for any $\eta_1 > 0$, if L is large enough we have

$$cT_{\kappa^{(L)}}(\phi^{(L)})(x) \ge (1-\eta_1)\phi^{(L)}(x)$$

for every x, i.e.,

$$c \sum_j \frac{K_{ij}^{(L)} \phi_j^{(L)}}{L} \ge (1-\eta_1)\phi_i^{(L)} \tag{26}$$

for each $i = 1, 2, \ldots, L$.

Let us fix throughout an arbitrary $0 < \gamma_1 < \gamma$. At the end we shall let $\gamma_1 \to \gamma$; indeed we shall think of γ_1 and γ as equal for most of the proof. Recalling that $c' = (1+\gamma)c$, choosing η_1 small enough, we have $(c'/c)(1-\eta_1) =$

$(1+\gamma)(1-\eta_1) > 1+\gamma_1$. For the rest of the proof we fix any L large enough that (26) holds, and write K_{ij} for $K_{ij}^{(L)}$ and ϕ_i for $\phi_i^{(L)}$, noting that

$$c' \sum_j \frac{K_{ij}\phi_j}{L} \geq (1+\gamma_1)\phi_i. \tag{27}$$

Instead of $G_n^{\kappa}(c')$ we shall consider the random graph $G = G_n^{\kappa^{(L)}}(c')$. Now $\kappa^{(L)} \leq \kappa$ holds pointwise by definition of $\kappa^{(L)}$. Thus we may couple G and $G_n^{\kappa}(c')$ so that $G \subset G_n^{\kappa}(c')$, and any lower bound on the size of the largest component in G carries over to $G_n^{\kappa}(c')$.

In the graph G a vertex s, $1 \leq s \leq n$, has *type* $\lceil sL/n \rceil$. We write C_i for the set of vertices of type i. As usual we ignore rounding, so there are $|C_i| = n/L$ vertices of each type i, $i = 1, \ldots, L$. Edges are present independently, and vertices of types i, j are joined with probability $c'K_{ij}/n$. Hence, for a given vertex s of type i, its expected number of neighbours of type j is $c'K_{ij}/L$, with an $O(n^{-1})$ correction. Thus, the neighbourhood expansion process in G may be compared to a finite type branching process with kernel $c'K_{ij}/L$. Actually, we shall argue by hand, since it is not so convenient to analyse the branching process directly: we do not have the maximum eigenvalue, for example. We shall construct a modified branching process tuned to what we do know, i.e., to (27). Also, we shall give ourselves some elbow room.

Let $0 < \gamma_2 < \gamma_1/(1+\gamma_1)$ be arbitrary. In the end, with γ fixed we shall let $\gamma_1 \to \gamma$ and

$$\gamma_2 \to \gamma/(1+\gamma). \tag{28}$$

We shall compare with a finite type branching process with kernel $(1-\gamma_2)c'K_{ij}/L$, noting that

$$(1-\gamma_2)c' \sum_j \frac{K_{ij}\phi_j}{L} \geq (1+\eta)\phi_i, \tag{29}$$

where $\eta = (1-\gamma_2)(1+\gamma_1) - 1 > 0$, and η can be made arbitrarily small by choosing γ_2 large enough.

We consider the following constrained neighbourhood expansion process in G. At each generation $t \geq 0$ we have a set S_t of *exposed* vertices already visited during the process, or excluded for some other reason, and a current generation X_t disjoint from S_t. The nature of the process will be such that at stage t we have only 'looked at' edges with at least one end in S_t. More precisely, the event that the process follows a certain sequence of states $(S_s, X_s)_{s=0}^t$ will be independent of edges between vertices outside S_t. We shall constrain ourselves to look for at most a certain number $N_{t,i}$ of vertices in $X_t \cap C_i$ for each $t \geq 1$ and each i, where

$$N_{t,i} = \lceil (1+\eta^2)^t \phi_i \rceil. \tag{30}$$

The process starts with some initial set S_0: we choose the initial generation X_0 to consist of a single vertex of $V(G) \setminus S_0$ chosen arbitrarily.

Going from X_t to X_{t+1}, we first set $S_{t+1} = S_t \cup X_t$. We assume throughout that

$$|S_{t+1} \cap C_i| \leq \gamma_2 n/L \tag{31}$$

holds for each $i = 1, 2, \ldots, L$. We will also assume a much stronger condition on S_0, that

$$|S_0 \cap C_i| \leq \sqrt{n} \tag{32}$$

holds for all i. Given disjoint S_t and X_t such that (31) holds for S_{t+1}, we define X_{t+1} as follows, by defining $X_{t+1} \cap C_i$ for each i:

List the vertices of $C_i \setminus S_{t+1}$ in some order. (These vertices are so far indistinguishable.) Go through each of these vertices checking whether it has a neighbour in X_t, and putting it into $X_{t+1} \cap C_i$ if so, stopping when we reach $|X_{t+1} \cap C_i| = N_{t+1,i}$. If we never reach this many vertices in C_i, we say the process *fails*, and abort it. If this step succeeds for each i, then we continue to the next value of t, unless we reach $t = t_0$, a certain stopping point defined below. If we reach $t = t_0$, we say the process *succeeds*, and stop.

The key observation is as follows: the probability that the process above fails at step t decreases rapidly with t, so it is very likely that the process either fails early, or succeeds. As the process has a positive probability of succeeding, and early failures do not 'use up' too many vertices, we can keep running the process until it succeeds **whp**. When the process succeeds we have found a large component in G.

We now make this precise. Let t_0 be maximal subject to

$$\max_i \sum_{t=0}^{t_0} N_{t,i} \leq \gamma_2 n/L - \sqrt{n}. \tag{33}$$

Note that our rule for stopping ensures that at every stage condition (31) holds: from the initial condition (32) we start with at most $\sqrt{n}$ exposed vertices in any C_i, and in the process we expose at most $\sum_{t=0}^{t_0} N_{t,i}$ vertices in C_i.

Let us ignore the first step for the moment, and suppose that $t > 0$. Thus, by construction, $|X_t \cap C_i| = N_{t,i}$ for each i. Now given X_t and S_t, everything we have done so far has only depended on the presence or absence of certain edges incident with S_t. As X_t is disjoint from S_t, conditional on the process so far, each edge between X_t and $V(G) \setminus S_{t+1}$ is present in G independently with its original unconditioned probability. Thus, the expected number of edges from a fixed vertex $v \in C_i \setminus S_{t+1}$ to X_t is exactly

$$E(v) = \sum_j \sum_{w \in X_t \cap C_j} c' K_{ij}/n = \sum_j N_{t,j} c' K_{ij}/n. \tag{34}$$

Now for t large enough, the $N_{t,j}$ grow by at most a factor $1 + 2\eta^2$ each time t increases. From (33) it follows that $N_{t,j} \leq 3\eta^2 \gamma_2 n/L$ for each j, so

$$E(v) \leq c' \sum_j K_{ij} 3\eta^2 \gamma_2/L \leq 3c'\eta^2 \gamma_2 \sup_{x,y} \kappa(x, y) \leq A\eta^2$$

for some constant A. Here we use $K_{ij} \leq \sup \kappa(x,y)$, and the constant A depends on κ, ϕ, c and γ which are all fixed throughout the proof. Since $E(v) \leq A\eta^2$, which will be much less than one as we take $\eta \to 0$, and since different potential edges from v to X_t are present independently, the probability that v has a neighbour in X_t is $(1-O(\eta^2))E(v) \sim E(v)$. Hence, using the assumption (31), and the formula (34) for $E(v)$, the number y_i of vertices of $C_i \setminus S_{t+1}$ adjacent to a vertex in X_t has expectation at least

$$(1-O(\eta^2))\frac{(1-\gamma_2)n}{L}\sum_j N_{t,j}c'K_{ij}/n = (1-O(\eta^2))(1-\gamma_2)c'\sum_j \frac{K_{ij}}{L}N_{t,j}.$$

Now, for t large, we have $N_{t,j} \sim \phi_j(1+\eta^2)^t$. Thus, from (29), the expectation of y_i above is at least $(1+\eta/2)N_{t+1,i}$. Now the number of vertices in $C_i \setminus S_{t+1}$ with a neighbour in X_t has a binomial distribution, so by a Chernoff bound, for t large enough, the probability of failure in step t of our process is at most $q_t = \sum_i \exp(-\eta N_{t+1,i}/10)$, say. Since $\min_i N_{t+1,i}$ grows rapidly with t (in fact exponentially), the sum $\sum_{t=1}^{\infty} q_t$ is convergent.

Let $p_t(n)$ be the supremum of the probability of failure of our process at step t, conditioned on X_t and S_t, over all legal X_t and S_t. We have argued that for t larger than some constant independent of n, we have $p_t(n) \leq q_t$. In particular, $\sum_t p_t(n)$ is bounded by a convergent sum independent of n, and for any $t_1(n) \to \infty$ the probability $p_{\geq t_1}(n)$ that our process fails at or after step t_1 is $o(1)$ as $n \to \infty$.

For the early steps in the process, including the first, note that each $N_{t+1,i}$ is by definition $\Theta(1)$, as $t = \Theta(1)$, while the expectation of y_i is also $\Theta(1)$. Thus it is easy to check that the probability of finding the required number of neighbours at each step is bounded away from zero. It follows that for each fixed t, for n large enough we have $p_t(n) \leq p_t$, for some $p_t < 1$. Combined with the remark above, there is a non-zero lower bound p_{suc} for the probability that the process succeeds, which is independent of n, and holds for all sufficiently large n.

Let $t_1 = \lfloor \log\log n \rfloor$, and let $\omega \to \infty$ sufficiently slowly, so that $\omega p_{\geq t_1}(n) \to 0$ as $n \to \infty$. Then we may run the process ω times, after each failure accumulating all exposed vertices into the initial set S_0 for the next run, until either a run succeeds, when we stop, or a run fails after step t_1. The latter happens with probability $o(1)$. The set S_0 never gets too big, as a run stopping before step t_1 exposes $n^{o(1)}$ vertices.

It only remains to show that success does guarantee a giant component: from the definition of the process, if we succeed, then we have $|X_t \cap C_i| = N_{t,i}$ for every $1 \leq t \leq t_0$ and each i. Thus we have found a component in the graph of size at least

$$S(t_0) = \sum_i \sum_{t=0}^{t_0} N_{t,i}. \tag{35}$$

For large t we have $N_{t,i} \sim (1+\eta^2)^t \phi_i$. Comparing (33) and (35), we see that as $n \to \infty$ with everything else fixed we have

$$S(t_0) \sim \frac{\sum_i \phi_i}{\max_i \phi_i} \frac{\gamma_2 n}{L}.$$

Finally, as noted at (28), we may take γ_2 arbitrarily close to $\gamma/(1+\gamma)$. Also, as $L \to \infty$ we have $\max_i \phi_i \to \sup_x \phi(x)$ and $\sum_i \phi_i/L \to \int_{x=0}^1 \phi(x)\mathrm{d}x$, completing the proof. □

Remark. In future work we hope to obtain a stronger result along these lines using similar methods, obtaining the normalized size of the giant component in terms of the solution to an appropriate non-linear integral equation, under weaker assumptions on the kernel than boundedness on the closed set. Such a result might apply directly to the kernel $\kappa(x, y) = 1/\max\{x, y\}$ defined on $(0, 1]^2$. Of course, it is also likely that the non-linear equation cannot be solved exactly in any given case.

9 Applying the general result

In this section we apply Theorem 10 to $G_n = G_n(c)$, $c = 1/4 + \varepsilon$, to prove Theorem 3 by showing that $G_n(c)$ does indeed have a giant component, and that this has at least the required order.

To avoid problems with singularities in the kernel, we work instead with G'_n, the subgraph of G_n induced by the vertices i with index $\rho n < i \le n$. Fix any positive $\eta < 1$, and let $\delta = \sqrt{(1-\eta)\varepsilon}$. We take

$$\rho = \exp\left(-\frac{2}{\delta}\tan^{-1}\left(\frac{1}{2\delta}\right)\right).$$

For $\rho n < i < j \le n$, the edge ij is present in G'_n with probability $\kappa(i/n, j/n)/n$, where $\kappa(x, y) = 1/\max\{x, y\}$ is defined on $[\rho, 1]^2$, and strictly positive and continuous on this set.

Now Theorem 10 requires a continuous kernel defined on $[0, 1]^2$. Rather than go through the straightforward but tedious re-scaling argument, we just apply the result by extending κ to $[0, 1]^2$, setting $\kappa = 0$ outside $[\rho, 1]^2$. (The legality of this can be checked either by rescaling, or by noting that the proof goes through for this non-continuous kernel.)

From the results in Section 7 there is a positive function $\phi \in C[\rho, 1]$ (which we extend by zero to $[0, 1]$) such that $(1/4+\delta^2)T_\kappa(\phi) = \phi$, with $\phi(x)$ integrating to 1, and with

$$\sup_x \phi(x) = \exp\left(\frac{\pi + o(1)}{2\delta}\right).$$

Define γ by $1 + \gamma = (1/4 + \varepsilon)/(1/4 + \delta^2)$. Then γ is positive and tends to zero as $\eta \to 0$, with $\gamma = \Theta(\eta\varepsilon)$. We may apply Theorem 10, obtaining that **whp** the graph $G_n(1/4+\varepsilon)$ has a component of order at least

$$\frac{\gamma}{1+\gamma}\exp\left(-\frac{\pi+o(1)}{2\delta}\right)n-o(n).$$

Letting $\eta \to 0$ appropriately as $\varepsilon \to 0$, e.g., taking $\eta = \varepsilon^2$, for large n this bound is $\exp(-(\pi/2+o(1))\varepsilon^{-1/2})n$. This completes the proof of Theorem 3.

10 The second largest component

We now prove Theorem 5, giving a rather crude bound on the size of the second largest component in $G_n(c)$, $c > 1/4$.

Proof of Theorem 5. Let $c > 1/4$ be given, let η be a small positive constant, and set $n' = (1-\eta)n$. As usual we ignore rounding to integers where this does not affect the argument.

We shall again classify the vertices of $G_n(c)$ into early and late vertices. This time, a vertex i is *early* if $1 \le i \le n'$, and *late* otherwise, so there are many early vertices, and few late vertices. Applying Theorem 3 with n' in place of n, we see that **whp** the largest component C of $G_{n'}(c)$, the subgraph of $G_n(c)$ induced by the early vertices, has size $|C| \ge \varepsilon n$, for some $\varepsilon > 0$ independent of n.

We shall condition on $G_{n'}(c)$, and generate $G_n(c)$ from $G_{n'}(c)$ by adding the new vertices one by one. Suppose that C' is any component of $G_{n'}(c)$ distinct from C, with $|C'| \ge M = 100\varepsilon^{-1}\eta^{-1}\log n$. Then when the vertex j is added, it sends an edge to C with probability at least $\frac{1}{2}c\varepsilon n/j$, and, independently, sends an edge to C' with probability at least $\frac{1}{2}c|C'|/j$. (In each case the factor 1/2 is to allow for the small difference between the expected number of edges, and the probability of at least one edge.) Hence, each vertex added has probability at least $\frac{1}{4}c^2\varepsilon|C'|n/j^2 \ge c^2\varepsilon|C'|/(4n)$ of joining C to C', and the probability that C and C' are not joined in $G_n(c)$ is at most

$$\left(1-c^2\varepsilon|C'|/(4n)\right)^{\eta n} \le \exp(-c^2\varepsilon\eta|C'|/4) = o(n^{-1}).$$

Since $G_{n'}(c)$ has at most n components, we see that the event $E_1 = \{$ every component of $G_{n'}(c)$ with order at least M is joined to C in $G_n(c)$ $\}$ holds **whp**.

Thus if there are two large components in $G_n(c)$, one, that not containing C, must involve late vertices. We shall show that it must involve many late vertices. It is easy to check that $E_2 = \{$ every late vertex sends at most $10\log n$ edges to early vertices $\}$ and $E_3 = \{$ the subgraph of $G_n(c)$ induced by the late vertices contains no component of order at least $10\log n$ $\}$ both hold **whp**. For E_3, a comparison with the standard random graph $G_{n'',a/n''}$ can be used, where $n'' = n - n' = \eta n$ and $a = c\eta/(1-\eta) < 1$.

If E_1, E_2 and E_3 hold and there is a second component C_2 in $G_n(c)$ of order at least $(\log n)^4$, then this component must contain at least $(\varepsilon\eta/2000)(\log n)^2 = \gamma(\log n)^2$ late vertices. We show that the existence of such a component is very unlikely by uncovering $G_n(c)$ as follows.

First, we uncover all edges between two early vertices, determining C, and all edges between late vertices. Next, we uncover all edges between early vertices *not in* C and late vertices. Given the graph G' consisting of the edges uncovered so far, $G_n(c)$ is obtained by adding certain edges between C and late vertices. Furthermore, given G', each late vertex j is joined independently to each vertex of C with probability c/j. Hence, each late vertex is joined to some vertex of C with probability at least $c\varepsilon/2 = \Theta(1)$, and these events are independent for different late vertices.

Now any component C_2 as described above sends no edges to C, and is hence a component in G'. But G' has at most n such components, and each contains at least $\gamma(\log n)^2$ late vertices, and hence has probability at most

$$(1 - c\varepsilon/2)^{\gamma(\log n)^2} = o(n^{-1})$$

of not sending an edge to C. Hence, **whp** there is no component C_2 in $G_n(c)$ with the given properties. Since E_1, E_2 and E_3 hold **whp**, **whp** the second largest component of $G_n(c)$ has order at most $(\log n)^4$. □

The above argument was rather crude, as we were aiming for brevity rather than the best bound. Being slightly more careful, this method gives a bound of order $O((\log n)^3)$. It seems likely that the real answer is $O(\log n)$, as suggested, for example, by the exponential decay of component sizes described in [9]. In any case, the second largest component above the critical probability is much smaller than the largest below, so a 'duality principle' does not hold. This is in contrast to $G_{n,c/n}$, where the distribution of small components above the critical probability is very similar to that below. See [2].

11 Comparison with the CHKNS model

In this section we compare our model $G_n(c)$ with the CHKNS model of Callaway, Hopcroft, Kleinberg, Newman and Strogatz [8], showing that almost all the results of Section 2 carry over with appropriate normalization. Before doing so, let us note that our aims here are different from those of [8] - we aim to give rigorous proofs. While the analysis in [8] is interesting, and parts of it can be made rigorous, new ideas and much hard work would be needed to turn the arguments into mathematical proofs. For example, below the critical point only components of bounded size are considered in [8], and only their expected number, not the concentration results needed. It is not excluded that there are many vertices in intermediate components, of some size between $\log\log n$, say, and order n. Above the critical point, there are not even heuristic results in [8]; rather data points taken from numerical integration of a certain equation are used to estimate the growth rate of the giant component.

Dorogovtsev, Mendes and Samukhin [9] have given an interesting and difficult heuristic argument for an infinite order phase transition in various growing network models. Parts of their work, based on evaluating the generating function for the limiting proportion of vertices in components of each (finite) size, are rigorous, but other parts are not. It may be, however, that their method can be made rigorous with not too much work; some steps in this direction have been made by Durrett [10]. However, much remains to be done, including showing separately that almost all vertices not in 'finite size' components are in a single giant component, proving required concentration results, and analyzing the differential equation for the generating function obtained with rigorous error bounds. (As pointed out by Durrett, a factor $1/(1-\xi)$, which may in the end not be important, is omitted from their equation (C3) without comment.) As noted in Section 2, if their results could be made rigorous they would imply much stronger bounds on the size of giant component than Theorems 3 and 4, and in particular that the constant in the exponent in Theorem 3 is best possible.

As an aside, let us note that there is some (unimportant) vagueness in the description of the model considered in [9]; it is stated in Section II that b new edges (between old vertices) are added at each time step, but b is not an integer. Clearly a random number of edges must be added, with expectation b. In this case, it is easy to see that any (sensible) distribution will do, and will not affect the result. Note, however, that in a similar situation apparently minor details can be very important. In particular, we shall show in future work [7] that in the related model where edges are added only from the new vertex to old vertices, it is not just the average number of edges added from each new vertex to earlier vertices that determines the critical probability: the second moment of the distribution is also important.

The CHKNS model is defined as follows: at each time step, a new vertex is added. Then, with probability δ, two vertices are chosen uniformly at random and joined by an undirected edge. In [8] loops and multiple edges are allowed; of course the component structure of a multigraph is the same as that of the underlying simple graph, so these make no difference.

It will be more convenient to work with the following very slight modification of the CHKNS model considered by Durrett [10]: we start with no vertices. At each time step, we add a new vertex, and then join each pair of vertices with probability $\delta/\binom{t}{2}$, where t is the number of vertices, and the decisions are independent of each other for different pairs and/or different times. Let us write $F_n(\delta)$ for the graph obtained at time n, a graph on $\{1, 2, \ldots, n\}$. Note that for the moment we allow multiple edges.

In the process above, on average δ edges are added at each step. Apart from the exclusion of loops, the only difference between $F_n(\delta)$ and the CHKNS model is that a binomial number of edges is added, rather than zero or 1. For the analysis in [8], this change makes no difference. In terms of a model for a growing graph, $F_n(\delta)$ seems if anything more natural. The great advantage is that, for pairs $\{i_1, j_1\} \neq \{i_2, j_2\}$, in $F_n(\delta)$ the number of $i_1 j_1$ edges is independent of the number of $i_2 j_2$ edges.

Remark. This simplification was used by Durrett [10] in earlier work published just after the first draft of the present paper was written. Durrett [10] obtains rigorously the percolation threshold in this model, and comments on the non-rigorous work of [9]. He also gives a lower bound on the size of the giant component above the threshold, with C/ε in the exponent rather than $C/\sqrt{\varepsilon}$. He does not attempt to prove that the transition has infinite order, stating that making the arguments of Dorogovtsev, Mendes and Samukhin [9] rigorous would be a thankless task!

For any $1 \leq i < j \leq n$, the expected number of edges between i and j in $F_n(\delta)$ is exactly

$$E_n(i,j) = \sum_{s=j}^{n} \frac{2\delta}{(s-1)s} = 2\delta\left(\frac{1}{j-1} - \frac{1}{n}\right). \tag{36}$$

We would like to compare $F_n(\delta)$ with $G_n(c)$, where $c = 2\delta$; actually we will have to take c very slightly larger than 2δ. Note that, since we shall only consider the sizes of components, we may collapse each multiple edge to a single edge. For upper bounds on component sizes, the comparison is very simple, using the following lemma, combining two simple observations.

Lemma 11. *Suppose that we have two random graph models, generating random graphs $G_n^{(1)}$ and $G_n^{(2)}$ on the vertex set $\{1, 2, \ldots, n\}$. Suppose that for $k = 1, 2$, within the graph $G_n^{(k)}$ each possible edge ij, $i < j$, is present with probability $0 < p_{ij}^{(k)} < 1$, independently of all other edges, where the $p_{ij}^{(k)}$ do not depend on n. Finally, suppose that $p_{ij}^{(1)} \leq p_{ij}^{(2)}$ holds for all but a constant number of pairs $\{i,j\}$. If $C_1(G_n^{(2)}) \leq f(n)$ holds* **whp** *for some function $f(n)$, then $C_1(G_n^{(1)}) \leq f(n)$ holds* **whp** *for the same function.*

Proof. Choose j_0 such that $p_{ij}^{(1)} \leq p_{ij}^{(2)}$ for $j > j_0$, and let $G_n^{(k)}[j_0]$ denote the subgraph of $G_n^{(k)}$ induced by the vertices $\{1, 2, \ldots, j_0\}$. Now there are only finitely many (in fact $2^{\binom{j_0}{2}}$) possibilities for $G_n^{(2)}[j_0]$, and, since the $p_{ij}^{(2)}$s are bounded away from zero and one, each has positive probability. It follows that for any graph G_0 on $\{1, 2, \ldots, j_0\}$ we have

$$\Pr\Big(C_1(G_n^{(2)}) > f(n) \mid G_n^{(2)}[j_0] = G_0\Big) \to 0 \tag{37}$$

as $n \to \infty$. However, conditional on the subgraph induced by the first j_0 vertices being equal to G_0, in each of the graphs $G_n^{(k)}$, edges are present independently, and each edge is at least as likely in $G_n^{(2)}$ as in $G_n^{(1)}$. It follows from (37) and an appropriate coupling that

$$\Pr\Big(C_1(G_n^{(1)}) > f(n) \mid G_n^{(1)}[j_0] = G_0\Big) \to 0.$$

As this holds for every G_0, we have $\Pr(C_1(G_n^{(1)}) > f(n)) \to 0$, as required. □

The edge ij is present in $F_n(\delta)$ with probability at most $E_n(i,j)$, which in turn is at most $2\delta/(j-1)$, and different edges of $F_n(\delta)$ are present independently. For any $\eta > 0$, setting $c = 2\delta(1+\eta)$, there is a j_0 such that for $j > j_0$ we have $2\delta/(j-1) \leq c/j$. Hence, by Lemma 11, any upper bound on $C_1(G_n(2\delta(1+\eta)))$ that holds **whp**, also holds **whp** as an upper bound on $C_1(F_n(\delta))$.

Theorem 12. *Let $\delta < 1/8$ be fixed. Then $F_n(\delta)$ has no giant component: indeed for any $\omega \to \infty$*

$$C_1(F_n(\delta)) \le \omega\sqrt{n \log n}$$

holds **whp** *as $n \to \infty$.*

In other words, $C_1(F_n(\delta)) = O_p(\sqrt{n \log n})$.

Proof. Given δ, choose η small enough that $2\delta(1+\eta) < 1/4$. Then the bound $C_1(.) \le \omega\sqrt{n \log n}$ holds **whp** for $G_n(2\delta(1+\eta))$ by Theorem 1. Hence, from the argument above, this bound holds **whp** for $F_n(\delta)$. □

Similarly, corresponding to Theorem 4, for any $a < 1/(2\sqrt{2})$ there is an $\varepsilon(a) > 0$ such that, for $0 < \varepsilon < \varepsilon(a)$,

$$C_1(F_n(1/8+\varepsilon)) \le \exp\left(-a/\sqrt{\varepsilon}\right) n \tag{38}$$

holds **whp** as $n \to \infty$.

We now turn to the lower bound on the giant component in $F_n(\delta)$, following the proof of Theorem 3 in Sections 7, 8 and 9. Throughout we will take $\delta = 1/8 + \varepsilon$. We aim to show that for any $b > \pi/(2\sqrt{2})$, there is an $\varepsilon(b) > 0$ such that, for $0 < \varepsilon < \varepsilon(b)$,

$$C_1(F_n(1/8+\varepsilon)) \ge \exp\left(-b/\sqrt{\varepsilon}\right) n \tag{39}$$

holds **whp** as $n \to \infty$.

This time we need a lower bound on edge probabilities in $F_n(\delta)$. In fact, we have an exact formula for the probability of an edge ij, $i < j$:

$$p_n(i,j) = 1 - \prod_{s=j}^{n} \left(1 - \frac{2\delta}{(s-1)s}\right).$$

For $j \ge \sqrt{n}$, say, the probability of more than one ij edge is negligible, and we have $p_n(i,j) \sim E_n(i,j)$. Since we wish for a lower bound of the form c/j, we must deal with the $1/n$ correction in (36), which now works against us. This is important compared to $1/j$ if j is close to n, so we simply ignore vertices with indices larger than $n' = \varepsilon^2 n$. As in the proof of Theorem 3, we shall only consider vertices with indices larger than $\rho n'$, where ρ is defined as in Section 9, so (for n large enough) the condition $j \ge \sqrt{n}$ is satisfied. In summary, for $\rho n' \le i < j \le n'$ it is easy to check that $2\delta(1-2\varepsilon^2)/j$ is a strict lower bound on the probability of the edge ij in $F_n(\delta)$, and we can follow the proof of Theorem 3 with n' in place of n, taking $c = 2\delta(1-2\varepsilon^2) = 1/4 + \varepsilon + O(\varepsilon^2)$. In the end, the giant component we find will be a factor $n'/n = \varepsilon^2$ smaller than that found in $G_n(c)$. However, this does not matter: as we are aiming for a bound of the form $\exp(-A/\sqrt{\varepsilon})$ we can absorb a factor ε^2 into the exponential, changing A by $o(1)$.

Putting together the upper and lower bounds (38) and (39), we have proved the following result.

Theorem 13. *There are functions $f_1(\varepsilon)$, $f_2(\varepsilon)$ both satisfying*

$$f_i(\varepsilon) = \exp(-\Theta(1/\sqrt{\varepsilon}))$$

such that, for any $\varepsilon > 0$,

$$f_1(\varepsilon)n \leq C_1(F_n(1/8 + \varepsilon)) \leq f_2(\varepsilon)n$$

holds **whp** *as $n \to \infty$*

In other words, the model $F_n(\delta)$ does indeed display an 'infinite-order phase transition' at $\delta = 1/8$, confirming the heuristic and numerical results of [8], and the heuristic result of [9]. The latter suggests that as far as the constant in the exponent is concerned, our lower bounds (Theorem 3 and, for the CHKNS model, (39)) are essentially correct.

Remark. If we insist on working with the exact formulation of the CHKNS model given in [8], we can obtain results analogous to Theorems 12 and 13, but more work is needed to deal with the dependence between edges. This dependence arises because, given that a set S of edges is present in the CHKNS graph, these edges were added at certain times, and no other edges could have been added then. Since this dependence is purely negative, the upper bounds on components go through.

For the lower bound, we can argue as follows. Let ρ and γ be fixed small positive numbers, and set $\rho' = \rho\gamma$. Given $\delta = 1/8 + \varepsilon$, $\varepsilon > 0$, we shall take $\gamma = \Theta(\varepsilon^2)$, and then choose ρ as in Section 9. Let $L = 1/(\rho'\gamma)$, which we shall take to be an integer. (As usual, we ignore rounding where this does not affect the argument.) We divide the growth of the CHKNS graph into L steps; in each step n/L vertices are added, together with a random number of edges. Let V_k be the set $\{1, \dots, (k-1)n/L\}$ of vertices that are present at the start of step k.

In the CHKNS model, the number of edges added in step k has exactly a Binomial $\mathrm{Bi}(n/L, \delta)$ distribution. Hence, by the Chernoff bounds, there is a constant $a > 0$ such that with probability $1 - O(\exp(-an))$ we add at least $(1-\gamma)\delta n/L$ edges during step k. Some of these edges are incident with a vertex added earlier during step k, and have one or both endvertices outside V_k. Let us call these edges *bad*, and edges with both ends in V_k *good*. During step k the proportion of vertices that lie outside V_k is at most $1/k$ (this proportion is maximal at the end of the step), so the number of bad edges is stochastically dominated by a $\mathrm{Bi}(n/L, \delta 2/k)$ distribution. Hence, by a Chernoff bound, there is a constant $a' > 0$ such that for each $k \geq \rho' L = 1/\gamma$, with probability $1 - O(\exp(-a'n))$ we add at most $3\gamma\delta n/L$ bad edges during step k. Hence, **whp**, in every step $k \geq 1/\gamma$ we add at least $(1-4\gamma)\delta n/L$ good edges. Note that given that an edge added is good, its endvertices are chosen uniformly from V_k, independently of each other. (The edge may be a loop.)

Let G_1 be the original CHKNS model, and let $G_2 \subseteq G_1$ consist only of the good edges. Let G_3 be the model where in step k, $\rho' L \leq k \leq L$, we add a Poisson distributed number $N_k \sim \mathrm{Po}((1-5\gamma)\delta n/L)$ of edges, with endpoints chosen independently and uniformly from V_k. (We add no edges before step $\rho' L$; we shall not in the end consider vertices before $\rho' n$.)

By a Chernoff bound, $N_k \le (1-4\gamma)\delta n/L$ holds with probability $1 - O(\exp(-a''n))$ for some constant $a'' > 0$. Hence, we can couple the graphs so that **whp** $G_3 \subseteq G_2$.

In G_3 distinct edges are present independently. An edge ij with $i < j$ appears a Poisson number of times, with expectation

$$E_{ij} = \sum_{k=\max(\rho' L, \lceil Lj/n \rceil + 1)}^{L} \frac{(1-5\gamma)\delta n}{L} \frac{2}{|V_k|^2}$$

For $j \ge \rho' n$ we have, recalling that $|V_k| = (k-1)n/L$,

$$\begin{aligned} E_{ij} &= \sum_{k=\lceil Lj/n\rceil+1}^{L} \frac{(1-5\gamma)\delta n}{L} \frac{2L^2}{(k-1)^2 n^2} = \frac{(1-5\gamma)2\delta L}{n} \sum_{k=\lceil Lj/n\rceil+1}^{L} \frac{1}{(k-1)^2} \\ &\ge \frac{(1-5\gamma)2\delta L}{n} \sum_{k=\lceil Lj/n\rceil+1}^{L} \frac{1}{(k-1)k} \\ &= \frac{(1-5\gamma)2\delta L}{n} \left(\frac{1}{\lceil Lj/n \rceil} - \frac{1}{L} \right) \ge (1-5\gamma)2\delta \left(\frac{1}{j(1+1/(L\rho'))} - \frac{1}{n} \right) \\ &= (1-5\gamma)2\delta \left(\frac{1}{j(1+\gamma)} - \frac{1}{n} \right). \end{aligned}$$

Finally, for $i < j$ and $j \ge \rho' n$, the probability that the edge ij is present is exactly $1 - \exp(-E_{ij}) \ge E_{ij}(1 - E_{ij})$, which is at least

$$(1-1/j)(1-5\gamma)2\delta \left(\frac{1}{j(1+\gamma)} - \frac{1}{n} \right) \ge (1-6\gamma)2\delta \left(\frac{1}{j(1+\gamma)} - \frac{1}{n} \right),$$

for n sufficiently large.

For $\rho' n = \rho\gamma n \le j \le \gamma n$, this is at least $(1-9\gamma)2\delta/j$. We compare with our model $G_{\gamma n}((1-9\gamma)2\delta)$. Since the proof of our lower bound on the giant component in $G_n(c)$ only involved considering vertices with index at least ρn, all relevant edges are present with higher probability in the subgraph of G_3 induced by the first γn vertices than in $G_{\gamma n}((1-9\gamma)2\delta)$. Hence, we obtain a lower bound on the giant component of G_3 that holds **whp** for G_3, and thus **whp** for G_1. As before, changing the constant in the exponent slightly, the bound (39) is unaffected by inserting the extra factor of $\gamma = \Theta(\varepsilon^2)$, this proves the equivalent of (39) for the CHKNS model. Hence Theorem 13 holds for the CHKNS model in place of $F_n(\delta)$.

References

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

[2] B. Bollobás, *Random Graphs, Second Edition,* Cambridge studies in advanced mathematics, vol. 73, Cambridge University Press, Cambridge, 2001, xviii + 498 pp.

[3] B. Bollobás, The evolution of random graphs, *Trans. Amer. Math. Soc.* **286**, 257–274 (1984).

[4] B. Bollobás, C. Borgs, J.T. Chayes, J.H. Kim and D.B. Wilson, The scaling window of the 2-SAT transition, *Random Structures and Algorithms* **18** (2001), 201–256.

[5] B. Bollobás and O. Riordan, The diameter of a scale-free random graph, *Combinatorica* **24** (2004), 5–34.

[6] B. Bollobás and O. Riordan, Robustness and vulnerability of scale-free random graphs, *Internet Mathematics* **1** (2003), 1–35.

[7] B. Bollobás and O. Riordan, Slow emergence of the giant component in the growing m-out graph, to appear in *Random Structures and Algorithms.* (Preprint available from http://www.dpmms.cam.ac.uk/∼omr10/.)

[8] D.S. Callaway, J.E. Hopcroft, J.M. Kleinberg, M.E.J. Newman and S.H. Strogatz, Are randomly grown graphs really random?, *Phys. Rev. E* **64** (2001), 041902.

[9] S.N. Dorogovtsev, J.F.F. Mendes and A.N. Samukhin, Anomalous percolation properties of growing networks, *Phys. Rev. E* **64** (2001), 066110.

[10] R. Durrett, Rigorous result for the CHKNS random graph model, Proceedings, Discrete Random Walks 2003, Cyril Banderier and Christian Krattenthaler, Eds. *Discrete Mathematics and Theoretical Computer Science* **AC** (2003), 95–104. `http://dmtcs.loria.fr/proceedings/`

[11] R. Durrett and H. Kesten, The critical parameter for connectedness of some random graphs, in *A Tribute to Paul Erdős* (A. Baker, B. Bollobás and A. Hajnal, eds), Cambridge Univ. Press, Cambridge, 1990, pp 161–176,

[12] P. Erdős and A. Rényi, On the evolution of random graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5**, 17–61 (1960).

[13] P. Erdős and A. Rényi, On the evolution of random graphs, *Bull. Inst. Internat. Statist.* **38**, 343–347, (1961).

[14] S. Janson, D. Knuth, T. Łuczak and B. Pittel, The birth of the giant component, *Random Structures and Algorithms* **4**, 231–358 (1994).

[15] S. Kalikow and B. Weiss, When are random graphs connected?, *Israel J. Math.* **62** (1988), 257–268.

[16] R.M. Karp, The transitive closure of a random digraph, *Random Structures and Algorithms* **1**, 73–93 (1990).

[17] T. Łuczak, Component behavior near the critical point of the random graph process, *Random Structures and Algorithms* **1**, 287–310 (1990).

[18] M.E.J. Newman, The structure and function of complex networks, *SIAM Review* **45**, 167-256 (2003).

[19] L.A. Shepp, Connectedness of certain random graphs. *Israel J. Math.* **67** (1989), 23–33.

[20] B. Söderberg, General formalism for inhomogeneous random graphs, *Phys. Rev. E* **66** (2002), 066121.

[21] B. Söderberg, Random graphs with hidden colour, *Phys. Rev. E* **68** (2003), 015102(R).

[22] B. Söderberg, Properties of random graphs with hidden colour, *Phys. Rev. E* **68** (2003), 026107.

[23] T.S. Turova, Dynamical random graphs with memory, *Phys. Rev. E* **65** (2002), 066102.

[24] Y. Zhang, A power law for connectedness of some random graphs at the critical point, *Random Structures and Algorithms* **2** (1991), 101–119.