The Annals of Probability
2012, Vol. 40, No. 1, 130–145
DOI: 10.1214/10-AOP610

SHARP THRESHOLD FOR PERCOLATION ON EXPANDERS

BY ITAI BENJAMINI, STÉPHANE BOUCHERON, GÁBOR LUGOSI AND RAPHAËL ROSSIGNOL

The Weizmann Institute, Université Paris-Diderot, Pompeu Fabra University and Université Paris-Sud and CNRS

We study the appearance of the giant component in random subgraphs of a given large finite graph $G = (V, E)$ in which each edge is present independently with probability p. We show that if G is an expander with vertices of bounded degree, then for any $c \in]0, 1[$, the property that the random subgraph contains a giant component of size $c|V|$ has a sharp threshold.

1. Introduction. Percolation theory studies the presence of "giant" connected components in a large graph G whose edges are deleted independently at random. The meaning of "giant" differs whether one looks at infinite or finite graphs. For the former, it means an infinite connected component while for finite graphs a giant component usually means a component of size linear in the number of vertices of the original graph G. Whether the graph is finite [see Pittel (2008), Nachmias and Peres (2010)] or infinite [see Benjamini and Schramm (1996)], symmetry of G has often played a key role in the study of percolation. In this paper, we are concerned with weakening these symmetry assumptions in the case of finite graphs, replacing them by a more geometric assumption. To do so, we follow the path of Alon, Benjamini and Stacey (2004), where finite graphs satisfying an isoperimetric inequality (the so-called "expanders") are studied without any symmetry assumption.

As Alon, Benjamini, and Stacey, we study percolation in expanders of bounded degree. Consider a finite graph $G_n = (V_n, E_n)$ with $|V_n| = n$ vertices, and let $G_n(p)$ denote the spanning subgraph of G_n obtained by retaining each edge of G_n independently with probability p. In this paper we consider "large" graphs, and the term "*with high probability*" refers to events whose probability is bounded from below by a number that tends to 1 as $n \to \infty$. For any two sets of vertices A and B in G_n, let $E_n(A, B)$ be the set of all edges with one endpoint in A and the other in B. The *edge-isoperimetric number* $c(G_n)$ of G_n, also called its *Cheeger constant*, is defined by

$$\min_{\substack{A \subset V_n: \\ 0<|A|\leq n/2}} \frac{|E_n(A, A^c)|}{|A|}.$$

Received July 2009; revised June 2010.
MSC2010 subject classifications. 60K35, 05C80.
Key words and phrases. Percolation, random graph, expander, giant component, sharp threshold.

Let $b, d > 0$ be constants. A (b, d)-*expander* is a graph $G_n = (V_n, E_n)$ such that the maximal degree in G_n is not greater than d, and $c(G_n) > b$. A d-regular expander is a (b, d)-expander which is d-regular. Theorem 2.1 in Alon, Benjamini and Stacey (2004) shows that in an expander, there is, with high probability, never more than one giant component. This allows one to speak about the giant component. A more precise statement is obtained in Theorem 2.8 in Alon, Benjamini and Stacey (2004), showing that, uniformly over p, with probability tending to 1, the second largest component cannot have size larger than $|n|^{\omega(b,d)}$ for some $\omega(b, d) < 1$. Although the same result is conjectured to hold under less stringent isoperimetric assumptions, it is important to note the absence of any symmetry assumption.

Arguably the most interesting phenomenon in random graphs is the emergence of the giant component as p is increased gradually from 0 to 1. In Alon, Benjamini and Stacey (2004), Theorem 3.2, the following result is shown:

THEOREM 1.1 [Alon, Benjamini and Stacey (2004)]. *Let* $d \geq 2$ *and let* $(G_n)_{n\geq 0}$ *be a sequence of* d*-regular* (b, d)*-expanders with girth* $(g_n) \to \infty$.

If $p > 1/(d - 1)$, *then there exists a* $c > 0$ *such that*

$$\lim_{n\to\infty} \mathbb{P}\big(G_n(p) \text{ contains a component of size at least } c|V_n|\big) = 1.$$

If $p < 1/(d - 1)$, *then for any* $c > 0$,

$$\lim_{n\to\infty} \mathbb{P}\big(G_n(p) \text{ contains a component of size at least } c|V_n|\big) = 0.$$

It is tempting to conjecture that the regularity and high-girth assumptions in Theorem 1.1 can be removed at the price of losing the precise location of the threshold, thus showing that *a giant component emerges in an interval of length* $o(1)$ *in any expander*.

CONJECTURE 1.2. *Let* G_n *be a* (b, d)*-expander. There exist* $0 < q_1 < q_2 < 1$, *depending on b and d and* $p_n^* \in [q_1, q_2]$, *such that, for every* $\varepsilon > 0$, *if* $p_n \geq p_n^* + \varepsilon$, *then there exists a* $c > 0$ *such that, with high probability,*

$$G_n(p_n) \text{ contains a component of size at least } cn,$$

and if $p_n \leq p_n^* - \varepsilon$, *then for any* $c > 0$, *with high probability,*

$$G_n(p_n) \text{ does not contain a component of size at least } cn.$$

However, it is not entirely clear whether Conjecture 1.2 can hold without a minimum of homogeneity of the underlying graph G_n. Benjamini, Nachmias and Peres (2009), Theorem 1.3, establishes Conjecture 1.2 under the additional "homogeneity" assumption of weak convergence of the sequence of graphs G_n to an infinite bounded-degree graph.

The main result of this paper is a sharp threshold result for the events "$G_n(p)$ contains a component of order at least cn" for every $c \in]0, 1[$. In a sense, it can be seen as a weakening of Conjecture 1.2. The weakness is that we cannot assert the existence of the "threshold function" p_n^*, but we can still interpret this result as the fact that *in any expander, every giant component of given proportion emerges in an interval of length* $o(1)$. This is formalized as follows.

THEOREM 1.3. *Let G_n be a (b, d)-expander and let $c \in]0, 1[$. There exist constants $q_1 = q_1(d) > 0$ and $q_2 = q_2(c) \in]q_1, 1[$ and $p_n^*(c) \in [q_1, q_2]$ such that, for every $\varepsilon > 0$.*

If $p_n \geq p_n^*(c) + \varepsilon$, then, with high probability,

$$G_n(p_n) \text{ contains a component of size at least } cn.$$

If $p_n \leq p_n^*(c) - \varepsilon$, then, with high probability,

$$G_n(p_n) \text{ does not contain a component of size at least } cn.$$

The rough idea of the proof is the following. First we show that the expansion and bounded-degree properties are sufficient to imply that the value of p for which the probability that $G_n(p)$ contains a component of size cn equals some fixed constant $\alpha \in (0, 1)$ stays bounded away from zero and one. This may be proved by putting together some arguments of Benjamini and Schramm (1996) and Alon, Benjamini and Stacey (2004). Now to show that the threshold is sharp, it suffices to prove that the threshold width is bounded by a function of n that tends to zero. The proof of this fact has two main components. First we show that around the critical probability, the derivative (with respect to p) of the expected size of the largest component is proportional to n. This is done by using Russo's lemma and the expansion property. In other words, the expected size of the largest component grows quickly (with p) around the threshold value. Then to show that the probability of existence of a component of size cn has a rapid growth around the threshold, it suffices to show that the size of the largest component is concentrated in the sense that its standard deviation is of a smaller order of magnitude than its mean. This may be achieved by applying a general bound for the variance of a function of independent Bernoulli variables due to Falik and Samorodnitsky (2007).

Note that if the underlying graph G_n is transitive, then after showing that the critical probability is bounded away from zero and one, the theorem follows immediately from a general result of Friedgut and Kalai (1996) which implies that the threshold width is at most $O(1/\log n)$. However, because of the lack of symmetry assumption, the theorem of Friedgut and Kalai is not directly applicable. Some of the main ingredients of our proof are similar to those of Friedgut and Kalai as in both proofs Russo's lemma and an appropriate variance bound (based on either hypercontractivity or a logarithmic Sobolev inequality) play a key role. We use an elegant inequality of Falik and Samorodnitsky that can also be used to prove the Friedgut–Kalai theorem.

The paper is organized as follows. In Section 2 we introduce some notation. In Section 3 we collect some arguments existing in the literature to show that the thresholds $p_n^*(c)$ are bounded away from zero and one. Finally, Section 4 is devoted to the proof of our main result, Theorem 1.3. In fact, we prove something more in Theorem 4.1, namely that the threshold width is at most of order $O((\log n)^{-1/3})$.

We end this Introduction with an open question. It would be interesting to relax the isoperimetric condition and replace the positivity of the Cheeger constant by a weaker condition of the type

$$\min_{\substack{A\subset V_n:\\ 0<|A|\le n/2}} \frac{|E(A, A^c)|}{|A|^\alpha} > 0$$

for some $\alpha < 1$.

2. Notation. Let $b > 0$ and d a positive integer. Throughout the paper, $G_n = (V_n, E_n)$ denotes a (b, d)-expander. For a subset W of vertices, we denote by $\partial_E W = E(W, W^c)$ the exterior edge-boundary of W.

Each point (or "configuration") $x \in \{0, 1\}^{E_n}$ is identified with the subgraph of G_n with vertex set V_n and edge set obtained by removing from E_n all edges e such that $x(e) = 0$. For a given $p \in [0, 1]$, we equip the space $\{0, 1\}^{E_n}$ with the product probability measure $\mu_{n,p}$ under which every $x(e)$ is independently 1 (resp., 0) with probability p [resp., $(1 - p)$]. For any function $f : \{0, 1\}^{E_n} \to \mathbb{R}$, we denote by $\mathbb{E}_{n,p}(f) = \int f(x)\, d\mu_{n,p}(x)$ the mean, and for any $\alpha \geq 1$, the norm $\|\cdot\|_{\alpha,p}$ denotes

$$\|f\|_{\alpha,p} = (\mathbb{E}_{n,p}(|f|^\alpha))^{1/\alpha}.$$

For $x \in \{0, 1\}^{E_n}$ and $i \in \mathbb{N}^*$, define $\mathcal{C}_n^{(i)} = \mathcal{C}_n^{(i)}(x)$ to be the ith largest connected component in the configuration x, and let $L_n^{(i)} = L_n^{(i)}(x)$ be the number of vertices in $\mathcal{C}_n^{(i)}$. We also denote by $C(v)$ the connected component containing a vertex $v \in V_n$.

For any $c \in]0, 1[$, the subset of $\{0, 1\}^{E_n}$ defined by $L_n^{(1)} \geq cn$ is monotone (i.e., if the inequality holds for a graph then by adding any new edge it still holds), and therefore $\mu_{n,p}\{L_n^{(1)} \geq cn\}$ is a strictly increasing polynomial of p. Thus, for any $\alpha \in [0, 1]$, we may define $p_{n,\alpha}(c)$ as the unique real number p in $[0, 1]$ such that

$$\mu_{n,p}\{L_n^{(1)} \geq cn\} = \alpha.$$

The threshold function is defined as

$$p_n^*(c) = p_{n,1/2}(c).$$

When n is clear from the context, we omit the subscript n.

3. Thresholds of giant components are bounded away from zero and one. The fact that the values $p_{n,\alpha}(c)$ are bounded away from zero and one may be proved by putting together arguments of Benjamini and Schramm (1996) and Alon, Benjamini and Stacey (2004). Here we give a self-contained proof. We also add an estimate for the decay of the probability of not having a giant component of fixed size for p close enough to 1, which seems to us interesting on its own.

PROPOSITION 3.1. *Let $c \in]0,1[$. There is a constant q_1, depending only on d and there exists $q_2(c) \in]q_1, 1[$ such that for any $\alpha \in]0,1[$, for all n large enough, $p_{n,\alpha}(c) \in]q_1, q_2(c)[$.*

Furthermore, for any $c \in]0,1[$, there are strictly positive constants C_1 and C_2, depending only on b and d, such that for every $p \geq q_2(c)$,

$$\mu_{n,p}\big(L_n^{(1)} \geq cn\big) \geq 1 - C_1 e^{-C_2 n}.$$

PROOF. The fact that $p_{n,\alpha}(c)$ is bounded away from 0 may be proved using standard branching-process arguments as follows. Fix $q_1 < 1/(d-1)$ and suppose that $p \leq q_1$. Since the degrees are bounded by d, the connected component $C(v)$ of a vertex $v \in V_n$ has a size not larger than S, where S is the total number of descendants of the root in a (sub-critical) Galton–Watson process with offspring distribution $\mathcal{B}(d-1, p)$ [except for the first offspring, which has distribution $\mathcal{B}(d, p)$]. Since the binomial distribution $\mathcal{B}(d-1, p)$ possesses exponential moments, it is well known [see, e.g, Exercise 5.22 in Lyons and Peres (2011)] that there are some values $\lambda > 0$, $M < \infty$, depending only on d and q_1, such that, for every n and $p \leq q_1$,

$$\mathbb{E}_p(e^{\lambda S}) \leq M.$$

Thus, for any $t > 0$ and $p \leq q_1$,

$$\mu_p\big(L_n^{(1)} > t\big) \leq nMe^{-\lambda t}. \tag{1}$$

In particular,

$$\mu_p\bigg(L_n^{(1)} > \frac{2}{\lambda}\log(nM^{1/2})\bigg) \leq \frac{1}{n}.$$

Thus, for any $\alpha \in]0,1[$, for n large enough, $p_{n,\alpha}(c) > q_1$.

The fact that $p_{n,\alpha}(c)$ is bounded away from 1 follows essentially from Theorem 2 and Remark 2 in Benjamini and Schramm (1996) and Lemma 2.2 and the proof of Proposition 3.1 in Alon, Benjamini and Stacey (2004). To detail the proof, first we show that there are constants $p_0(b) < 1$, $a(b) > 0$ and $C(b,d) > 0$ such that for every $p \geq p_0$ and every $n \in \mathbb{N}^*$,

$$\mu_p\big(L_n^{(1)} \geq an\big) \geq 1 - e^{-Cn}. \tag{2}$$

Then, we slightly extend Lemma 2.2 in Alon, Benjamini and Stacey (2004) in proving that for every $c_1 \in]0, 1/2[$ and $c_2 \in]1/2, 1[$, there is a constant $q_3(c_1, c_2)$, depending only on c_1, c_2, b and d, such that, for every $p > q_3(c_1, c_2)$,

$$\begin{aligned}&\mathbb{P}\big(G_n(p) \text{ contains a component of size in } [c_1 n, c_2 n[\big)\\ &\qquad \le 4\left(1 + \frac{1}{c_1}\right)e^{-n}.\end{aligned} \tag{3}$$

The proof of this extension follows from an argument similar to Alon, Benjamini and Stacey (2004).

Once inequalities (2) and (3) are proved, Proposition 3.1 follows with the choice $q_2(c) = \max\{q_3(\min\{1/4, a\}, \max\{3/4, c\}), p_0(b)\}$.

PROOF OF (2). First we show that if $p > \frac{1}{1+b}$, then there is some $\delta > 0$, depending only on $p - \frac{1}{1+b}$, such that

$$\forall v \in V_n \qquad \mu_p\big(|C(v)| \ge n/2\big) \ge \delta. \tag{4}$$

To see this, we construct recursively the component $C(v)$ and its edge-boundary $W(v)$ as follows. First, we order the edges in E_n. Let $C_1 = \{v\}$ and $W_1 = \varnothing$. At each step of the algorithm, C_k denotes a subset of $C(v)$ and W_k a subset of $W(v)$. At step k, we explore the first (in the aforementioned order) edge $e_k = (y, z)$ from $E_n \setminus W_k$ that is adjacent to a vertex y from C_k and to a vertex z from $V_n \setminus C_k$, if there exists such an edge. Otherwise ($\partial_E C_{k-1} \subset W_{k-1}$), $C_{k+1} = C_k$ and $W_{k+1} = W_k$. If the edge $e_k = (y, z)$ is open [i.e., $x(e_k) = 1$], let $C_{k+1} = C_k \cup \{z\}$ and $W_{k+1} = W_k$. If the edge (y, z) is closed [i.e., $x(e_k) = 0$], let $C_{k+1} = C_k$ and $W_{k+1} = W_k \cup \{e_k\}$. We have $C(v) = \bigcup_{k=1}^{\infty} C_k$. If $|C(v)| < n$, then there is a smallest N such that $W_N = \partial_E C_N$. Notice that $N = |C_N| + |W_N|$ and $C_N = C(v)$. Thus, by the expansion assumption, if $|C_N| \le \frac{n}{2}$,

$$N - |W_N| = |C_N| \le \frac{|W_N|}{b},$$

so $|W_N| \ge Nb/(1 + b)$.

It is easy to see that, under $\mu_{n,p}$, $(x(e_1), \dots, x(e_N))$ can be completed so as to form an infinite i.i.d. sequence of Bernoulli random variables with parameter p. Thus, to construct $C(v)$, we flipped $N - 1$ independent $(p, 1 - p)$-coins and at least $Nb/(1 + b)$ among them turned out zero. But if $p > \frac{1}{1+b}$, then, with positive probability δ, depending only on $p - \frac{1}{1+b}$, a random infinite sequence of i.i.d. Bernoulli variables of parameter p does not have an n such that at least $(n + 1)b/(1 + b)$ among the N first coordinates equal 0. The last fact is a consequence of the law of large numbers. This proves inequality (4).

Now, fix $q \in](1 + b)^{-1}, 1[$, let R be a positive real number to be chosen later and define S_n to be the number of vertices which belong to a component of size at least $R/2$

$$S_n = \sum_{v \in V_n} \mathbb{1}_{|C(v)|>R/2}.$$

Denoting $X_v = \mathbb{1}_{|C(v)|>R/2}$, notice that X_v and $X_{v'}$ are independent as soon as $d(v, v') > R$ where $d(v, v')$ is the distance of vertices v and v' according to the shortest path metric in G_n. Thus, using the fact that the maximal degree in G_n is at most d, the maximal degree in the dependency graph of $(X_v)_{v \in V_n}$ is less than d^R. Recall that the dependency graph of the random variables $(X_v)_{v \in V_n}$ is given by the vertex set V_n and the edge set satisfying that if for two disjoint sets of vertices A and B there is no edge between A and B, then the families $(X_v)_{v \in A}$ and $(X_v)_{v \in B}$ are independent. Thus, by Theorem 2.1 in Janson (2004) for any $t > 0$,

$$\mu_p\big(S_n < \mathbb{E}_p(S_n) - t\big) \leq e^{-2t^2/(nd^R)}.$$

Notice that a similar result would be obtained from the method of bounded differences. From (4), we see that if $p \geq q$, $\mathbb{E}_p(S_n) \geq \delta n$. Choosing $t = \mathbb{E}_p(S_n)/2$ in the above inequality gives, for any $p \geq q$,

$$\mu_p(S_n < \delta n/2) \leq e^{-\delta^2 n/(2d^R)}.$$

This means that with probability at least $1 - e^{-\delta^2 n/(2d^R)}$, there are at least $\delta n/2$ vertices which belong to components of size at least $\frac{R}{2}$. Then, fix $p_0 \in]q, 1[$. The proof of Proposition 3.1 in Alon, Benjamini and Stacey (2004) shows that if R is chosen large enough, there is some positive constant C depending on b, d, δ, p_0 and q such that with probability at least $1 - e^{-\delta^2 n/(2d^R)} - e^{-Cn}$, for n large enough, there is a component of size at least $\delta n/6$ in $G_n(p_0)$. We recall their argument: fix a set of at most $r = \delta n/R$ components of size at least $R/2$ which contain together at least $\delta n/2$ vertices. If $\varepsilon = 1 - \frac{1-p_0}{1-q}$, $G(p_0)$ has the same law as $G(q) \cup G(\varepsilon)$, where $G(q)$ and $G(\varepsilon)$ are independent. Then, we claim that there is some C depending on b, d and ε such that with probability at least $1 - e^{-Can}$, in the random graph $G(\varepsilon)$, there is no way of splitting these components into two parts A and B, each containing at least $\delta n/6$ vertices, with no path of $G(\varepsilon)$ connecting the two parts. This will imply that, with the required probability, $G(q) \cup G(\varepsilon)$ contains a connected component consisting of at least $\delta n/6$ vertices. Now, we show the claim. Let us fix two parts A and B of the components above, each containing at least $\delta n/6$ vertices. Thanks to Menger's theorem, there are at least $b\delta n/6$ edge-disjoint paths between A and B. Since the total number of edges is less than $dn/2$, at least half of these paths are of length not larger than $6d/(b\delta)$. Thus, the probability that there is no path between A and B in $G(\varepsilon)$ is at most

$$\big(1 - \varepsilon^{6d/(b\delta)}\big)^{b\delta n/12} \leq e^{-b\delta n \varepsilon^{6d/(b\delta)}/12}.$$

Now, there are at most $2^r = 2^{\delta n/R}$ ways to choose A and B. Thus, the probability that there is a way to split the components into two parts A and B, each containing at least $\delta n/6$ vertices, with no path of $G(\varepsilon)$ connecting the two parts is at most

$$2^{\delta n/R} e^{-b\delta n \varepsilon^{6d/(b\delta)}/12} \leq e^{-b\delta n \varepsilon^{6d/(b\delta)}/24},$$

as soon as R is larger than $24\log 2/(b\varepsilon^{6d/(b\delta)})$. This finishes the proof of the claim, and thus the one of (2).

PROOF OF (3). It is well known [see Flajolet and Sedgewick (2009), Example I.14, page 68] that an infinite d-regular rooted tree contains precisely $\frac{1}{(d-1)r+1}\binom{r}{dr}$ rooted subtrees of size r. This number is at most $(de)^r$. To a graph G of maximum degree less than d and a vertex v of that graph, one may associate a subtree of the infinite d-regular tree rooted at v by considering the self-avoiding paths issued from v in G. Through this correspondence, any connected component of size r of G containing v is mapped to a different subtree of size r (through the choice of a spanning tree of the component in G). Thus, the total number of connected subsets of size r in V_n is at most $n(de)^r/r$. Now, using the expanding property, for any subset U of size r, the probability that all edges in $E(U, U^c)$ are absent is at most $(1-p)^{br}$ if $r \leq n/2$ and at most $(1-p)^{b(n-r)}$ if $r > n/2$. Thus, the probability that there is a connected component of size in $[c_1 n, c_2 n[$ is at most

$$\begin{aligned}
&\sum_{r=\lceil c_1 n\rceil}^{\lfloor n/2\rfloor} \frac{n(de)^r}{r}(1-p)^{br} + \sum_{r=\lfloor n/2\rfloor+1}^{\lfloor c_2 n\rfloor} \frac{n(de)^r}{r}(1-p)^{b(n-r)} \\
&\quad \leq \frac{1}{c_1}\frac{(de(1-p)^b)^{c_1 n}}{1-(de(1-p)^b)} + 2(1-p)^{nb}\frac{(de(1-p)^{-b})^{c_2 n+1}}{de(1-p)^{-b}-1} \\
&\quad \leq \frac{4}{c}e^{-n} + 4e^{-n},
\end{aligned}$$

provided that

$$de(1-p)^{-b} \geq 2, \qquad (de)^{c_2}(1-p)^{b(1-c_2)} \leq \frac{1}{e} \quad \text{and} \quad \bigl(de(1-p)^b\bigr)^{c_1} \leq \frac{1}{e}.$$

These conditions are satisfied if p is larger than some $q_3(c_1, c_2) < 1$. This defines the value of $q_3(c_1, c_2)$ for which (3) is valid for every $p > q_3(c_1, c_2)$. □

4. Threshold phenomenon for the appearance of a giant component. In this section, we prove our main result, Theorem 1.3. The main step is stated below in Theorem 4.1, showing that the threshold for having a component of size at least cn has width of order at most $O((\log n)^{-1/3})$. Theorem 4.1, together with Proposition 3.1 imply our main result, Theorem 1.3. [Recall that $p_n^*(c) = p_{n,1/2}(c)$.]

THEOREM 4.1. *Let $\alpha < 1/2$ and $c \in]0, 1[$. There is a constant C_3, depending only on c, α, b, and d, such that, for any n,*

$$p_{n,(1-\alpha)}(c) - p_{n,\alpha}(c) \le \frac{C_3}{(\log n)^{1/3}}.$$

Here is the idea of the proof. Let us call informally, the "super-critical phase" the set of values of p such that a giant component of size cn has appeared with probability greater than some $\alpha > 0$. The main idea of the proof is to show that for most values of p in the super-critical phase, the standard deviation of $L_n^{(1)}$, the size of the largest component, is small with respect to its mean (which is of the order n). This is shown essentially in Lemma 4.4 below. In this lemma, we crucially use an estimate of Alon, Benjamini and Stacey (2004) for the probability that the second largest component is "large" (greater than some n^ω with $\omega \in]0, 1[$). Next, it follows immediately from the expanding property that the mean of $L_n^{(1)}$ has a derivative at least of order n inside the super-critical phase. This is proved in Lemma 4.6. These two facts imply that the threshold is sharp: when p goes from $p_{n,1-\alpha}$ to $p_{n,\alpha}$, the size of the largest component increases by a positive fraction of the number of vertices, since the fluctuations of this size around its mean is small with respect to the number of vertices.

Now we turn to the proof which relies on a series of lemmas. First we need some technical definitions. For any function $f : \{0, 1\}^{E_n} \to \mathbb{R}$ and any $e \in E_n$, define the operator $\Delta_{e,p}$ as

$$\Delta_{e,p} f(x) = f(x) - \int f(x)\, dx(e),$$

where the integration with respect to $x(e)$ is understood with respect to the Bernoulli measure with parameter p. When there is no ambiguity, we write Δ_e instead of $\Delta_{e,p}$. Finally, the following notation will be useful: when X and $X' \in \{0, 1\}^{E_n}$ are independent and distributed according to $\mu_{n,p}$, and e belongs to E_n, we denote by $X^{(e)}$ the random configuration obtained from X by replacing X_e by X'_e.

LEMMA 4.2. *There exist $\delta(b, d) < 1$ and $K(b, d) < \infty$ such that, for every n,*

$$\sup_p \sup_{e \in E_n} \| \Delta_{e,p} L_n^{(1)} \|_{1,p} \le K(b, d) n^{\delta(b,d)}.$$

PROOF. To lighten notation, we write $f(x) = L_n^{(1)}(x)$ for the size of the largest component. A look at the proof of Theorem 2.8 in Alon, Benjamini and Stacey (2004) reveals that there are three positive real numbers $K_1(b, d) < \infty$, $\omega(b, d) < 1$, and $g(b, d) > 0$, depending only on b and d, such that

$$\sup_p \mathbb{P}\begin{pmatrix} G_n(p) \text{ contains more than one component} \\ \text{of size at least } n^{\omega(b,d)} \end{pmatrix} \le K_1(b, d) n^{-g(b,d)}.$$

The estimate for p close to 0 is not made explicit in Alon, Benjamini and Stacey (2004) but follows from (1). Since f is monotone, one may write for any fixed p,

$$\|\Delta_{e,p} f\|_{1,p} = 2\mathbb{E}\big[\big(f(X) - f\big(X^{(e)}\big)\big)_-\big]$$

[where $y_- = \max(0, -y)$ denotes the negative part of a real number y]. Let A_n be the event that the size of the second largest connected component is at most $n^{\omega(b,d)}$. Notice that $(f(X) - f(X^{(e)}))_-$ can only be positive if $x(e) = 0$, and the edge e is adjacent to the largest component in G_n. Then the difference between $f(X^{(e)})$ and $f(X)$ is the size of the component that gets attached to the largest component by adding the edge e to G_n. Thus $(f(X) - f(X^{(e)}))_-$ is always bounded by the size of the second largest connected component in G_n. It is also smaller than n. Thus, we have

$$\begin{aligned}\|\Delta_{e,p} L_n^{(1)}\|_{1,p} &\le 2\mathbb{E}\big[\big(L_n^{(1)}(X) - L_n^{(1)}\big(X^{(e)}\big)\big)_-\big]\\ &\le 2n^{\omega(b,d)} + 2n\mathbb{P}(A_n^c)\\ &\le \big(2 + K_1(b,d)\big)n^{\delta(b,d)},\end{aligned}$$

where $\delta(b,d) = \max\{\omega(b,d), 1 - g(b,d)\} < 1$. □

Next we establish an upper bound for the variance of the second largest component. Our main tool is a result of Falik and Samorodnitsky (2007) that gives an improved estimate over the Efron–Stein inequality for functions defined on the binary hypercube. Recall that the Efron–Stein inequality implies that if $f : \{0,1\}^{E_n} \to \mathbb{R}$, then

$$\mathsf{Var}(f) \le \sum_{e \in E_n} \|\Delta_e f\|_{2,p}^2$$

[see Efron and Stein (1981)]. The next inequality appears in this form in Benjamini and Rossignol (2008):

LEMMA 4.3 (Falik and Samorodnitsky). *Let f belong to $L^1(\{0,1\}^{E_n})$. Suppose that $\mathcal{E}_1(f)$ and $\mathcal{E}_2(f)$ are two real numbers such that*

$$\mathcal{E}_2(f) \ge \sum_{e \in E_n} \|\Delta_e f\|_2^2,$$

$$\mathcal{E}_1(f) \ge \sum_{e \in E_n} \|\Delta_e f\|_1^2$$

and

$$\frac{\mathcal{E}_2(f)}{\mathcal{E}_1(f)} \ge e.$$

Then

$$\mathsf{Var}(f) \le 2\frac{\mathcal{E}_2(f)}{\log(\mathcal{E}_2(f)/(\mathcal{E}_1(f)\log(\mathcal{E}_2(f)/\mathcal{E}_1(f))))}.$$

This inequality may be used to derive our next key lemma, implying that for most values of p in the super-critical phase, the standard deviation of $L_n^{(1)}$ is small with respect to its mean.

LEMMA 4.4. *There is a constant $C(b,d) < \infty$ such that, for any p and n,*

$$\mathsf{Var}_p(L_n^{(1)}) \le C(b,d)\frac{n}{\log n}\frac{d\mathbb{E}_p(L_n^{(1)})}{dp}.$$

PROOF. Fix $\beta \in]0, 1-\delta(b,d)[$, where $\delta(b,d)$ is given in Lemma 4.2. Define

$$\mathcal{E}_2(L_n^{(1)}) = \sum_{e\in E_n}\|\Delta_e L_n^{(1)}\|_2^2$$

and

$$\mathcal{E}_1(L_n^{(1)}) = \sum_{e\in E_n}\|\Delta_e L_n^{(1)}\|_1^2.$$

We distinguish two cases depending on the relationship between $\mathcal{E}_1(L_n^{(1)})$ and $\mathcal{E}_2(L_n^{(1)})$:

- If $\mathcal{E}_2(L_n^{(1)}) \le n^\beta \mathcal{E}_1(L_n^{(1)})$, then using Lemma 4.2,

$$\mathcal{E}_2(L_n^{(1)}) \le K(b,d)n^{\beta+\delta(b,d)}\sum_{e\in E_n}\|\Delta_e L_n^{(1)}\|_1.$$

  Since $L_n^{(1)}$ is a monotone increasing function on the binary hypercube, a straightforward generalization of Russo's lemma [see Rossignol (2006)] implies that

$$\frac{d\mathbb{E}_p(L_n^{(1)})}{dp} = \frac{1}{2p(1-p)}\sum_{e\in E_n}\|\Delta_e L_n^{(1)}\|_1. \tag{5}$$

  On the other hand, by the Efron–Stein inequality,

$$\mathsf{Var}_p(L_n^{(1)}) \le \mathcal{E}_2(L_n^{(1)}).$$

  Thus, ignoring the term $2p(1-p) < 1$,

$$\mathsf{Var}_p(L_n^{(1)}) \le K(b,d)n^{\beta+\delta(b,d)} \times \frac{d\mathbb{E}_p(L_n^{(1)})}{dp}.$$

- If $\mathcal{E}_2(L_n^{(1)}) > n^\beta \mathcal{E}_1(L_n^{(1)})$, then Lemma 4.3 implies that there is a constant C_1, depending only on β (i.e., on b and d), such that

$$\mathsf{Var}_p(L_n^{(1)}) \le C_1 \frac{\mathcal{E}_2(L_n^{(1)})}{\log n}.$$

But since $L_n^{(1)}$ is positive and always smaller than n, we have $\mathcal{E}_2(L_n^{(1)}) \le n \sum_{e \in E_n} \|\Delta_e L_n^{(1)}\|_1$. Thus, Russo's lemma implies that

$$\mathsf{Var}_p(L_n^{(1)}) \le C_1 \frac{n}{\log n} \frac{d\mathbb{E}_p(L_n^{(1)})}{dp}.$$

In both cases, the result follows. □

The following easy lemma states that, whatever $\gamma < 1$ and $\varepsilon \in]0, 1[$ are, there is always some $c < 1$ such that the probability of having a component of size cn is less than ε if $p \le \gamma$.

LEMMA 4.5. *Let $\gamma < 1$ and $\varepsilon \in]0, 1[$. Then there is some $c < 1$ such that, for n large enough,*

$$\mu_{n,\gamma}(L_n^{(1)} \ge cn) \le \varepsilon.$$

PROOF. The size of the largest connected component is less than $n - N$, where N is the number of isolated vertices (except when all vertices are isolated, in which case $L_n^{(1)}$ is 1). Let X_v denote the indicator function of the event "v is isolated." If $d(v, v') \ge 2$, X_v and X'_v are independent. Thus, the maximal degree in the dependency graph of $(X_v)_{v \in V_n}$ is less than d, and Theorem 2.1 in Janson (2004) shows that for any $t > 0$ and $p \in [0, 1]$,

$$\mu_{n,p}(N < \mathbb{E}_p(N) - t) \le e^{-2t^2/(nd)}.$$

On the other hand, by the bounded-degree assumption, $\mathbb{E}_\gamma(N) \ge (1-\gamma)^d n$, and therefore, for any $c > 1 - (1-\gamma)^d$,

$$\mu_{n,\gamma}(L_n^{(1)} \ge cn) \le \mu_{n,\gamma}(N < (1-c)n) \le e^{-2n((1-\gamma)^d - (1-c))^2/d}.$$

Choose c such that

$$1 - (1-\gamma)^d + \sqrt{\frac{d \log(1/\varepsilon)}{2n}} \le c < 1$$

(which is possible for n large enough), and get the desired inequality. □

The last piece we need for the proof is the fact that the mean grows at least linearly in the super-critical phase. This can be proved using the expansion property as follows:

LEMMA 4.6. *Let $\alpha \in]0, 1/2[$ and $c \in]0, 1[$. There is a positive constant C', depending only on α, c, b and d, such that for n large enough, and for every $p \in [p_{n,\alpha}(c), p_{n,1-\alpha}(c)]$,*

$$\frac{d\mathbb{E}_p(L_n^{(1)})}{dp} \geq C'n.$$

PROOF. Let us fix $0 < c < c_2 < 1$. Writing $f(x) = L_n^{(1)}(x)$ and using Russo's lemma (5),

$$\begin{aligned}\frac{d\mathbb{E}_p(L_n^{(1)})}{dp} &= \frac{1}{p(1-p)} \sum_{e \in E_n} \mathbb{E}_p\big[\big(f(X) - f\big(X^{(e)}\big)\big)_-\big] \\ &\geq \frac{1}{(1-p)} \mathbb{E}_p\big(\big|\partial_E\big(\mathcal{C}^{(1)}\big)\big|\big).\end{aligned}$$

By the expansion property,

$$\big|\partial_E\big(\mathcal{C}^{(1)}\big)\big| \geq b\big(L_n^{(1)} \mathbb{1}_{L_n^{(1)} \leq n/2} + \big(n - L_n^{(1)}\big) \mathbb{1}_{L_n^{(1)} > n/2}\big),$$

and therefore, for any $p \in [p_{n,\alpha}(c), p_{n,1-\alpha}(c)]$,

$$\begin{aligned}\frac{d\mathbb{E}_p(L_n^{(1)})}{dp} &\geq \frac{b}{(1-p)} \mathbb{E}\big[L_n^{(1)} \mathbb{1}_{L_n^{(1)} \leq n/2} + \big(n - L_n^{(1)}\big) \mathbb{1}_{L_n^{(1)} > n/2}\big] \\ &\geq bn \min\{c, (1-c_2)\} \mu_p\big(L_n^{(1)} \in [cn, c_2 n[\big).\end{aligned}$$

Now, thanks to Proposition 3.1, we know that there is some $q_2(c) < 1$ such that for n large enough, $p_{n,1-\alpha}(c) \leq q_2(c)$. Thus, applying Lemma 4.5 with $\gamma = q_2(c)$ and $\varepsilon = \alpha/2$, there is some $c_2(c) \in]c, 1[$ such that, for n large enough, $p_{n,1-\alpha}(c) \leq q_2(c) \leq p_{n,\alpha/2}(c_2)$. Therefore $\mu_p(L_n^{(1)} \geq cn) \geq \alpha$ and $\mu_p(L_n^{(1)} \geq c_2 n) \leq \alpha/2$ for any $p \in [p_{n,\alpha}(c), p_{n,1-\alpha}(c)]$. This leads to

$$\frac{d\mathbb{E}_p(L_n^{(1)})}{dp} \geq bn \min\{c, (1-c_2)\}\alpha/2$$

for any $p \in [p_{n,\alpha}(c), p_{n,1-\alpha}(c)]$. □

Now we are ready to wrap up our argument.

PROOF OF THEOREM 4.1. Let $c < 1$ and $\alpha < 1/2$ be fixed positive numbers. We show that there exists a constant $K = K(\alpha, c, b, d)$, such that if $\varepsilon_n = K \log^{-1/3} n$, then

$$p_{1/2}(c) - p_\alpha(c) \leq \varepsilon_n.$$

The proof that $p_{1-\alpha}(c) - p_{1/2}(c) \leq \varepsilon_n$ is completely similar.

Lemma 4.4 and the trivial bound $L_n^{(1)} \le n$, implies that no matter how ϵ_n is chosen,

$$\int_{p_{1/2}(c)-\varepsilon_n}^{p_{1/2}(c)-3\varepsilon_n/4} \mathsf{Var}_p(L_n^{(1)})\,dp \le C\frac{n^2}{\log n},$$

where $C = C(b,d)$. Thus there is some $q_1 \in [p_{1/2}(c) - \varepsilon_n, p_{1/2}(c) - 3\varepsilon_n/4]$ such that

$$\mathsf{Var}_{q_1}(L_n^{(1)}) \le \frac{4Cn^2}{\varepsilon_n \log n}.$$

Similarly, one finds q_2 such that

$$q_2 \in [p_{1/2}(c) - \varepsilon_n/2, p_{1/2}(c) - \varepsilon_n/4] \quad \text{and} \quad \mathsf{Var}_{q_2}(L_n^{(1)}) \le \frac{4Cn^2}{\varepsilon_n \log n}.$$

Observe that it suffices to prove that $q_1 \le p_\alpha(c)$. Note that $q_1 + \frac{\varepsilon_n}{4} \le q_2 \le p_{1/2}(c)$. Thanks to Lemma 4.6, there is a constant C', depending only on α, c, b and d such that for n large enough,

$$\mathbb{E}_{q_2}(L_n^{(1)}) - \mathbb{E}_{q_1}(L_n^{(1)}) \ge C'n\frac{\varepsilon_n}{4}.$$

On the other hand, denote by M_p the median of $L_n^{(1)}$ under μ_p (we assume that it is of the form $k + 1/2$, with $k \in \mathbb{N}$, which ensures its uniqueness). M_p is an increasing function of p and therefore

$$cn \ge M_{p_{1/2}(c)} - \tfrac{1}{2} \ge M_{q_2} - \tfrac{1}{2}.$$

By Lévy's inequality, the difference between the mean and median of any random variable is bounded by its standard deviation and therefore

$$\left|\mathbb{E}_{q_2}(L_n^{(1)}) - M_{q_2}\right| \le \sqrt{\mathsf{Var}_{q_2}(L_n^{(1)})} \le n\sqrt{\frac{4C}{\varepsilon_n \log n}}.$$

Summarizing, we can write

$$\begin{aligned}
&\mu_{q_1}(L_n^{(1)} \ge cn) \\
&\quad = \mu_{q_1}(L_n^{(1)} - \mathbb{E}_{q_1}(L_n^{(1)}) \ge cn - \mathbb{E}_{q_1}(L_n^{(1)})) \\
&\quad \le \mu_{q_1}\left(L_n^{(1)} - \mathbb{E}_{q_1}(L_n^{(1)}) \ge M_{q_2} - \frac{1}{2} - \mathbb{E}_{q_1}(L_n^{(1)})\right) \\
&\quad = \mu_{q_1}\left(L_n^{(1)} - \mathbb{E}_{q_1}(L_n^{(1)}) \ge M_{q_2} - \mathbb{E}_{q_2}(L_n^{(1)}) + \mathbb{E}_{q_2}(L_n^{(1)}) - \mathbb{E}_{q_1}(L_n^{(1)}) - \frac{1}{2}\right) \\
&\quad \le \mu_{q_1}\left(L_n^{(1)} - \mathbb{E}_{q_1}(L_n^{(1)}) \ge C'n\frac{\varepsilon_n}{4} - \frac{1}{2} - n\sqrt{\frac{4C}{\varepsilon_n \log n}}\right).
\end{aligned}$$

Now we choose $K = (256C/(C'^2\alpha))^{1/3}$, so

$$\varepsilon_n = \left(\frac{256C}{C'^2\alpha \log n}\right)^{1/3}.$$

Clearly, for n sufficiently large,

$$C'n\frac{\varepsilon_n}{4} - \frac{1}{2} - n\sqrt{\frac{4C}{\varepsilon_n \log n}} \geq C'n\frac{\varepsilon_n}{8}.$$

Thus, Chebyshev's inequality implies

$$\begin{aligned}\mu_{q_1}\big(L_n^{(1)} \geq cn\big) &\leq \frac{\mathsf{Var}_{q_1}(L_n^{(1)})}{(C'n\varepsilon_n/8)^2}\\ &\leq \frac{4Cn^2}{\varepsilon_n \log n}\frac{64}{C'^2n^2\varepsilon_n^2}\\ &= \alpha.\end{aligned}$$

This implies that $q_1 \leq p_\alpha(c)$, as desired. □

The bound on the size of the threshold width in Theorem 4.1 is quite likely not to be tight. Indeed, one would rather be inclined to compare it to what happens in the Erdős–Rényi random graph $\mathcal{G}(n, (1+\varepsilon)/n)$. In this case, the mean grows also linearly, and the fluctuations of the giant component are approximately Gaussian, with a variance of order $\Theta(n)$, which implies that the threshold width is of order $\Theta(1/\sqrt{n})$. A similar tight threshold seems to hold in random d-regular graphs as well [see Theorem 3 in Pittel (2008)]. Thus, it is natural to conjecture that the threshold is much smaller in our setting as well. Note that if the underlying graph G_n is transitive, then Friedgut and Kalai (1996) implies that the threshold width is at most $O(1/\log n)$, so our quantitative bound $O(1/\log^{1/3} n)$ seems to be very weak.

Finally, let us emphasize an open problem about "exponential decay" of the probabilities of abnormally small or abnormally large size of the giant cluster. Proposition 3.1 implies that for any c in $]0, 1[$, for p close enough to 1, the probability of not having a component of size at least cn decays as quickly as $e^{-\alpha n}$ for some $\alpha > 0$. An open problem is to find out whether this is the case as soon as $p > p_n^*(c) + \varepsilon$ for some fixed $\varepsilon > 0$. A similar question can be posed on the left of the threshold: at which speed does the probability of having a component of size at least cn decay to zero when $p < p_n^*(c) - \varepsilon$?

Acknowledgment. We thank Gady Kozma for useful discussions.

REFERENCES

ALON, N., BENJAMINI, I. and STACEY, A. (2004). Percolation on finite graphs and isoperimetric inequalities. *Ann. Probab.* **32** 1727–1745. MR2073175

BENJAMINI, I., NACHMIAS, A. and PERES, Y. (2009). Is the critical percolation probability local? *Probab. Theory Related Fields* **149** 261–269.

BENJAMINI, I. and ROSSIGNOL, R. (2008). Submean variance bound for effective resistance of random electric networks. *Comm. Math. Phys.* **280** 445–462. MR2395478

BENJAMINI, I. and SCHRAMM, O. (1996). Percolation beyond $\mathbb{Z}^d$, many questions and a few answers. *Electron. Comm. Probab.* **1** 71–82. MR1423907

EFRON, B. and STEIN, C. (1981). The jackknife estimate of variance. *Ann. Statist.* **9** 586–596.

FALIK, D. and SAMORODNITSKY, A. (2007). Edge-isoperimetric inequalities and influences. *Combin. Probab. Comput.* **16** 693–712. Available from http://arxiv.org/pdf/math.CO/0512636. MR2346808

FLAJOLET, P. and SEDGEWICK, R. (2009). *Analytic Combinatorics*. Cambridge Univ. Press, Cambridge.

FRIEDGUT, E. and KALAI, G. (1996). Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.* **124** 2993–3002. MR1371123

JANSON, S. (2004). Large deviations for sums of partly dependent random variables. *Random Structures Algorithms* **24** 234–248. MR2068873

LYONS, R. and PERES, Y. (2011). *Probability on Trees and Networks*. Cambridge Univ. Press, Cambridge. To appear. Available at http://mypage.iu.edu/~rdlyons/prbtree/prbtree.html.

NACHMIAS, A. and PERES, Y. (2010). Critical percolation on random regular graphs. *Random Structures Algorithms* **36** 111–148. MR2583058

PITTEL, B. (2008). Edge percolation on a random regular graph of low degree. *Ann. Probab.* **36** 1359–1389. MR2435852

ROSSIGNOL, R. (2006). Threshold for monotone symmetric properties through a logarithmic Sobolev inequality. *Ann. Probab.* **34** 1707–1725. MR2271478

I. BENJAMINI
THE WEIZMANN INSTITUTE OF SCIENCE
REHOVOT POB 76100
ISRAEL

S. BOUCHERON
LABORATOIRE PROBABILITÉS
ET MODÈLES ALÉATOIRES
UNIVERSITÉ PARIS-DIDEROT
175 RUE DU CHEVALERET
75013-F PARIS
FRANCE

G. LUGOSI
ICREA AND DEPARTMENT OF ECONOMICS
POMPEU FABRA UNIVERSITY
25-27 RAMON TRIAS FARGAS
BARCELONA
SPAIN

R. ROSSIGNOL
LABORATOIRE DE MATHÉMATIQUES
UMR 8628
UNIVERSITÉ PARIS-SUD
ORSAY, F91405
FRANCE
E-MAIL: raphael.rossignol@math.u-psud.fr
AND
CNRS, UMR 8628
ORSAY, F91405
FRANCE