

## PERCOLATION IN A HIERARCHICAL RANDOM GRAPH

D. A. DAWSON\* AND L. G. GOROSTIZA\*\*

ABSTRACT. We study asymptotic percolation as  $N \rightarrow \infty$  in an infinite random graph  $\mathcal{G}_N$  embedded in the hierarchical group of order  $N$ , with connection probabilities depending on an ultrametric distance between vertices.  $\mathcal{G}_N$  is structured as a cascade of finite random subgraphs of (approximate) Erdős–Rényi type. However, the results are different from those of classical random graphs, e.g., the average length of paths in the giant component of an ultrametric ball is much longer than in the classical case. We give a criterion for percolation, and show that percolation takes place along giant components of giant components at the previous level in the cascade of subgraphs for all consecutive hierarchical distances. The proof involves a hierarchy of “doubly stochastic” random graphs with vertices having an internal structure and random connection probabilities.

Received: 13th March 2017 Revised: 24th April 2017 Accepted: 10th October 2017

### 1. Introduction

Random graphs have been used to analyze percolation in some infinite systems (e.g., [6, 7] and references therein). On the other hand, hierarchical structures have been used in applications in physics, biology (in particular, genetics), etc., where an underlying ultrametric distance plays a basic role (e.g., [14, 15, 16, 26] and references therein). Hence it is natural to consider infinite random graphs embedded in hierarchical structures, with the probability of connection between two vertices depending on an ultrametric distance between them. In this paper we study percolation in an infinite random graph,  $\mathcal{G}_N$ , embedded in an ultrametric group,  $\Omega_N$ , called hierarchical group of order  $N$ , as  $N \rightarrow \infty$ . The structure of  $\mathcal{G}_N$  is a cascade of (approximate) Erdős–Rényi random subgraphs at consecutive hierarchical distances, and this allows using results from the classical theory of random graphs [1, 9, 17, 18, 22]. However, we use them only as a technical tool. As we shall see, for  $\mathcal{G}_N$  there are different results from the classical ones. An essential feature of  $\mathcal{G}_N$  is that, due to the tree-type form of  $\Omega_N$ , any path connecting two vertices must contain an edge (or 1-step connection) of size equal to the hierarchical distance between them. In order to prove that percolation occurs along giant components in the cascade at consecutive distances, we introduce a family of random graphs whose vertices have an internal structure involving giant components at the previous hierarchical levels. In these “doubly stochastic”

---

2000 *Mathematics Subject Classification.* Primary 05C80; Secondary 60C05, 60K35.

*Key words and phrases.* Percolation, hierarchical random graph, Erdős–Rényi type, ultrametric, giant component.

\* This research is supported by NSERC.

\*\* Partially supported by CONACyT grant 45684-F.

random graphs the connection probabilities between vertices are random variables which are highly dependent, but asymptotically deterministic as  $N \rightarrow \infty$ . There are papers that consider random graphs with connection probabilities depending on a distance between vertices or some function of pairs of vertices (e.g., [5, 8, 10, 13]), and/or involving hierarchical structures or ultrametrics (e.g., [3, 8, 20, 21, 23, 24, 25, 27]), but we have not seen the setup we consider here.

**Definition 1.1.** The *hierarchical group of order  $N$*  (integer  $\geq 2$ ) is defined as

$$\Omega_N = \{ \mathbf{x} = (x_1, x_2, \dots) : x_i \in \{0, 1, \dots, N-1\}, i = 1, 2, \dots; \\ x_i \neq 0 \text{ only for finitely many } i \},$$

with addition componentwise mod  $N$ , i.e.,  $\Omega_N$  is the direct sum of a countable number of copies of the cyclic group of order  $N$ . The *hierarchical distance* on  $\Omega_N$  is given by

$$d(\mathbf{x}, \mathbf{y}) = \begin{cases} 0 & \text{if } \mathbf{x} = \mathbf{y}, \\ \max\{i : x_i \neq y_i\} & \text{if } \mathbf{x} \neq \mathbf{y}. \end{cases}$$

Note that  $d(\cdot, \cdot)$  is translation-invariant, and it is an ultrametric, i.e., it satisfies the strong (or non-Archimedean) triangle inequality: for any  $\mathbf{x}, \mathbf{y}, \mathbf{z}$ ,

$$d(\mathbf{x}, \mathbf{y}) \leq \max\{d(\mathbf{x}, \mathbf{z}), d(\mathbf{z}, \mathbf{y})\},$$

and for each integer  $k > 1$ , a  $k$ -ball in  $\Omega_N$  (i.e., a set of points at distance at most  $k$  from each other) contains  $N^k$  points, and it is the union of  $N$   $(k-1)$ -balls which are at distance  $k$  from each other (a 0-ball is a single point).

The group  $\Omega_N$  has been used as a state space for stochastic models in several applications (e.g., [15] and references therein).

We are interested in studying percolation properties on  $\Omega_N$ , and to this end we consider a random graph  $\mathcal{G}_N$  whose vertex set is  $\Omega_N$ , with connection probabilities depending on the hierarchical distance between points. We parameterize the connection probabilities so as to characterize the critical regime for percolation.

The random graph  $\mathcal{G}_N$  is defined as follows.

**Definition 1.2.** We define an infinite random graph  $\mathcal{G}_N$  with the points of  $\Omega_N$  as vertices, and for each  $k \geq 1$  the probability of connection between  $\mathbf{x}$  and  $\mathbf{y}$  such that  $d(\mathbf{x}, \mathbf{y}) = k$  is given by

$$\frac{c_k}{N^{2k-1}}, \tag{1.1}$$

where  $c_k$  is a positive constant (independent of  $N$ ), all connections being independent.

In this paper we do not investigate percolation for fixed  $N$ , we obtain a result for large  $N$ , i.e., asymptotic percolation as  $N \rightarrow \infty$ , by means of the theory of random graphs. This method is natural for the framework of the hierarchy of random graphs which constitutes  $\mathcal{G}_N$ .

Recall that in the classical theory of Erdős–Rényi random graphs  $\mathcal{G}(N, p)$  with  $N$  vertices and connection probability  $p$ , the right parameterization for existence of giant components as  $N \rightarrow \infty$  is  $p = c/N$ ,  $c > 1$ . Analogously, the choice of connection probabilities (1.1) will allow us to characterize percolation in  $\mathcal{G}_N$  as  $N \rightarrow \infty$  in terms of the sequence  $(c_k)$  with all  $c_k > 1$ . One of our main tools

is the *ER* theory, which we will apply recursively along consecutive hierarchical distances. Since a  $k$ -ball has  $N^k$  points, the *ER* theory would require connection probabilities  $c_k/N^k, c_k > 1$ . The fact that percolation in  $\mathcal{G}_N$  can occur with connection probabilities of the form (1.1) is due to the internal hierarchical structure of  $k$ -balls. We shall see in Section 3.2 that the average length of shortest paths between randomly chosen points in the giant component of a  $k$ -ball is of order  $(\log N)^k$ , which should be compared with  $k \log N$  in the classical case.

To begin, we observe why the theory of *ER* random graphs can be used to investigate  $\mathcal{G}_N$ . By (1.1), two different  $(k-1)$ -balls in a given  $k$ -ball are connected within the  $k$ -ball with probability

$$p^{N,k} = 1 - \left(1 - \frac{c_k}{N^{2k-1}}\right)^{N^{2(k-1)}}, \quad k > 1, \quad p^{N,1} = \frac{c_1}{N}. \quad (1.2)$$

Hence  $p^{N,k} \sim c_k/N$  as  $N \rightarrow \infty$ . More precisely, using the elementary inequality

$$0 < \frac{m}{n}y - \left[1 - \left(1 - \frac{y}{n}\right)^m\right] < \left(\frac{m}{n}\right)^2 \frac{y^2}{2}, \quad 0 < y < n, \quad m \geq 2, \quad (1.3)$$

with  $y = c_k, n = N^{2k-1}, m = N^{2(k-1)}$ , we have from (1.2)

$$p^{N,k} = \frac{c_k}{N} + o\left(\frac{1}{N}\right) \quad \text{as } N \rightarrow \infty, \quad k > 1, \quad p^{N,1} = \frac{c_1}{N}. \quad (1.4)$$

Two  $(k-1)$ -balls in a  $k$ -ball can also be connected through points outside the  $k$ -ball. However, as we shall see, this does not add more than  $o(1/N)$  to the probability of connection between the  $(k-1)$ -balls. On the other hand, we are interested in what happens inside the  $k$ -ball. Therefore we will disregard connections through points outside the  $k$ -ball. By (1.4), for any  $k > 1$ , a  $k$ -ball with its  $N$   $(k-1)$ -balls as vertices is a random graph  $\mathcal{G}(N, p^{N,k})$  which approximates the *ER* random graph  $\mathcal{G}(N, c_k/N)$  as  $N \rightarrow \infty$ . Note that (1.1) is equal to  $c_k/N$  with a norming which is the product of the sizes of two  $(k-1)$ -balls. (See Remark 2.4).

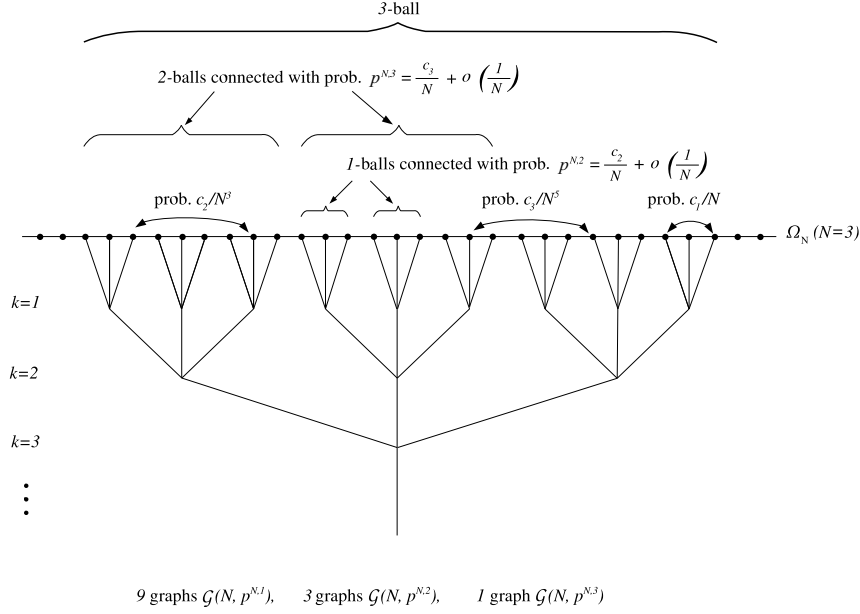
The tree representation of  $\Omega_N$  (See **Figure 1**) shows  $\mathcal{G}_N$  as a cascade of random subgraphs  $\mathcal{G}(N, p^{N,k}), k = 1, 2, \dots$ , where the form (1.4) of the connection probabilities is inherited throughout the hierarchy. Note that the hierarchical distances  $k$  are the same for all  $N$ .

We consider asymptotic percolation in  $\mathcal{G}_N$ , meaning that, as  $N \rightarrow \infty$ , there exists a path going from  $\mathbf{0} \in \Omega_N$  to infinity, i.e., to points at arbitrarily large hierarchical distances, with positive probability. This is made precise in the following definition.

**Definition 1.3.** We say that there is *asymptotic percolation* in  $\mathcal{G}_N$  if

$$P_{\text{perc}} := \inf_k \liminf_{N \rightarrow \infty} P[\mathbf{0} \in \Omega_N \text{ is connected by a path to a point at distance } k] > 0. \quad (1.5)$$

In Theorem 2.2 we give a criterion for asymptotic percolation in  $\mathcal{G}_N$ , and we show that percolation occurs along giant components of giant components at the previous level in the cascade of subgraphs  $\mathcal{G}(N, p^{N,k})$  for all hierarchical distances  $k$ . This involves replacing the connection probabilities  $p^{N,k}$ , given by (1.2), by

FIGURE 1. Tree representation of  $\Omega_N$ 

random connection probabilities between giant components, which generates a hierarchy of doubly stochastic random graphs. Although we use the basic results on the sizes of components in *ER* random graphs, most of the work in the proof of the theorem consists in formulating the problem in such a way that those results can be employed to obtain upper and lower bounds for the probability of percolation. Another part of the proof consists in showing that percolation occurs along the cascade of giant components at consecutive hierarchical levels. We will also discuss the possible form of percolation paths, and comment on approximate scale-free degree sequences in the graphs  $\mathcal{G}(N, c_k/N)$  for large  $k$ .

In a final section we make some heuristic comments regarding doubly stochastic random graphs of the type used for the proof of Theorem 2.2 and their application, in particular referring to average distance and central limit theorem for the cascade of giant components. These random graphs give rise to questions that suggest further research which would be of independent interest.

## 2. Asymptotic percolation and cascade percolation

We will use the following fundamental result on *ER* random graphs  $\mathcal{G}(N, c/N)$  (see, e.g., [22], Theorem 5.4; recall that “a.a.s.” for a graph property means that the probability that the random graph possesses the property tends to 1 as  $N \rightarrow \infty$  [22]):

(A) If  $c > 1$ , then a.a.s.  $\mathcal{G}(N, c/N)$  has a unique giant component, its size is  $\beta N$ , where  $\beta \in (0, 1)$  satisfies  $\beta = 1 - e^{-c\beta}$ , and all the other components have sizes at most  $\frac{16c}{(c-1)^2} \log N$ .

For each  $k \geq 1$  and each  $k$ -ball, we have a random graph  $\mathcal{G}(N, p^{N,k})$  whose vertices are the  $N$   $(k-1)$ -balls contained in the  $k$ -ball, with connection probability  $p^{N,k}$  given by (1.2). We assume that  $c_k > 1$  for all  $k$ , hence, because of (1.4) and (A), each  $\mathcal{G}(N, p^{N,k})$  has a unique giant component a.a.s.. We may assume that  $N$  is large enough so that the giant components have emerged in all  $k$ -balls for all  $k$ . Note that, by (1.3),  $p^{N,k} > q^{N,k}/N$ , where  $q^{N,k} = c_k(1 - c_k/2N) > 1$  for  $N > c_k^2/2(c_k - 1)$ , therefore unique giant components emerge in  $\mathcal{G}(N, p^{N,k})$  for all  $k$  as  $N \rightarrow \infty$ . Since percolation involves  $c_k \rightarrow \infty$  as  $k \rightarrow \infty$ , our method is restricted to the limiting situation  $N \rightarrow \infty$ .

If  $S$  is a non-empty set of vertices of  $\mathcal{G}(N, p^{N,k})$ , we denote by  $\|S\|$  the number of its elements (hence  $0 < \|S\| \leq N$ ), and by  $|S|$  the number of points of  $\Omega_N$  contained in  $S$ , thus  $|S| = \|S\|N^{k-1}$ .

It is reasonable to consider percolation with connections along giant components in the subgraphs  $\mathcal{G}(N, p^{N,k})$  at consecutive hierarchical distances  $k$ . Indeed, it may happen that  $\mathbf{0}$  is in the giant component of the 1-ball it belongs to, and that this 1-ball is in the giant component of the 2-ball it belongs to, but it may also happen that  $\mathbf{0}$  is not in the giant component of the 1-ball it belongs to, but that it is connected to the giant component of the 2-ball, which may still be good for percolation. However, by (A), the second possibility is negligible with respect to the first one as  $N \rightarrow \infty$ . The same argument can be made for the vertices of the subgraphs  $\mathcal{G}(N, p^{N,k})$  and any two consecutive values of  $k$ . In addition, if  $Y^{(k)}$  denotes the number of 1-step connections from a point to points at distance  $k$  from it, we shall see that

$$P[Y^{(k)} > 0] \sim c_k N^{-(k-1)}$$

as  $N \rightarrow \infty$  for each  $k \geq 1$ , so, the probability of connections at distance  $k$  is negligible w.r.t. the probability of connections at distance  $k-1$ , and all external connections from a  $k$ -ball will be at distance  $k+1$  from it for all  $k$  as  $N \rightarrow \infty$ . Therefore we will look at asymptotic percolation along giant components at consecutive distances  $k$ . Moreover, we will consider asymptotic percolation along the cascade of giant components of giant components at the previous level for all hierarchical distances  $k$ . This precisely means that  $\mathbf{0}$  is in the giant component of the 1-ball it belongs to; this giant component is connected to other giant components of 1-balls in the 2-ball in they belong to, forming a level 2 giant component; and so on for every hierarchical distance  $k$ . For brevity, we call *cascade percolation* this special form of percolation.

In order to study cascade percolation, for each  $k > 1$  and large  $N$  we consider a random graph  $\mathcal{G}_{N,k}$  whose vertices are the  $(k-1)$ -balls that are the vertices of  $\mathcal{G}(N, p^{N,k})$ , but now each vertex has an internal structure which involves all the components in the  $j$ -balls,  $j = 1, \dots, k-2$ , in the  $(k-1)$ -ball. Clearly, the giant components are the ones that will determine percolation. These random graphs are described in the proof of Theorem 2.2, and their composition is reformulated after the proof in terms of giant components alone.

For each  $k \geq 1$ , let  $\beta_k \in (0, 1)$  satisfy

$$\beta_k = 1 - e^{-c_k \beta_{k-1}^2 \beta_k}, \quad \beta_0 = 1, \quad (2.1)$$

where  $c_k \beta_{k-1}^2 > 1$ . We will show that the probability of asymptotic percolation  $P_{\text{perc}}$  is given by  $\prod \beta_k$ . Hence we need a condition for positivity of this product in terms of the  $c_k$ , which are the data of the graph  $\mathcal{G}_N$ . This is given in the following lemma (proved in the Appendix).

**Lemma 2.1.** *Assume  $c_k \nearrow \infty$  as  $k \rightarrow \infty$ ,  $c_1 > 2 \log 2$  and  $c_2 > 8 \log 2$ . Then*

$$\beta_k > 1/2 \quad \text{and} \quad c_k \beta_{k-1}^2 > 1 \quad \text{for all } k, \quad (2.2)$$

and

$$\prod_{k=1}^{\infty} \beta_k > 0 \quad \text{if and only if} \quad \sum_{k=1}^{\infty} e^{-c_k} < \infty. \quad (2.3)$$

The main result in this paper is the following theorem.

**Theorem 2.2.** *Assume  $c_k \nearrow \infty$  as  $k \rightarrow \infty$ ,  $c_1 > 2 \log 2$ ,  $c_2 > 8 \log 2$ . Then there is asymptotic percolation in  $\mathcal{G}_N$  if and only if  $\sum_{k=1}^{\infty} e^{-c_k} < \infty$ , asymptotic percolation occurs in the form of cascade percolation, and the probability of percolation (1.5) is given by*

$$P_{\text{perc}} = \prod_{k=1}^{\infty} \beta_k. \quad (2.4)$$

**Proof.** Let us assume that percolation takes place only with connections along consecutive pairs of hierarchical distances  $k$ . In the last part of the proof we will show that this is so.

We prove first that  $\sum e^{-c_k} = \infty$  implies there is no percolation in  $\mathcal{G}_N$  along consecutive hierarchical distances. By the left-hand inequality in (1.3), the connection probability  $p^{N,k}$  given by (1.2) satisfies

$$p^{N,k} < \frac{c_k}{N}.$$

Hence it suffices to show that there is no percolation along the random graphs  $\mathcal{G}(N, c_k/N)$  at distance  $k$  for every  $k$  in the weaker sense that  $\mathbf{0}$  is in the giant component of the 1-ball it belongs to, this 1-ball is in the giant component of the 2-ball it belongs to, and so on. The probability of percolation along the graphs  $\mathcal{G}(N, c_k/N)$  is given by  $\prod_{k=1}^{\infty} \gamma_k$  (connections at different distances are independent), where, by (A),  $\gamma_k > 0$  satisfies

$$\gamma_k = 1 - e^{-c_k \gamma_k} \quad (2.5)$$

for each  $k \geq 1$ . Now,  $\sum e^{-c_k} = \infty$  implies  $\sum e^{-c_k \gamma_k} = \infty$ , since all  $\gamma_k < 1$ , which by (2.5) implies  $\sum (1 - \gamma_k) = \infty$ , and therefore  $\prod \gamma_k = 0$ . Hence there is no percolation. So, percolation along consecutive hierarchical distances implies  $\sum e^{-c_k} < \infty$ .

Note that if for some  $k$  the vertex  $((k-1)$ -ball) in  $\mathcal{G}(N, c_k/N)$  containing  $\mathbf{0}$  is not in the giant component of the  $k$ -ball, then by (A) the corresponding  $\gamma_k$  in  $\prod \gamma_k$  is 0. More precisely, if the  $(k-1)$ -ball containing  $\mathbf{0}$  is not in the giant component

in the  $k$ -ball, then, by (A),  $\mathbf{0}$  is connected to at most  $O(N^{k-1} \log N)$  points within the  $k$ -ball. But the probability of connection between a set of  $N^{k-1} \log N$  points in the  $k$ -ball and points in the  $(k+1)$ -ball outside the  $k$ -ball is, by (1.1),

$$1 - \left(1 - \frac{c_{k+1}}{N^{2k+1}}\right)^{(N^{k-1} \log N)N^k(N-1)} \sim c_{k+1} \frac{\log N}{N} \rightarrow 0, \quad \text{as } N \rightarrow \infty.$$

Therefore there is no percolation. This observation allows us to omit non-giant components in the proofs.

Now we assume  $\sum e^{-c_k} < \infty$ , hence  $\prod \beta_k > 0$  by Lemma 2.1. We denote by  $Q_{\text{perc}}$  the probability of cascade percolation. We will show that

$$Q_{\text{perc}} = \prod_{k=1}^{\infty} \beta_k. \quad (2.6)$$

This obviously implies that there is asymptotic percolation in  $\mathcal{G}_N$ .

We prove first that

$$Q_{\text{perc}} \geq \prod_{k=1}^{\infty} \beta_k. \quad (2.7)$$

We proceed by steps.

For  $k = 2$ , a given 2-ball and large  $N$ , the graph  $\mathcal{G}_{N,2}$  has  $N$  vertices, which are the 1-balls in the 2-ball, and the internal structure of a 1-ball is the family of its components. Although only the giant component  $G^{(1)}$  will be relevant for percolation, the other ones complete the size of the 1-ball, and this plays a role in the proof. By (1.1), two giant components  $G_i^{(1)}, G_j^{(1)}, i \neq j$  (which are at distance 2 from each other) are connected with conditional probability (given the sizes  $|G_i^{(1)}|, |G_j^{(1)}|$ )

$$p_{ij}^{N,2} = 1 - \left(1 - \frac{c_2}{N^3}\right)^{|G_i^{(1)}||G_j^{(1)}|}. \quad (2.8)$$

We write  $p_{ij}^{N,2}$  as

$$p_{ij}^{N,2} = \frac{c_2 \beta_1^2}{N} + A_{ij}^{N,2} + B_{ij}^{N,2}, \quad (2.9)$$

where

$$A_{ij}^{N,2} = 1 - \left(1 - \frac{c_2}{N^3}\right)^{|G_i^{(1)}||G_j^{(1)}|} - \frac{|G_i^{(1)}||G_j^{(1)}|}{N^3} c_2, \quad (2.10)$$

$$B_{ij}^{N,2} = \left(\frac{|G_i^{(1)}||G_j^{(1)}|}{N^3} - \frac{\beta_1^2}{N}\right) c_2. \quad (2.11)$$

By (1.3),  $A_{ij}^{N,2} \leq 0$  and

$$|A_{ij}^{N,2}| \leq \left(\frac{|G_j^{(1)}||G_j^{(1)}|}{N^3}\right)^2 \frac{c_2^2}{2} \leq \frac{c_2^2}{2N^2},$$

since  $|G_i^{(1)}| \leq N$ . Hence

$$-\frac{c_2^2}{2N^2} \leq A_{ij}^{N,2} \leq 0. \quad (2.12)$$

We take  $0 < \delta_2 < 1$  and delete from  $\mathcal{G}_{N,2}$  the vertices (1-balls) with giant components  $G_i^{(1)}$  such that

$$|G_i^{(1)}| \leq \beta_1(1 - \delta_2)N, \quad (2.13)$$

where  $\beta_1$  satisfies (2.1) with  $k = 1$ , except the one that contains  $\mathbf{0}$  if it happens to be in this case. For a pair  $ij$  of giant components that do not satisfy (2.13) we have, by (2.11) and (2.13),

$$B_{ij}^{N,2} > \frac{c_2\beta_1^2}{N}((1 - \delta_2)^2 - 1) = \frac{c_2\beta_1^2}{N}(\delta_2^2 - 2\delta_2). \quad (2.14)$$

Then, from (2.9), (2.12) and (2.14), for the remaining giant components we have

$$p_{ij}^{N,2} > \frac{c_2\beta_1^2}{N} \left( (1 - \delta_2)^2 - \frac{c_2}{2\beta_1^2 N} \right). \quad (2.15)$$

Since  $c_2\beta_1^2 > 1$ , by Lemma 2.1, for any  $0 < \varepsilon_2 < 1$  such that  $c_2\beta_1^2(1 - \varepsilon_2) > 1$  we can take  $\delta_2$  small enough and  $N_2$  large enough so that, from (2.15),

$$p_{ij}^{N,2} > \frac{q_2^{(\varepsilon_2)}}{N}, \quad \text{where } q_2^{(\varepsilon_2)} = c_2\beta_1^2(1 - \varepsilon_2) > 1, \quad (2.16)$$

for all  $N \geq N_2$ , for all pairs  $ij$  of giant components which do not satisfy (2.13).

Let  $M^{N,2}$  denote the number of 1-balls of  $\mathcal{G}_{N,2}$  whose giant components satisfy (2.13).  $M^{N,2}$  is distributed  $\text{Bin}(N, P[|G^{(1)}| \leq \beta_1(1 - \delta_2)N])$  (the  $|G_i^{(1)}|$  are i.i.d.). Hence, for any  $0 < \eta < 1$ ,

$$\begin{aligned} P\left[\frac{M^{N,2}}{N} > \eta\right] &\leq \frac{1}{N\eta} EM^{N,2} = \frac{1}{\eta} P[|G^{(1)}| \leq \beta_1(1 - \delta_2)N] \\ &\rightarrow 0 \quad \text{as } N \rightarrow \infty, \end{aligned}$$

by (2.1) for  $k = 1$  and (A) for the *ER* graph  $\mathcal{G}(N, c_1/N)$  (see also [1], p.167). So,  $M^{N,2} = o(N)$  as  $N \rightarrow \infty$  ( $o_p(N)$  in the notation of [22], p.11). Therefore, deleting from  $\mathcal{G}_{N,2}$  the 1-balls whose giant components satisfy (2.13) does not impair the emergence of a giant component in  $\mathcal{G}_{N,2}$ , which actually emerges due to (2.16). The giant component of  $\mathcal{G}_{N,2}$  is a *level 2* giant component, whose elements are 1-balls with their internal structures.

For  $k = 3$ , a given 3-ball and large  $N$ , the graph  $\mathcal{G}_{N,3}$  has  $N$  vertices, which are the 2-balls in the 3-ball, and the internal structure of a 2-ball is the family of its components, in particular the giant component of the corresponding graph  $\mathcal{G}_{N,2}$ , as described above. Having deleted from each 2-ball an  $o(N)$  number of 1-balls as above, we are left only with 1-balls whose connection probabilities between their giant components  $p_{ij}^{N,2}$  satisfy (2.16) (and the 1-ball containing  $\mathbf{0}$ ). We now connect these 1-balls with the smaller probability  $q_2^{(\varepsilon_2)}/N$  (recall that we are looking for a lower bound for  $Q_{\text{perc}}$ ), and we consider the resulting level 2 giant components  $G_i^{(2)}$ .



By (1.1), two giant components  $G_i^{(2)}, G_j^{(2)}, i \neq j$  (which are at distance 3 from each other) are connected with probability

$$p_{ij}^{N,3} = 1 - \left(1 - \frac{c_3}{N^5}\right)^{|G_i^{(2)}||G_j^{(2)}|}, \quad (2.17)$$

where  $|G_i^{(2)}| = ||G_i^{(2)}||N$  (each 1-ball in  $G_i^{(2)}$  contains  $N$  points), and, by (A),

$$||G_i^{(2)}|| \sim \beta_2^{(\varepsilon_2)} N \quad \text{a.a.s.}, \quad (2.18)$$

where  $\beta_2^{(\varepsilon_2)} > 0$  satisfies

$$\beta_2^{(\varepsilon_2)} = 1 - e^{-q_2^{(\varepsilon_2)} \beta_2^{(\varepsilon_2)}}. \quad (2.19)$$

Now we are in a similar situation as in the previous step ( $k = 2$ ), with 1-balls playing the role of points and  $q_2^{(\varepsilon_2)}$  playing the role of  $c_1$ . The fact that  $o(N)$  1-balls have been deleted from each 2-ball has no effect, and in any case it helps towards a lower bound. So, proceeding as above with (2.17), (2.18), (2.19), we take  $0 < \delta_3 < 1$  and delete from  $\mathcal{G}_{N,3}$  the 2-balls with giant components  $G_i^{(2)}$  such that

$$||G_i^{(2)}|| \leq \beta_2^{(\varepsilon_2)}(1 - \delta_3)N, \quad (2.20)$$

(except the one containing  $\mathbf{0}$ ). Choosing  $\delta_3$  small enough and  $N_3$  large enough (and larger than  $N_2$ ), we have that the connection probabilities  $p_{ij}^{N,3}$  for the remaining giant components satisfy

$$p_{ij}^{N,3} > \frac{q_3^{(\varepsilon_2, \varepsilon_3)}}{N}, \quad \text{where} \quad q_3^{(\varepsilon_2, \varepsilon_3)} = c_3(\beta_2^{(\varepsilon_2)})^2(1 - \varepsilon_3) > 1, \quad (2.21)$$

with small enough  $\varepsilon_3$ , for all  $N \geq N_3$ . Note that since  $q_2^{(\varepsilon_2)} \nearrow c_2\beta_1^2$  as  $\varepsilon_2 \searrow 0$ , then, from (2.19),  $\beta_2^{(\varepsilon_2)} \nearrow \beta_2$  as  $\varepsilon_2 \searrow 0$ , where  $\beta_2$  satisfies (2.1) with  $k = 2$ , and since  $c_3\beta_2^2 > 1$  by Lemma 2.1, then  $\varepsilon_2$  and  $\varepsilon_3$  can be taken small enough so that  $q_3^{(\varepsilon_2, \varepsilon_3)} > 1$ .

Since  $\beta_2^{(\varepsilon_2)} < \beta_2 < \gamma_2$  (see (2.5)), we have as above that the number of deleted 2-balls from  $\mathcal{G}_{N,3}$  is  $o(N)$  as  $N \rightarrow \infty$ , and we connect the remaining ones with the smaller probability  $q_3^{(\varepsilon_1, \varepsilon_3)}/N$ .

Iterating this scheme, for each  $k$ , each given  $k$ -ball and large  $N$ , we have a graph  $\mathcal{G}_{N,k}$  whose  $N$  vertices are the  $(k-1)$ -balls in the  $k$ -ball with their internal structures, which involve the level  $(k-1)$  giant components in the corresponding graphs  $\mathcal{G}_{N,k-1}$ . A number  $o(N)$  of vertices of  $\mathcal{G}_{N,k}$  are deleted, as above, the remaining ones are connected with the smaller probability  $q_k^{(\varepsilon_2, \dots, \varepsilon_k)}/N$ , where

$$q_k^{(\varepsilon_2, \dots, \varepsilon_k)} = c_k(\beta_{k-1}^{(\varepsilon_2, \dots, \varepsilon_{k-1})})^2(1 - \varepsilon_k) > 1, \quad (2.22)$$

and  $\beta_k^{(\varepsilon_2, \dots, \varepsilon_k)} > 0$  satisfies

$$\beta_k^{(\varepsilon_2, \dots, \varepsilon_k)} = 1 - e^{-q_k^{(\varepsilon_2, \dots, \varepsilon_k)} \beta_k^{(\varepsilon_2, \dots, \varepsilon_k)}}. \quad (2.23)$$

Clearly, because of the deletions we have made and the use of smaller connection probabilities, we have

$$Q_{\text{perc}} \geq \beta_1 \prod_{k=1}^{\infty} \beta_k^{(\varepsilon_2, \dots, \varepsilon_k)} \quad (2.24)$$

(connections at different distances are independent).

Letting  $\varepsilon_2, \varepsilon_k, \dots \searrow 0$  (in this order), we have from (2.22) and (2.23):  $q_2^{(\varepsilon_2)} \nearrow c_2 \beta_1^2$ , hence  $\beta_2^{(\varepsilon_2)} \nearrow \beta_2$ , where  $\beta_2$  satisfies (2.1) for  $k = 2$ , hence  $q_3^{(\varepsilon_2, \varepsilon_3)} \nearrow c_3 \beta_2^2$ , hence  $\beta_3^{(\varepsilon_2, \varepsilon_3)} \nearrow \beta_3$ , where  $\beta_3$  satisfies (2.1) for  $k = 3$ , and so on. Therefore, we obtain (2.7) from (2.24).

Note that the cascade of giant components containing  $\mathbf{0}$  will always remain, and in any case it would not be deleted according to (2.13), (2.20), etc., a.a.s..

We now prove that

$$Q_{\text{perc}} \leq \prod_{k=1}^{\infty} \beta_k. \quad (2.25)$$

The approach is analogous to the one used for proving (2.7), but now we replace the random connection probabilities between giant components by upper bounds. Since non-giant components in  $k$ -balls have no effect on percolation, we consider only the connections between giant components.

For  $k = 2$ , from (2.9), (2.11), (2.12), and (A) for the *ER* graph  $\mathcal{G}(N, c_1/N)$  we have

$$\begin{aligned} p_{ij}^{N,2} &\leq \frac{c_2 \beta_1^2}{N} + c_2 \left| \frac{|G_i^{(1)}| |G_j^{(1)}|}{N^3} - \frac{\beta_1^2}{N} \right| \\ &= \frac{c_2 \beta_1^2}{N} + \frac{c_2}{N} \left| \frac{|G_i^{(1)}| |G_j^{(1)}|}{N^2} - \beta_1^2 \right| \\ &< \frac{c_2}{N} (\beta_1^2 + \delta_2) \quad \text{a.a.s.} \end{aligned}$$

for any  $\delta_2 > 0$ . Now we connect the 1-balls in a 2-ball with probability  $c_2(\beta_1^2 + \delta_2)/N$ , and we consider the resulting level 2 giant components  $G_i^{(2)}$  in 2-balls (we keep the notation  $G_i^{(2)}$  used before, but now the connection probabilities are different).

For  $k = 3$ , two giant components  $G_i^{(2)}, G_j^{(2)}$ ,  $i \neq j$ , in a 3-ball are connected, by (1.1), as above, with probability

$$p_{ij}^{N,3} = 1 - \left( 1 - \frac{c_3}{N^3} \right)^{|G_i^{(2)}| |G_j^{(2)}|}, \quad (2.26)$$

where  $|G_i^{(2)}| = |G_i^{(2)}|/N$ , and, by (A),

$$|G_i^{(2)}| \sim \beta_2^{(\delta_2)} N \quad \text{a.a.s.}, \quad (2.27)$$

where  $\beta_2^{(\delta_2)} > 0$  satisfies

$$\beta_2^{(\delta_2)} = 1 - e^{-q_2^{(\delta_2)} \beta_2^{(\delta_2)}}, \quad (2.28)$$

with

$$q_2^{(\delta_2)} = c_2(\beta_1^2 + \delta_2) \quad (2.29)$$

Proceeding the same way as above, it follows from (2.26)-(2.29) that

$$\begin{aligned} p_{ij}^{N,3} &\leq \frac{c_2(\beta_2^{(\delta_2)})^2}{N} + c_3 \left| \frac{|G_i^{(2)}||G_j^{(2)}|}{N^5} - \frac{(\beta_2^{(\delta_2)})^2}{N} \right| \\ &= \frac{c_3(\beta_2^{(\delta_2)})^2}{N} + \frac{c_3}{N} \left| \frac{||G_i^{(2)}|| ||G_j^{(2)}||}{N^2} - (\beta_2^{(\delta_2)})^2 \right| \\ &< \frac{c_3}{N} ((\beta_2^{(\delta_2)})^2 + \delta_3) \quad \text{a.a.s.} \end{aligned}$$

for any  $\delta_3 > 0$ . Iteration of this scheme leads to

$$Q_{\text{perc}} \leq \beta_1 \prod_{k=1}^{\infty} \beta_k^{(\delta_2, \dots, \delta_k)}, \quad (2.30)$$

and  $\beta_k^{(\delta_2, \dots, \delta_k)} \searrow \beta_k$  as  $\delta_2, \delta_3, \dots \searrow 0$ , where  $\beta_k$  satisfies (2.1). We then obtain (2.25) from (2.30).

Since  $Q_{\text{perc}}$  refers to a special form of percolation, we have  $Q_{\text{perc}} \leq P_{\text{perc}}$ . In order to see that cascade percolation is equivalent to asymptotic percolation, and therefore  $Q_{\text{perc}} = P_{\text{perc}}$ , which proves (2.4) by (2.6), it remains only to show that, as we assumed at the beginning of the proof, for each  $k$  all 1-step connections from a  $k$ -ball are in the  $(k+1)$ -ball it belongs to a.a.s.. This is done later on in Lemma 2.5.  $\square$

In the proof of Theorem 2.2 the random graph  $\mathcal{G}_{N,k}$  has  $N$  vertices which are the  $(k-1)$ -balls in a  $k$ -ball, and each  $(k-1)$ -ball has an internal structure which contains the giant and the other components in it. We replaced the connection probabilities between  $(k-1)$ -balls by the random connection probabilities between their giant components because they determine percolation, and then we replaced these random connection probabilities by upper and lower (deterministic) connection probabilities in order to obtain upper and lower bounds for  $Q_{\text{perc}}$ . The non-giant components have no part in percolation, but still each  $(k-1)$ -ball contained all its  $N$   $(k-2)$ -balls (with their internal structures), and this was relevant in the proof of the theorem (see Remark 2.4). The random graphs  $\mathcal{G}_{N,k}$  can be redefined so that their vertices contain only giant components at previous levels, with random connection probabilities between the vertices. In this sense the graphs  $\mathcal{G}_{N,k}$  are ‘‘doubly stochastic’’. This way the *cascade of giant components in a  $k$ -ball* is more clearly depicted: all the points of  $\Omega_N$  in a vertex of  $\mathcal{G}_{N,k}$  are connected, and for each  $j = 2, \dots, k$ , the giant component at level  $j$  consists of connected giant components at level  $j-1$ . We denote by  $C^{N,k}$  the size of the cascade of giant components in a  $k$ -ball (i.e., the number of points of  $\Omega_N$  in the cascade). The growth of  $C^{N,k}$  is given in the following corollary.

**Corollary 2.3** *Let  $\beta_j$  satisfy (2.1). Then*

$$C^{N,k} \sim \left( \prod_{j=1}^k \beta_j \right) N^k \quad \text{a.a.s.} \quad (2.31)$$

**Proof.** For  $k = 1$  and a giant component  $G^{(1)}$  of a 1-ball, we have

$$|G^{(1)}| \sim \beta_1 N \quad \text{a.a.s.}, \quad (2.32)$$

by (2.1) and (A).

For  $k = 2$  and a level 2 giant component  $G^{(2)}$  (with the true connection probabilities, not upper and lower bounds as before) in a 2-ball, we have similarly, by (A), on the one hand,

$$\|G^{(2)}\| \gtrsim \beta_2^{(\varepsilon_2)} N \quad \text{a.a.s.}, \quad (2.33)$$

where  $\beta_2^{(\varepsilon_2)}$  satisfies (2.19), and on the other hand,

$$\|G^{(2)}\| \lesssim \beta_2^{(\delta_2)} N \quad \text{a.a.s.}, \quad (2.34)$$

where  $\beta_2^{(\delta_2)}$  satisfies (2.28). Since  $\varepsilon_2$  and  $\delta_2$  are both arbitrarily small, we have from (2.33) and (2.34),

$$\|G^{(2)}\| \sim \beta_2 N \quad \text{a.a.s.} \quad (2.35)$$

Now, each vertex of  $G^{(2)}$  contains  $\sim \beta_1 N$  points a.a.s. (not all  $N$  points, as in the proof of Theorem 2.2), therefore, from (2.35),

$$|G^{(2)}| \sim \beta_1 \beta_2 N^2 \quad \text{a.a.s.}$$

By induction we obtain (2.31).  $\square$

**Remark 2.4** We could consider the random graphs  $\mathcal{G}_{N,k}$  with giant components alone in the proof of Theorem 2.2, but then, instead of (1.1), which is  $c_k/N(N^{k-1})^2$ , by Corollary 2.3 we would use connection probabilities

$$\frac{c_k}{N((\prod_{j=1}^{k-1} \beta_j)N^{k-1})^2}$$

between points  $\mathbf{x}$  and  $\mathbf{y}$  such that  $d(\mathbf{x}, \mathbf{y}) = k$ , in order to restrict the internal structures to the cascades of giant component in the  $(k-1)$ -balls containing the points. The approach we have taken is more natural.

To complete the proof of Theorem 2.2 it remains to prove the following lemma.

**Lemma 2.5** *For each  $k \geq 0$ , all external 1-step connections from a given  $k$ -ball are at distance  $k+1$  from it a.a.s.*

The proof of Lemma 2.5 is a special case of the following calculations, which are intended to give some idea of what percolation paths might look like as  $N \rightarrow \infty$ .

Let  $Y^{(k)}$  denote the number of 1-step connections from  $\mathbf{0} \in \Omega_N$  to points at hierarchical distance  $k \geq 1$  (or from any given point of  $\Omega_N$  to points at distance  $k$  from it; recall that  $d(\cdot, \cdot)$  is translation-invariant). Since  $Y^{(k)}$  is distributed  $\text{Bin}(N^k - N^{k-1}, c_k/N^{2k-1})$ , then

$$Y^{(1)} \Rightarrow \text{Pois}(c_1) \quad \text{as } N \rightarrow \infty, \quad (2.36)$$

and

$$Y^{(k)} \Rightarrow 0 \quad \text{as } N \rightarrow \infty, \quad k > 1. \quad (2.37)$$

In the case  $k > 1$ ,

$$P[Y^{(k)} > 0] \sim \frac{c_k}{N^{k-1}} \quad \text{as } N \rightarrow \infty, \quad (2.38)$$

and

$$P[Y^{(k)} = n | Y^{(k)} > 0] \rightarrow \begin{cases} 1, & n = 1 \\ 0, & n > 1 \end{cases} \quad \text{as } N \rightarrow \infty. \quad (2.39)$$

Indeed, using  $1 - x \sim -\log x$  as  $x \rightarrow 1$ ,

$$\begin{aligned} P[Y^{(k)} > 0] &= 1 - \left(1 - \frac{c_k}{N^{2k-1}}\right)^{N^k - N^{k-1}} \\ &\sim -(N^k - N^{k-1}) \log \left(1 - \frac{c_k}{N^{2k-1}}\right), \end{aligned}$$

and now using  $x \log(1 - a/x) \sim -a$  as  $x \rightarrow \infty$ ,  $a > 0$ ,

$$N^{k-1} P[Y^{(k)} > 0] \sim -(N^{2k-1} - N^{2k-2}) \log \left(1 - \frac{c_k}{N^{2k-1}}\right) \sim c_k,$$

which proves (2.38). For the proof of (2.39) we have

$$P[Y^{(k)} = n | Y^{(k)} > 0] = \frac{P[Y^{(k)} = n]}{P[Y^{(k)} > 0]},$$

and

$$\begin{aligned} P[Y^{(k)} = n] &= \binom{N^k - N^{k-1}}{n} \left(\frac{c_k}{N^{2k-1}}\right)^n \left(1 - \frac{c_k}{N^{2k-1}}\right)^{N^k - N^{k-1} - n} \\ &\sim \frac{(N^k - N^{k-1})!}{n!(N^k - N^{k-1} - n)!} \left(\frac{c_k}{N^{2k-1}}\right)^n, \end{aligned}$$

hence, by (2.38),

$$\begin{aligned} \frac{P[Y^{(k)} = n]}{P[Y^{(k)} > 0]} &\sim \frac{(N^k - N^{k-1})!}{n!(N^k - N^{k-1} - n)!} \frac{N^{k-1}}{c_k} \left(\frac{c_k}{N^{2k-1}}\right)^n \\ &\rightarrow \begin{cases} 1, & n = 1 \\ 0, & n > 1 \end{cases} \quad \text{as } N \rightarrow \infty. \end{aligned}$$

The results (2.36)-(2.39) show that a.a.s. almost all 1-step connections from  $\mathbf{0}$  will be at distance 1 and the degree of  $\mathbf{0}$ , and of any point, will be distributed  $\text{Pois}(c_1)$ . Also, for each  $k > 1$  the probability of 1-step connections from  $\mathbf{0}$  to points at distance  $k$  is negligible w.r.t. the probability of connections to points at distance  $k - 1$ , and given that it has connections at distance  $k$ , there is only one such connection, as  $N \rightarrow \infty$ . In a similar way it can be shown that the number of points in the giant component of any 1-ball which have external 1-step connections at distance 2 from it is asymptotically  $\text{Pois}(c_2\beta_1)$  as  $N \rightarrow \infty$ .

Similarly, let  $Y^{(k,j)}$  denote the number of 1-step connections from a given  $k$ -ball to  $k$ -balls at distance  $j \geq k + 1$  from it.  $Y^{(k,j)}$  is distributed  $\text{Bin}(N^{j-k} - N^{j-1-k}, p^{N,k,j})$ , where

$$p^{N,k,j} = 1 - \left(1 - \frac{c_j}{N^{2j-1}}\right)^{N^{2k}} \quad (2.40)$$

is the probability of 1-step connections between two  $k$ -balls at distance  $j$  from each other. In the same way as above we have

$$p^{N,k,j} \sim \frac{c_j}{N^{2(j-k)-1}} \quad \text{as } N \rightarrow \infty, \quad (2.41)$$

and it follows that

$$Y^{(k,k+1)} \Rightarrow \text{Pois}(c_{k+1}) \quad \text{as } N \rightarrow \infty, \quad (2.42)$$

$$Y^{(k,j)} \Rightarrow 0 \quad \text{as } N \rightarrow \infty, \quad j > k + 1, \quad (2.43)$$

and for  $j > k + 1$ ,

$$P[Y^{(k,j)} > 0] \sim \frac{c_j}{N^{j-k-1}} \quad \text{as } N \rightarrow \infty, \quad (2.44)$$

and

$$P[Y^{(k,j)} = n | Y^{(k,j)} > 0] \rightarrow \begin{cases} 1, & n = 1 \\ 0, & n > 1 \end{cases} \quad \text{as } N \rightarrow \infty. \quad (2.45)$$

Analogous results can be obtained, using (A), for the number of 1-step connections between a giant component in a  $k$ -ball and other such giant components at distance  $j \geq k + 1$  from it, with  $c_j$  replaced by  $c_j \beta_k^2$  in (2.41), (2.42), (2.44). It can also be shown that the number of points in the cascade of giant components in a  $k$ -ball with external 1-step connections at distance  $k + 1$  is asymptotically  $\text{Pois}(c_{k+1} \prod_{j=1}^k \beta_j)$  as  $N \rightarrow \infty$ .

From (2.36)-(2.39) and (2.42)-(2.45), we see that the same pattern described above for connections between individual points is repeated for connections between giant components of giant components at the previous level in  $k$ -balls for all  $k$ . In particular, for each  $k$  all the external 1-step connections from the giant component in a given  $k$ -ball are at distance  $k + 1$  from it a.a.s.

Note that these arguments also prove Lemma 2.5.

With these results one can also see that connections between  $(k - 1)$ -balls in a  $k$ -ball through points outside the  $k$ -ball have probability  $o(1/N)$ , as stated in the Introduction.

These results suggest the following picture for the most likely form of cascade percolation paths:  $\mathbf{0}$  is in the giant component of the 1-ball it belongs to. A  $\text{Pois}(c_2 \beta_1)$  number of points in this giant component have external single 1-step connections to points at distance 2 in such a way that the giant component is connected to the giant components of the other 1-balls, forming the level 2 giant component of the 2-ball they belong to. This level 2 giant component has a  $\text{Pois}(c_3 \beta_1 \beta_2)$  number of points with external single 1-step connections to other level 2 giant components in the 3-ball they belong to, forming a level 3 giant component. And so on along the cascade of giant components for every distance  $k$ . Each path starting from  $\mathbf{0}$  and going to infinity this way is a percolation path.

**Example.** Let  $c_k = a \log k$  for large  $k$ ,  $a > 0$ . Then there is asymptotic percolation if and only if  $a > 1$ .

**Remarks 2.6** (1) Each of the following is a sufficient condition for percolation:

$$\sum_{k=1}^{\infty} c_k^{-\delta} < \infty \quad \text{for some } \delta > 0, \quad \liminf_{k \rightarrow \infty} c_k/k > 0, \quad \liminf_{k \rightarrow \infty} (c_{k+1} - c_k) > 0.$$

(2) The degree sequence of *ER* random graphs has an approximate Poisson distribution. On the other hand, the random graphs of [2] have scale-free degree distributions. A rigorous study of scale-free random graphs, including a coupling with *ER* random graphs, is contained in [11, 12]. We will see that in the graphs  $\mathcal{G}(N, p^{N,k})$  the degrees in a neighborhood of  $c_k$  have an approximate scale-free behavior with exponent  $1/2$ , for large  $k$ .

Recall that in the random graph  $\mathcal{G}(N, p^{N,k})$  the vertices are the  $N$   $(k-1)$ -balls in a  $k$ -ball with  $p^{N,k}$  given by (1.2). Neglecting the term  $o(1/N)$  in (1.4), we consider  $\mathcal{G}(N, c_k/N)$ .

Let  $X_j$  denote the number of vertices in  $\mathcal{G}(N, c_k/N)$  of degree  $j = 0, 1, \dots$ . Then the proportion  $X_j/N$  has an approximate Poisson distribution [11] (Theorem 3), i.e., for each  $j$  and any  $0 < \varepsilon < 1$ ,

$$P \left[ (1 - \varepsilon) \frac{c_k^j e^{-c_k}}{j!} \leq \frac{X_j}{N} \leq (1 + \varepsilon) \frac{c_k^j e^{-c_k}}{j!} \right] \rightarrow 1 \text{ as } N \rightarrow \infty. \quad (2.46)$$

Let  $B_k = \{j : |j - c_k| < M\}$  for some constant  $M$ . Recall that  $c_k \rightarrow \infty$  as  $k \rightarrow \infty$ . Using the Stirling formula (in the form in [19], p.52), and the standard inequalities

$$\left(1 + \frac{x}{y}\right)^y < e^x < \left(1 - \frac{x}{y}\right)^{-y}, \quad 0 < x < y, \quad y > 1,$$

it is easy to show that for any  $0 < \delta < 1$ ,

$$(1 - \delta) \frac{1}{\sqrt{2\pi}} j^{-1/2} \leq \frac{c_k^j e^{-c_k}}{j!} \leq \frac{1}{\sqrt{2\pi}} j^{-1/2} \quad (2.47)$$

for all  $j \in B_k$  and sufficiently large  $k$ .

Combining (2.46) and (2.47) we have, for any  $0 < \varepsilon < 1$ ,

$$P \left[ (1 - \varepsilon) \frac{j^{-1/2}}{\sqrt{2\pi}} \leq \frac{X_j}{N} \leq (1 + \varepsilon) \frac{j^{-1/2}}{\sqrt{2\pi}} \right] \rightarrow 1 \text{ as } N \rightarrow \infty \quad (2.48)$$

for all  $j \in B_k$  and sufficiently large  $k$ .

### 3. Some open problems

**3.1. A class of doubly stochastic random graphs.** The essence of the proof of Theorem 2.2 is that the random dependent connection probabilities  $p_{ij}^{N,k}$  between giant components of  $(k-1)$ -balls in a  $k$ -ball obey

$$p_{ij}^{N,k} \sim \frac{c_k \beta_{k-1}^2}{N} \text{ a.a.s., } c_k \beta_{k-1}^2 > 1, \quad (3.1)$$

in the sense that

$$\frac{c_k \beta_{k-1}^2}{N} (1 - \varepsilon) < p_{ij}^{N,k} < \frac{c_k \beta_{k-1}^2}{N} (1 + \varepsilon) \quad (3.2)$$

a.a.s. for each  $k$  and arbitrary  $\varepsilon$  (such that  $c_k \beta_{k-1}^2 (1 - \varepsilon) > 1$ ). Therefore asymptotic percolation in  $\mathcal{G}_N$  is as if it took place along the giant components of the *ER* random graphs  $\mathcal{G}(N, c_k \beta_{k-1}^2 / N)$  instead of  $\mathcal{G}(N, p^{N,k})$ , considering only connections between  $(k-1)$ -balls.

This suggests the study of “doubly stochastic” random graphs of the form  $\mathcal{G}(N, \{p_{ij}^N\}_{i,j=1,\dots,N})$  with vertices  $\{v_1, \dots, v_N\}$  and connection probabilities  $p_{ij}^N$  for the vertices  $v_i$  and  $v_j$ , where the  $p_{ij}^N$  are random variables which may be dependent, such that  $p_{ij}^N \sim c/N$ ,  $c > 1$ , for all  $ij$  a.a.s.. A natural question is which results of the theory of *ER* random graphs  $\mathcal{G}(N, c/N)$  can be extended to this more general class of random graphs.

The usefulness of such extended results is illustrated in the following two problems, which could be treated if they were available.

### 3.2. Average distance in the cascade of giant components in a $k$ -ball.

Let  $D(x, y)$  denote the (graph) distance between two points  $x$  and  $y$  (i.e. the length of the shortest path connecting the two points) chosen at random in the cascade of giant components in a given  $k$ -ball,  $k > 1$ . If the two points are at hierarchical distance 1 from each other (i.e., they belong to the giant component of the same 1-ball), then  $D(x, y)$  grows like  $\log N / \log c_1$  as  $N \rightarrow \infty$ , according to Theorem 2.4.1 in [17] and (1.1). If the hierarchical distance between the two points is larger, say  $k$  (i.e., they are in different  $(k-1)$ -balls; note that this is the most likely outcome of the random choice of points, with probability  $1 - 1/N$ ), then the expected result would be that  $D(x, y)$  grows like

$$\frac{(\log N)^k}{\prod_{j=1}^k \log(c_j \beta_{j-1}^2)} \quad \text{as } N \rightarrow \infty,$$

where the  $\beta_{j-1}$  satisfy (2.1) and (2.2), provided that Theorem 2.4.1 in [17] could be extended to the random graphs with random connection probabilities  $p_{ij}^{N,k}$  obeying (3.1).

Heuristically, the reasoning is as follows. Starting from a given point in the giant component of a 1-ball, the shortest path to one of the points in the 1-ball that have external connections at distance 2 grows like  $\log N / \log c_1$  (these points are chosen at random in the giant component of the 1-ball and all the shortest paths grow the same way). This happens in every 1-ball. Next, in the giant component in the 2-ball the shortest path from a given level 1 giant component to one of the level 1 giant components that have connections at distance 3 should grow like  $\log N / \log(c_2 \beta_1^2)$ , according to (3.1), if the extended result holds. Hence the (graph) distance between two points chosen at random in the level 2 giant component of level 1 giant components in a 2-ball grows like the product  $(\log N / c_1)(\log N / \log(c_2 \beta_1^2)) = (\log N)^2 / \prod_{j=1}^2 \log(c_j \beta_{j-1}^2)$ . Etc.

In contrast, the average distance in the giant component of the *ER* random graph  $\mathcal{G}(N^k, c/N^k)$ ,  $c > 1$ , grows like  $k \log N / \log c$ . Hence, although the sizes of the giant clusters in a  $k$ -ball of  $\Omega_N$  (which has  $N^k$  points) and in the *ER* graph  $\mathcal{G}(N^k, c/N^k)$  are both of order  $N^k$  (by (A) and Corollary 2.3), the typical length of paths is longer in the hierarchical case (of order  $(\log N)^k$ ). This happens because two points chosen at random in the cascade in a  $k$ -ball are most likely hierarchical distance  $k$  apart, and the paths connecting them have to go consecutively through  $k$  hierarchical levels (by cascade percolation).



**3.3. Central limit theorem for the size of the cascade of giant components in a  $k$ -ball.** The following central limit theorem is proved in [4] for the fluctuation of the size  $|G^N|$  of the giant component  $G^N$  of a  $ER$  random graph  $\mathcal{G}(N, c/N)$ ,  $c > 1$  :

$$\frac{|G^N| - \beta N}{N^{1/2}} \Rightarrow \mathcal{N}(0, \sigma^2) \quad \text{as } N \rightarrow \infty,$$

where  $\Rightarrow$  denotes convergence in distribution, and  $\mathcal{N}(0, \sigma^2)$  is the normal distribution with mean 0 and variance  $\sigma^2 = \beta(1-\beta)/\mu^2$ , where  $\beta > 0$  satisfies  $1-\beta = e^{-c\beta}$ , and  $-\mu$  is the slope of  $1-t-e^{-ct}$  at  $t = \beta$  (see also [17], Theorem 2.5.3).

If this result could be extended to random graphs with random connection probabilities obeying (3.1), we would have the following central limit result. Recall that  $C^{N,k}$  denotes the size of the cascade of giant components in a  $k$ -ball (see Corollary 2.3). Let

$$X^{N,k} = \frac{C^{N,k} - \left(\prod_{j=1}^k \beta_j\right) N^k}{N^{k-1/2}}, \quad k = 1, 2, \dots$$

Then

$$X^{N,k} \Rightarrow \mathcal{N}\left(0, \left(\prod_{j=1}^{k-1} \beta_j\right)^2 \sigma_k^2\right) \quad \text{as } N \rightarrow \infty$$

for each  $k$ , where  $\beta_k > 0$  satisfies (2.1),  $\sigma_k^2 = \beta_k(1-\beta_k)/\mu_k^2$ , and  $-\mu_k$  is the slope of  $1-t-e^{-c_k\beta_{k-1}^2 t}$  at  $t = \beta_k$ .

We argue by induction. For  $k = 1$  the result is obtained directly by the c.l.t. of [4]. Now we assume the result is true for  $k-1$  and consider the case  $k$ . Note that  $C^{N,k} = \sum_{j=1}^{\|G^{N,k}\|} C_j^{N,k-1}$ , where  $\|G^{N,k}\|$  is the size of the giant component of  $\mathcal{G}(N, c_k\beta_{k-1}^2/N)$ , hence  $\|G^{N,k}\|$  grows like  $\beta_k N$ . We write  $X^{N,k}$  as

$$\begin{aligned} X^{N,k} &= \frac{1}{N^{k-1/2}} \left( \sum_{j=1}^{\|G^{N,k}\|} C_j^{N,k-1} - \left(\prod_{j=1}^k \beta_j\right) N^k \right) \\ &= \frac{1}{N} \sum_{j=1}^{\|G^{N,k}\|} \frac{1}{N^{(k-1)-1/2}} \left( C_j^{N,k-1} - \left(\prod_{i=1}^{k-1} \beta_i\right) N^{k-1} \right) \\ &\quad + \left(\prod_{i=1}^{k-1} \beta_i\right) \frac{\|G^{N,k}\| - \beta_k N}{N^{1/2}}, \end{aligned} \quad (3.3)$$

where  $C_j^{N,k-1}$ ,  $j = 1, \dots, \|G^{N,k}\|$ , are the sizes of the cascades in the  $(k-1)$ -balls forming the giant component of  $\mathcal{G}(N, c_k\beta_{k-1}^2/N)$ . The  $C_j^{N,k-1}$  are i.i.d, and by the induction assumption,

$$\frac{1}{N^{(k-1)-1/2}} \left( C_j^{N,k-1} - \left(\prod_{i=1}^{k-1} \beta_i\right) N^{k-1} \right) \Rightarrow \mathcal{N}\left(0, \left(\prod_{i=1}^{k-1} \beta_i\right)^2 \sigma_{k-1}^2\right) \quad (3.4)$$

for each  $j$ , as  $N \rightarrow \infty$ . It is easy to show from (3.4) that the first term on the r.h.s. of (3.3) converges to 0 in probability as  $N \rightarrow \infty$ . The result then follows

from the second term in the r.h.s. of (3.3), if the extended central limit theorem for the random graph with connection probabilities  $p_{ij}^N$  obeying (3.1) holds.

## Appendix

**Proof of Lemma 2.1** (2.1) with  $k = 1$  and  $c_1 > 2 \log 2$  imply  $\beta_1 > 1/2$ . Then  $c_2 \beta_1^2 > 2 \log 2$ , which by (2.1) with  $k = 2$  implies  $\beta_2 > 1/2$ , and so on, so  $\beta_k > 1/2$  for all  $k$ .

From (2.1) we have

$$\begin{aligned} 1 - \beta_k &= \exp\{-c_k \beta_{k-1}^2 (1 - e^{-c_k \beta_{k-1}^2 \beta_k})\} \\ &= e^{-c_k \beta_{k-1}^2} \exp\left\{\frac{1}{\beta_k} c_k \beta_{k-1}^2 \beta_k e^{-c_k \beta_{k-1}^2 \beta_k}\right\} \\ &\quad (\text{using } \beta_k > 1/2 \text{ and } x e^{-x} \leq e^{-1}) \\ &\leq C e^{-c_k \beta_{k-1}^2}, \end{aligned}$$

where  $C = e^{2e^{-1}}$ . Then, by (2.1) and  $c_k \geq c_{k-1}$ ,

$$\begin{aligned} 1 - \beta_k &\leq C \exp\{-c_{k-1} (1 - e^{-c_{k-1} \beta_{k-2}^2 \beta_{k-1}})^2\} \\ &\leq C \exp\{-c_{k-1} (1 - 2e^{-c_{k-1} \beta_{k-2}^2 \beta_{k-1}})\} \\ &\leq C e^{-c_{k-1}} \exp\{2c_{k-1} e^{-c_{k-1}/8}\} \\ &\leq C_1 e^{-c_{k-1}}, \end{aligned}$$

where  $C_1 = C e^{16e^{-1}}$ . Hence  $\sum e^{-c_k} < \infty$  implies  $\sum (1 - \beta_k) < \infty$ , and therefore  $\prod \beta_k > 0$ . The reverse inequality is clear, since all  $\beta_k < 1$ .  $\square$

**Acknowledgement.** LGG thanks the Laboratory for Research in Statistics and Probability, Carleton University, Ottawa, and the Swiss Federal Institute of Technology (ETH), Zürich, for their hospitality.

## References

1. Alon, N. and Spencer, J.H.: *The Probabilistic Method*, Wiley, (2000).
2. Albert, R. and Barabási, A.L.: Emergence of scaling in random networks, *Science* 286 (1999), 509-512.
3. Barabási, A.L. and Ravasz, E: Hierarchical organization in complex networks, *Phys. Rev. E* 67, 026112 (2003).
4. Barraez, D., Boucheron, S. and Fernández de la Vega, W.: On the fluctuations of the giant component, *Comb., Probab. and Comput.* 9 (2000), 287-304.
5. Benjamini, I. and Berger, N.: The diameter of long-range percolation clusters on finite cycles, *Random Structures Algorithms* 19, No. 2 (2001), 102-111.
6. Benjamini, I., Lyons, R. and Peres, Y: Group-invariant percolation on graphs, *Geom. Funct. Anal.* 9, No. 1 (1999), 29-66.
7. Benjamini, I. and Schramm, O.: Percolation beyond  $Z^d$ , many questions and a few answers, *Elect. Commun. Probab.* 1 (1996), 71-82.
8. Biskup, M.: On the scaling of chemical distance in long-range percolation models, *Ann. Probab.* 32, No. 4 (2004), 2938-2977.
9. Bollobás, B.: *Random Graphs*, 2nd. edition, Cambridge Univ. Press, (2001).
10. Bollobás, B., Janson, S. and Riordan, O.M.: The phase transition in inhomogeneous random graphs, *Random Structures and Algorithms* (to appear).

11. Bollobás, B. and Riordan, O.M.: Mathematical results on scale-free random graphs, in *Handbook of Graphs and Networks*, 1-34, Wiley, (2003).
12. Bollobás, B. and Riordan, O.M.: Coupling scale-free and classical random graphs, *Internet Math.* 1, No. 2, (2004), 215-255.
13. Coppersmith, D., Gamarnik, D. and Sviridenko, M.: The diameter of a long-range percolation graph, *Random Structures Algorithms* 21, No. 1 (2002), 1-13.
14. Cox, J.T., Dawson, D.A. and Greven, A.: Mutually catalytic super branching random walks: large finite systems and renormalization analysis, *Mem. Amer. Math. Soc.* 171, No. 809 (2004).
15. Dawson, D.A., Gorostiza, L.G. and Wakolbinger, A.: Hierarchical random walks, in *Asymptotic Methods in Stochastics*, (L. Horváth and B. Szyszkowicz, eds.), Fields Institute Communications 44 (2004), 173-193.
16. Dawson, D.A. Gorostiza, L.G. and Wakolbinger, A.: Hierarchical equilibria of branching populations, *Elect. J. Probab.* 9 (2004), Paper No. 12, 316-381.
17. Durrett, R.: *Random Graph Dynamics*, Cambridge Univ. Press, (2006).
18. Erdős, P. and Rényi, A.: On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* 5 (1960), 17-61.
19. Feller, W.: *An Introduction to Probability Theory and Its Applications*, 2nd. Edition, Wiley, (1962).
20. Hambly, B.M. and Jordan, J.: A random hierarchical lattice: the series parallel graph and its properties, *Adv. Appl. Probab.* 36, No. 3 (2004), 824-838.
21. Hogg, T.: Refining the phase transition in combinatorial search, *Artificial Intelligence* 81, No. 1-2 (1996), 127-154.
22. Janson, S., Łuczak, T. and Ruciński, A.: *Random Graphs*, Wiley, (2000).
23. Kleinberg, J.: Small-world phenomena and the dynamics of information, in *Advances in Neural Information Processing Systems (NIPS)* 14 (2001), 431-438.
24. Kleinberg, J.: Complex networks and decentralized search algorithms, *Proc. ICM, 2006*, Volume 3, 1019-1044, European Mathematical Society, Zurich, 2006.
25. Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N. and Barabási, A.L.: Hierarchical organization of modularity in metabolic networks, *Science* 297 (2002), 1551-1554.
26. Sawyer, S. and Felsenstein, J.: Isolation by distance in a hierarchically clustered population, *J. App. Probab.* 20 (1983), 1-10.
27. Van Custem, B. and Ycart, B.: Indexed dendograms on random dissimilarities, *J. Classification* 15, No. 1 (1998), 93-127.

D.A. DAWSON: SCHOOL OF MATHEMATICS AND STATISTICS, OTTAWA, CANADA K1S 5B6  
E-mail address: ddawson@math.carleton.ca

L.G. GOROSTIZA: CINVESTAV, MEXICO CITY, MEXICO  
E-mail address: lgorosti@math.cinvestav.mx

