The Evolution of Random Subgraphs of the Cube

B. Bollobás[1, 2], Y. Kohayakawa[1, 2] and T. Łuczak[1, 3]

[1] Department of Pure Mathematics and Mathematical Statistics,
University of Cambridge, 16 Mill Lane, Cambridge CB2 1SB, England

[2] Department of Mathematics,
Louisiana State University, Baton Rouge, LA 70803, USA

[3] Department of Discrete Mathematics,
Adam Mickiewicz University, ul. Matejki 48/49, Poznań, Poland

Abstract. Let $(Q_t)_0^M$ be a random Q^n-process, that is let Q_0 be the empty spanning subgraph of the cube Q^n and, for $1 \le t \le M = nN/2 = n2^{n-1}$, let the graph Q_t be obtained from Q_{t-1} by the random addition of an edge of Q^n not present in Q_{t-1}. When t is about $N/2$, a typical Q_t undergoes a certain 'phase transition': the component structure changes in a sudden and surprising way. Let $t = (1+\epsilon)N/2$ where ϵ is independent of n. Then all the components of a typical Q_t have $o(N)$ vertices if $\epsilon < 0$, while if $\epsilon > 0$ then, as proved by Ajtai, Komlós and Szemerédi, a typical Q_t has a 'giant' component with at least $\alpha(\epsilon)N$ vertices, where $\alpha(\epsilon) > 0$. In this note we give essentially best possible results concerning the emergence of this giant component close to the time of phase transition. Our results imply that if $\eta > 0$ is fixed and $t \le (1 - n^{-\eta})N/2$ then all components of a typical Q_t have at most $n^{\beta(\eta)}$ vertices, where $\beta(\eta) > 0$. More importantly, if $60(\log n)^3/n \le \epsilon = \epsilon(n) = o(1)$ then the largest component of a typical Q_t has about $2\epsilon N$ vertices, while the second largest component has order $O(n\epsilon^{-2})$. Loosely put, the evolution of a typical Q^n-process is such that shortly after time $N/2$ the appearance of each new edge results in the giant component acquiring 4 new vertices.

1. Introduction

Let H be a graph with m edges. A *random H-process* $\widetilde{H} = (H_t)_0^m$ is a Markov chain whose states are spanning subgraphs of H. The process starts with the empty subgraph and for $1 \le t \le m$ the subgraph H_t is obtained from H_{t-1} by the random addition of an edge of H which is not present in H_{t-1}, all of such edges being equiprobable. Thus H_t has exactly t edges and $H_m = H$. The K^n-processes are the usual random graph processes $\widetilde{G} = (G_t)$.

A random H-process is intimately related to two spaces of random subgraphs of H, namely $\mathcal{G}(H,p)$ and $\mathcal{G}(H,t)$. These spaces are defined for $0 \le p \le 1$ and $0 \le t \le m$. A random element of the space $\mathcal{G}(H,p)$ is a spanning subgraph of H obtained by selecting its edges from the edges of H with probability p, all such selections being independent from one another. A random element of the space $\mathcal{G}(H,t)$ is simply a spanning subgraph of H with t edges, all such subgraphs being equiprobable.

Let us recall some of the major discoveries of Erdős and Rényi ([11] and [12]) concerning random graph processes, that is K^n-processes. As is customary, given $j \geq 1$ let us denote by $L_j(G)$ the order of the jth largest component of G; if G has fewer than j components we set $L_j(G) = 0$. Let $c > 0$ be a constant. Erdős and Rényi proved that a.e. random graph process $\widetilde{G} = (G_t)$ is such that if $c < 1$ then

$$L_1(G_{\lfloor cn/2 \rfloor}) = O(\log n).$$

On the other hand, if $c > 1$ then a.e. such process satisfies

$$L_1(G_{\lfloor cn/2 \rfloor}) = (1 - t(c) + o(1))n,$$

where $t(c)$ is a certain decreasing function of c. Moreover, a.e. $\widetilde{G}$ is such that $G_{\lfloor cn/2 \rfloor}$ has a 'giant' component: one whose order is much greater than the order of any other component.

The following considerable refinements of the above results were proved in [3]. (We shall be very sketchy and somewhat imprecise.) If t is not much greater than $n/2$, then

$$L_1(G_t) = (1 + o(1))(4t - 2n)$$

for a.e. $\widetilde{G}$, *i.e.* the giant component grows about four times as fast as the number of edges. As the process continues, the larger components are swallowed up by the giant component at such a rate that the order of the second largest component decreases. By time $cn/2$, where $c > 1$, the second largest component has order $O(\log n)$ for a.e. $\widetilde{G}$. The reader is referred to Chapters V and VI of [4] for a very detailed account of these and other related results. Also, recent further improvements are given in [18].

The spaces $\mathcal{G}(H, p)$ and $\mathcal{G}(H, t)$ have not yet been studied systematically for arbitrary finite non-complete graphs H. There is however a reasonable number of results (and very natural and challenging conjectures) concerning these spaces for the n-dimensinal cube Q^n. In this note we shall give some results concerning Q^n-processes that have the same flavour as the results about $L_1(G_t)$ mentioned above. Let us recall that the *n-dimensional cube* Q^n, or simply the *n-cube*, is the graph whose vertices are the subsets of $[n] = \{1, \ldots, n\}$, two of them being adjacent in Q^n if and only if their symmetric difference is a singleton. Thus Q^n has $N = 2^n$ vertices and $M = nN/2$ edges (throughout this note N and M will stand for these quantities). Note that Q^n is rather sparse.

Burtin was the first to study the space $\mathcal{G}(Q^n, p)$. In [9], he showed that $p = 1/2$ is the critical value for connectedness: for fixed values of p, a.e. $Q_p \in \mathcal{G}(Q^n, p)$ is disconnected if $p < 1/2$ while a.e. $Q_p \in \mathcal{G}(Q^n, p)$ is connected if $p > 1/2$. Erdős and Spencer [13] (see also Toman [20]) analysed the case $p = 1/2$ and proved that the probability that $Q_{1/2}$ is connected tends to 1/e, thus

$$\lim_{n\to\infty} \mathbb{P}\left(Q_p \in \mathcal{G}(Q^n, p) \text{ is connected}\right) = \begin{cases} 0 & \text{if } p < 1/2 \\ 1/\mathrm{e} & \text{if } p = 1/2 \\ 1 & \text{if } p > 1/2. \end{cases}$$

(For considerable refinements of these results, see [2], [5], and Dyer, Frieze and Foulds [10].) Now, Erdős and Spencer also showed that all components of $Q_p \in \mathcal{G}(Q^n, p)$ are a.s. of order $o(N)$ provided $p = c/n$ with $c < 1$ fixed. They conjectured that a 'jump' occurs at $p = 1/n$, *i.e.* when the average degree is one, just as in the case of ordinary random graphs: for any fixed $c > 1$, if $p = c/n$ then a.e. $Q_p \in \mathcal{G}(Q^n, p)$ contains a component of order at least ηN, where $\eta = \eta(c) > 0$. Ajtai, Komlós and Szemerédi [1] proved this conjecture and they also suggested a possible way of showing that the second largest component of such a Q_p is a.s. $O(n)$, as conjectured by Erdős.

The main results in this note considerably improve our knowledge of the behaviour of the orders of the largest and second largest components of a typical $Q_t \in \mathcal{G}(Q^n, t)$. Roughly speaking, we shall substantially improve Erdős and Spencer's remark on the order of the components before time $N/2$ and we shall show that some of the results in [3] carry over to cube processes.

More precisely, we shall show that for a.e. $Q_p \in \mathcal{G}(Q^n, p)$ the largest component of Q_p has order

$$-(1 + o(1))\frac{n \log 2}{\epsilon + \log(1 - \epsilon)} = (2 \log 2 + o(1))n\epsilon^{-2}, \tag{1}$$

where $p = (1-\epsilon)/n$, and $\epsilon = \epsilon(n) \geq (\log n)^2/(\log\log n)\sqrt{n}$ is bounded away from 1. In ordinary random graph processes $\widetilde{G} = (G_t)$, the critical value for t is $n/2$ in the sense that the behaviour of $L_1(G_t)$ starts to change suddenly at time $t = n/2$. Hence one might at first think that the critical time for Q^n-processes $\widetilde{Q} = (Q_t)_{t=0}^M$ should be $N/2$. Our results concerning $L_1(Q_p)$ will show that this is not the case; in fact, we shall see that (1) is an upper bound for $L_1(Q_p)$ for $p = (1-\epsilon)/(n-1)$, $\epsilon > 0$, and hence the critical time for Q^n-processes is at least $(1 + 1/(n-1))N/2$ (see Section 8).

We shall see that for a.e. Q^n-process a giant component emerges soon after time $N/2$. In fact, we shall see that a.e. $\widetilde{Q} = (Q_t)_0^M$ is such that for

$$t \geq \left\{1 + \frac{60(\log n)^3}{n}\right\}\frac{N}{2},$$

the graph Q_t contains a distinguished largest component whose order is much greater than the order of any other component (see Theorem 25).

We shall also estimate the order of the giant component of Q_t. We shall see that if $t = N/2 + s$ and $30N(\log n)^3/n \leq s = o(N)$ then, as in the ordinary processes, the giant component grows four times as fast as the number of edges. Indeed,

$$L_1(Q_t) = (4 + o(1))s$$

will turn out to hold a.s. in this case (see Theorem 28). Also, we shall see that in this range of s we almost surely have

$$L_2(Q_t) = O(\epsilon^{-2}n),$$

where $\epsilon = \epsilon(s) = 2s/N$. Therefore the order of the second largest component decreases as s increases.

We shall also consider times $t = cN/2$ for constants $c > 1$. We shall see that then the giant component is of order $(\eta + o(1))N$, where $\eta = \eta(c) > 0$ will be computed explicitly (see Theorem 29). It will in fact turn out that $\eta(c) = 1 - t(c)$, which shows another similarity between ordinary graph processes and cube processes. These similarities between the two processes are at least at a first glance rather surprising since Q^n is a very sparse graph.

Finally, we present more precise results concerning the order of the second largest component of random subgraphs of Q^n. However, for the sake of simplicity we shall only look at $Q_p \in \mathcal{G}(Q^n, p)$, since the computations involved in the proof of the corresponding results for $Q_t \in \mathcal{G}(Q^n, t)$ are fierce and not at all enlightening. The main result here is that, for fixed $k \geq 2$, the order of the kth largest component of Q_p is

$$(1 + o(1))\frac{n\log 2}{\epsilon - \log(1+\epsilon)} = (2\log 2 + o(1))n\epsilon^{-2},$$

where $p = (1+\epsilon)/n$ and $(\log n)^2/(\log\log n)\sqrt{n} \leq \epsilon \leq 1$ (cf. Theorem 32).

This note is organised as follows. We first give some preliminary lemmas in Section 2; some well-known simple facts concerning branching processes are also briefly discussed in this section, as they will be used in the proof of the estimate of the order of the giant component. In Section 3 we study the behaviour of $L_1(Q_p)$ for $p < 1/(n-1)$. In the subsequent sections we deal with $p > 1/(n-1)$. In Section 4 we prove that shortly after time $N/2$ a gap in the sequence of the orders of the components develops in a.e. cube process. In Section 5 we turn our attention to the problem of estimating the number of vertices that lie in large components. The main results are given in Section 6. In Section 7 we give the results concerning the second largest component of Q_p ($p > 1/(n-1)$). We close with some comments and open problems.

2. Preliminaries

We shall need to know reasonably precisely the number of certain subgraphs of the cube Q^n, and it will also be important for us to know the number of certain subtrees of the rooted n-regular infinite tree T_n. We shall start this section with some lemmas that give us good enough estimates for these quantities.

Let k, ℓ and m be given. For a vertex $v \in Q^n$, let $C_v(k, \ell, m)$ denote the family of connected subgraphs C of Q^n that contain v, have order $|C| = |V(C)| = k$, size $e(C) = |E(C)| = \ell$, and such that the number of edges $e(Q^n[V(C)])$ induced by $V(C)$ in Q^n is m. Let us write $c(k, \ell, m) = |C_v(k, \ell, m)|$.

Lemma 1. *For all n, k, ℓ and m satisfying $1 \leq k-1 \leq \ell \leq m \leq nk-m$, we have*

$$c(k,\ell,m) \leq \frac{(nk-m)^{nk-m}}{\ell^\ell (nk-m-\ell)^{nk-m-\ell}} \leq 3(nk)^{1/2}\binom{nk-m}{\ell}.$$

Proof. Let us set $p = \ell/(nk-m)$; note that $0 \leq p \leq 1$. Given $Q_p \in \mathcal{G}(Q^n, p)$, let us denote by $C_v = C_v(Q_p)$ the connected component of Q_p that contains v. Then the probability that C_v is a member of $C_v(k,\ell,m)$ equals

$$c(k,\ell,m)p^\ell(1-p)^{nk-m-\ell}.$$

Our estimate on $c(k,\ell,m)$ follows from the fact that this probability is less than 1. □

Let T_n be the infinite n-regular rooted tree. Let us denote by $t(k,n)$ the number of subtrees of T_n that contain the root of T_n and have order k.

Lemma 2. *For all n, $k \geq 2$, we have*

$$t(k,n) = \frac{n}{k-1}\binom{k(n-1)}{k-2}.$$

Proof. Let us for simplicity write t_k for $t(k,n)$. Let us delete the edges of T_n independently with probability $1-p$. The probability that, in the random graph thus obtained, the component T_n^p of the root v_0 of T_n is of order k is $t_k p^{k-1}(1-p)^{nk-2k+2}$, and hence we have that

$$\sum_{k\geq 1} t_k p^{k-1}(1-p)^{nk-2k+2} = 1, \tag{2}$$

provided p is such that T_n^p is finite with probability 1. Let us write $X_\ell = X_\ell(T_n^p)$ for the number of vertices of T_n at distance ℓ from v_0 that belong to T_n^p. Then if $p \leq 1/2n$ and $\ell \geq 1$ we have that

$$\mathbb{P}(|C_v| = \infty) \leq \mathbb{P}(X_\ell \geq 1) \leq \mathbb{E}(X_\ell) \leq 2^{-\ell},$$

and hence $\mathbb{P}(|C_v| = \infty) = 0$. Therefore (2) holds for all $0 \leq p \leq 1/2n$. Equivalently, we have that for all $0 \leq x \leq 1/2n$

$$f(x) = \sum_{k\geq 1} t_k y^k, \tag{3}$$

where $f(x) = x(1-x)^{-2}$ and $y = x(1-x)^{n-2}$. We can now use the Lagrange inversion formula to find $t_k = t(k,n)$ explicitly. Indeed, let $\varphi(x) = (1-x)^{-(n-2)}$, and note that both φ and f are analytic in a neighbourhood of 0. Since $x = y\varphi(x)$, there is a function $\xi(y)$, analytic in a neighbourhood of $y = 0$, such that $x = \xi(y)$. In fact, The Lagrange inversion formula tells us that if $g(y) = f(\xi(y))$ then

$$g(y) = f(0) + \sum_{k\geq 1} \frac{y^k}{k!}\left[\mathrm{D}^{k-1}\left((\mathrm{D}\, f(x))(\varphi(x))^k\right)\right]_{x=0}$$

where $\mathrm{D} = d/dx$. Simple computations show that

$$\begin{aligned}(\mathrm{D}\, f(x))(\varphi(x))^k &= (1+x)(1-x)^{-k(n-2)-3}\\ &= \sum_{j\geq 0}\left\{\binom{k(n-2)+2+j}{j} - \binom{k(n-2)+1+j}{j-1}\right\}x^j\end{aligned}$$

and hence

$$\begin{aligned}\mathrm{D}^{k-1}\left[(\mathrm{D}\, f(x))(\varphi(x))^k\right]\Big|_{x=0} &= (k-1)!\left\{\binom{k(n-1)+1}{k-1} - \binom{k(n-1)}{k-2}\right\}\\ &= (k-1)!\frac{kn}{k-1}\binom{k(n-1)}{k-2},\end{aligned}$$

where for the second equality we need that $k \geq 2$. Thus

$$g(y) = y + \sum_{k \geq 2} \frac{n}{k-1} \binom{k(n-1)}{k-2} y^k, \tag{4}$$

and the result follows by comparing (3) and (4). Indeed, the power-series on the right-hand side of (3) defines an analytic function $g'(y)$ of y in a neighbourhood of $y = 0$, and moreover we have that $g'(y) = f(\xi(y))$ for all small enough non-negative real y. Since by definition $g(y) = f(\xi(y))$, we have that the zeros of the analytic function $g - g'$ are not isolated, and hence we have that $g = g'$. □

Standard estimates for the binomial coefficients imply the following.

Corollary 3. *For $k \geq 3$ and $n \geq 2$ we have that $t(k,n) \leq (3/2kn)(en)^k$. Moreover, if $k = k(n) \to \infty$ as $n \to \infty$, we have that*

$$\begin{aligned} t(k,n) &= (1+o(1)) \frac{n}{k\sqrt{(2\pi k)}} \frac{(k(n-1))^{k(n-1)}}{(k-2)^{k-2}(kn-2k+2)^{kn-2k+2}} \\ &= (1+o(1)) \frac{1}{k\sqrt{(2\pi k)}} \frac{(n-1)^{k(n-1)+1}}{(n-2)^{kn-2k+2}} \end{aligned}$$

□

Note that in T_n the root has n descendants while all the other vertices have $n-1$ descendants. Let T'_{n-1} be the rooted $(n-1)$-ary tree, *i.e.* the infinite rooted tree all whose vertices have $n-1$ descendants. The tree T'_{n-1} is in a sense more natural than T_n, but when studying Q_p it will be clear that T_n is the tree we should look at, since there is a canonical projection map $T_n \to Q^n$ once we distinguish a vertex of Q^n, where the root of T_n should be mapped to.

Let $t'(k,n)$ be the number of subtrees of T'_{n-1} containing the root of T'_{n-1} and having order k. Then it is intuitively clear that $t(k,n)$ and $t'(k,n)$ should be essentially the same if both k and n are large. Let us remark that one may easily compute $t'(k,n)$ by using a beautiful lemma of Raney (see Theorem 2.1 in [19]).

Indeed, this lemma implies that given a cyclic sequence of k non-negative integers not larger than $n-1$, adding up to $k-1$, there is a unique place where we can break the sequence so that every initial segment of the sequence obtained in this way add up to at least the length of the segment. On the other hand, such a sequence corresponds to a unique rooted plane tree that can be embedded in T'_{n-1}. (Given such a tree, visit the vertices following a breadth-first search from the root, visiting the descendants of a vertex from left to right, say. Let us write down the number of descendants of the vertex v when we first visit v. The sequence obtained in this way determines the tree.)

Therefore $kt'(k,n)$ is the number of cyclic sequences satisfying the condition given above, where each sequence is counted with multiplicity, which corresponds to the number of embeddings that the tree corresponding to the sequence admits. Hence we have that

$$t'(k,n) = \frac{1}{k} \binom{k(n-1)}{k-1},$$

since the number of such cyclic sequences counted with multiplicity is the coefficient of x^{k-1} in

$$(1+x)^{k(n-1)} = \left\{ \sum_{i=0}^{n-1} \binom{n-1}{i} x^i \right\}^k.$$

Let us now go back to T_n. As in the proof of Lemma 2, let us consider $T_p \in \mathcal{G}(T_n, p)$ and let us again denote by T_n^p the component of the root of T_n in T_p. We shall be interested in the probability $P_k(n,p) = \mathbb{P}(|T_n^p| = k)$ that T_n^p has order k.

Lemma 4. *Let $k = k(n) \to \infty$ as $n \to \infty$ and $0 < p = p(n) < 1$. Then*

$$P_k(n,p) = (1+o(1))\frac{(p(n-1))^{k-1}}{k\sqrt{(2\pi k)}}\left\{\frac{(n-1)(1-p)}{(n-2)}\right\}^{nk-2k+2}.$$

Proof. We clearly have that $P_k(n,p) = t(k,n)p^{k-1}(1-p)^{nk-2k+2}$, and the result follows from Corollary 3. □

We shall also be interested in the probability $P_0 = \mathbb{P}(|T_n^p| < \infty)$ that T_n^p is finite. It is clear that

$$\begin{aligned} P_0 &= \sum_{k\geq 1} t(k,n)p^{k-1}(1-p)^{nk-2k+2} \\ &= (1-p)^n + \sum_{k\geq 2}\frac{n}{k-1}\binom{k(n-1)}{k-2}p^{k-1}(1-p)^{nk-2k+2}, \end{aligned}$$

and hence we know P_0 exactly. However, in order to gain a better insight into the behaviour of the order of T_n^p, we shall look at certain branching processes. It will turn out that by using some very basic facts about such processes, we shall be able to derive the asymptotic behaviour of P_0 rather easily. More importantly, we shall see that our results on Q_p are rather natural if we keep branching processes in mind. Let us remark that the use of branching processes in the theory of random graphs goes back to Füredi [14], and that Karp [16] has successfully used them to study random directed graphs.

Let us now very briefly give the relevant definitions and some well-known facts concerning branching processes. For the proofs of the results that we shall use, we refer the reader to the monograph by Harris [15] and to Kolchin [17], Chapter 2. The branching processes we shall be interested in will always be homogeneous, discrete-time processes. Let Z be an integer, non-negative random variable with distribution $p_i = \mathbb{P}(Z = i)$ $(i \geq 0)$. A *Galton–Watson branching process* whose *particles* generate Z *offspring* at a time is a Markov chain $(Z_t)_{t=0}^{\infty}$ that may be described as follows. We start at time $t = 0$ with one particle, that is $Z_0 = 1$; then at time $t \geq 1$, each of the Z_{t-1} particles in *generation* $t-1$ generates Z offspring, independently from all the others. The total number of offspring generated then is Z_t. The most basic fact about such processes is that they are unstable in the sense that with probability 1 a process $(Z_t)_0^{\infty}$ is either such that $Z_t = 0$ for large enough t, or else $Z_t \to \infty$ as $t \to \infty$. In the former case, we say that $(Z_t)_0^{\infty}$ *dies out*.

In studying the probability that a particular process should die out, it is useful to consider the generating function of the distribution of Z. Let us write

$$f(s) = \sum_{i=1}^{\infty} p_i s^i.$$

Let us write π for the probability that our process does not die out, and $\pi' = 1 - \pi$ for the probability that it does die out. Then one can show that $f(\pi') = \pi'$, and in fact, if $\mathbb{E}(Z) > 1$ then π' is the unique solution of $f(s) = s$ strictly between 0 and 1.

We shall need to consider *binomial branching processes*; that is, processes for which the offspring of a particle are generated according to a binomial distribution. Let $0 < p = p(n) < 1$ be given. We again denote the number of offspring of a particle by Z. In our process $\Pi_m = \Pi_m(p)$ $(m \geq 1)$ the r.v. Z obeys the law

$$\mathbb{P}(Z = i) = \binom{m}{i}p^i(1-p)^{m-i}. \tag{5}$$

Let the probability that the process Π_m does not die out be $\pi_m = \pi_m(p)$. We shall also make use of the following branching process $\Pi_0 = \Pi_0(p)$. We start with one particle that generates offspring according to (5) with $m = n$, and all other particles generate offspring according to the same law except that we now have $m = n - 1$. Let the probability that Π_0 does not die out be $\pi_0 = \pi_0(p)$.

We may look at the branching process $\Pi_0(p)$ defined above in the following way. Let $Z_t = Z_t(T_n^p)$ be the number of vertices of T_n^p at distance t from the root of T_n. Then it is a simple observation that $(Z_t)_{t=0}^{\infty}$

is a Markov chain that behaves in the same way as $\Pi_0(p)$. Thus $P_k(n,p) = \mathbb{P}(|T_n^p| = k)$ is the probability that $\Pi_0(p)$ ends up with total progeny k and $P_0 = \mathbb{P}(|T_n^p| < \infty) = \pi_0'(p) = 1 - \pi_0(p)$. Hence we may estimate P_0 by estimating $\pi_0(p)$. In order to estimate $\pi_0(p)$ we shall compare $\Pi_0(p)$ with $\Pi_{n-1}(p)$ and $\Pi_n(p)$.

We shall also consider *Poisson branching processes*; that is, Galton–Watson processes in which the particles generate offspring according to a Poisson distribution. Let us denote by $\Pi_{\mathrm{P}}(\lambda)$ $(\lambda > 0)$ the branching process in which the particles generate Z offspring at a time where

$$\mathbb{P}(Z = i) = \mathrm{e}^{-\lambda}\lambda^i/i!\,,$$

for $i \geq 0$. Let us denote by $\pi_{\mathrm{P}} = \pi_{\mathrm{P}}(\lambda)$ the probability that $\Pi_{\mathrm{P}}(\lambda)$ does not die out.

Note that the generating functions associated with the branching processes $\Pi_m = \Pi_m(p)$ and $\Pi_{\mathrm{P}} = \Pi_{\mathrm{P}}(\lambda)$ are

$$f_m(s) = (1 - (1-s)p)^m$$

and

$$g_\lambda(s) = \mathrm{e}^{(s-1)\lambda},$$

respectively. In particular, if $p = \lambda/m$ and λ is bounded then

$$f_m(s) = g_\lambda(s)\mathrm{e}^{O(1/m)}.$$

The next lemma gives us estimates for the various probabilities of death for the various branching processes defined above.

Lemma 5.

(i) *For all* $0 \leq p \leq 1$, *we have* $\pi_{n-1}(p) \leq \pi_0(p) \leq \pi_n(p)$.

(ii) *If* $\lambda > 1$ *is fixed then* $\pi_{\mathrm{P}}(\lambda)$ *is the unique solution of*

$$x + \mathrm{e}^{-\lambda x} = 1$$

in the interval $0 < x < 1$.

(iii) *Let* $p = \lambda/n$ *where* $\lambda = 1 + \epsilon$ *and* $0 < \epsilon = \epsilon(n) = o(1)$. *Then*

$$\pi_n(p) = \frac{2n\epsilon}{n-1} + O(\epsilon^2).$$

In particular, if $m = n - k$ *then*

$$\pi_m(p) = 2\epsilon + O(\epsilon/n) + O(k/n) + O(\epsilon^2),$$

and hence if $k = o(\epsilon n)$ *then* $\pi_m(p) = (1 + o(1))\pi_0(p)$. □

We may finally state our estimates for $\pi_0(p)$.

Corollary 6. *Let* $p = \lambda/n$.

(i) *If* $\lambda > 1$ *is fixed then* $\pi_0(p) = (1 + o(1))\pi_{\mathrm{P}}(\lambda)$.

(ii) *Let* $\lambda = 1 + \epsilon$ *where* $0 < \epsilon = \epsilon(n) = o(1)$. *Then, if* $m = n - k$ *and* $k = o(\epsilon n)$,

$$\pi_0(p) = (1 + o(1))\pi_m(p) = (2 + o(1))\epsilon$$ □

We close this section with two lemmas concerning graph-theoretic properties of the cube Q^n. The first one is the following compact form of the edge-isoperimetric inequality in the cube Q^n, which will be used often in the sequel (see [7]).

Lemma 7. *Let* $A \subset Q^n$ *be a non-empty set of* k *vertices of* Q^n. *Then the number of edges induced by* A *in* Q^n *is at most* $(k/2)\log_2 k$.

The following lemma gives us an upper estimate for the number of cycles of the n-cube that contain a fixed vertex and have a given length.

Lemma 8. *Let $v \in Q^n$ be a vertex of the cube and $\ell \geq 2$. The number of cycles of length 2ℓ in Q^n containing v is at most*

$$\binom{2\ell}{\ell}\ell! n^{\ell}. \tag{6}$$

Moreover, if $\ell \geq n/32$ then the number of such cycles is at most $2^{-n/11}n(n-1)^{2\ell-1}$.

Proof. We may clearly assume that v is the empty set. Given a cycle C of length 2ℓ containing v, say

$$C: \quad v = v_0, v_1, \ldots, v_{2\ell} = v,$$

let us associate with it the sequence

$$s(C) = (\epsilon_1 i_1, \ldots, \epsilon_{2\ell} i_{2\ell}),$$

where $v_j \bigtriangleup v_{j-1} = \{i_j\}$ and

$$\epsilon_j = \begin{cases} +1 & \text{if } v_j = v_{j-1} \cup \{i_j\} \\ -1 & \text{if } v_j = v_{j-1} \setminus \{i_j\}, \end{cases}$$

$j = 1, \ldots, 2\ell$. Clearly, if $C \neq C'$ then $s(C) \neq s(C')$, and therefore we may bound the number of such cycles by finding an upper bound for the number of possible sequences $s(C)$.

Note that a sequence $s(C)$ has exactly ℓ positive entries, and that the absolute values of these entries are chosen from $[n]$. Also, the set of absolute values of the negative entries equals the corresponding set of the positive ones. Hence the number of such sequences is clearly bounded by (6), which completes the proof of the first part of our lemma.

Let us now assume that $\ell \geq n/32$. We shall now prove our bound by considering a random walk W_0 starting at $v = v_0 = \emptyset$. More precisely, let

$$W_0: \quad v_0, v_1, \ldots, v_{2\ell}$$

be a random walk in Q^n defined as follows. The vertex v_1 is chosen at random from the neighbours of v_0, each with probability $1/n$. For $2 \leq i \leq 2\ell$ the vertex v_i is chosen from the neighbours of v_{i-1}, but it is not allowed to be v_{i-2}, and hence it is one $n-1$ vertices, all such vertices being equiprobable. Note that to prove our estimate it suffices to show that the probability that $v_{2\ell} = v_0$ is at most $2^{-n/11}$.

Let B_0 be the closed Hamming ball of radius $r = \lfloor n/32 \rfloor$ centred at v_0, *i.e.*

$$B_0 = \{v \in Q^n : d(v, v_0) \leq n/32\}.$$

The probability that our random walk W_0 stays in B_0 and $v_{2\ell} = v_0$ is at most

$$\binom{2\ell}{\ell}(1/32)^{\ell} = 2^{-3\ell} \leq 2^{-3n/32} \leq 2^{-n/11}/2,$$

since in such a case there are ℓ steps among the 2ℓ steps in W_0 which are 'towards' v_0. More precisely, for ℓ indices i $(1 \leq i \leq 2\ell)$ we have $|v_i| = |v_{i-1}| - 1$, but this event occurs with probability at most $(r/n)^{\ell} \leq (1/32)^{\ell}$.

Let us now assume that W_0 does go out of B_0 but $v_{2\ell} = v_0$. Set

$$i_0 = \min\{i : v_j \in B_0,\ j \geq i\}.$$

Then clearly $|v_{i_0}| = r = \lfloor n/32 \rfloor$ and the number of remaining steps in W_0 is $t = 2\ell - i_0 \geq r$. Amongst these, exactly $t_0 = (t+r)/2$ must be towards $v_0 = \emptyset$. Now, the probability that this event happens is at most

$$\sum_{t \geq r} \binom{t}{t_0}(r/(n-1))(1/32)^{t_0 - 1} = 2^{-5r/2} \sum_{t \geq r} 2^{-3t/2} = 2^{1-4r} \leq 2^{-n/11}/2,$$

and hence $\mathbb{P}(v_{2\ell} = v_0) \leq 2^{-n/11}$, which completes the proof. □

3. The subcritical phase

In this section we shall study the behaviour of the largest component of $Q_p \in \mathcal{G}(Q^n, p)$ for $p < 1/(n-1)$. We start with a result that tells us that, in this range of p, all components of Q_p are very small indeed. We remark that the upper bound on ϵ in the result below is merely a technical restriction, and it suffices to assume that ϵ should be bounded away from 1, say.

Theorem 9. *Let $p = (1-\epsilon)/(n-1)$, where $0 < \epsilon = \epsilon(n) \le 1/2$. Then a.e. $Q_p \in \mathcal{G}(Q^n, p)$ contains no components of order larger than*

$$k_0 = -\frac{n \log 2}{\epsilon + \log(1-\epsilon)} + \frac{\log(1-(1-\epsilon)\mathrm{e}^{\epsilon})}{\epsilon + \log(1-\epsilon)}. \tag{7}$$

Proof. We start by noticing the following. Let C_v be the component of v in Q_p, and T_n^p be the component of the root of T_n in $T_p \in \mathcal{G}(T_n, p)$. Then

$$\mathbb{P}(|C_v| \le k) \ge \mathbb{P}(|T_n^p| \le k) = \sum_{1 \le \ell \le k} P_\ell(n, p).$$

Indeed, we may generate C_v by the following iterative procedure. At stage k, we test which of the edges of Q^n incident to the vertices of C_v, at distance k from v and going to vertices not in the component so far, are present in our random graph. Then, at each stage, for every vertex there are at most $n-1$ cube edges to test, and our claim follows, since if we generate T_n^p with the analogous procedure, for every such vertex the number of T_n-edges to test is exactly $n-1$.

It follows from the claim above that the probability that v belongs to a component of order larger than k_0 is at most

$$P_0 = 1 - \sum_{1 \le k \le k_0} P_k(n, p).$$

Since $p(n-1) = 1 - \epsilon < 1$, we see that $\sum_{k \ge 1} P_k(n,p) = 1$ (cf. the proof of Lemma 2). Hence $P_0 = \sum_{k > k_0} P_k(n,p)$. Thus the expected number of vertices in components of order $k > k_0$ is, by Lemma 4, at most

$$N \sum_{k \ge k_0} k^{-3/2} (1-\epsilon)^{k-1} \left(1 + \frac{\epsilon}{n-2}\right)^{k(n-2)+2}$$
$$\le 4N k_0^{-3/2} \sum_{k \ge k_0} [(1-\epsilon)\mathrm{e}^{\epsilon}]^k \le 4N k_0^{-3/2} \frac{[(1-\epsilon)\mathrm{e}^{\epsilon}]^{k_0}}{1-(1-\epsilon)\mathrm{e}^{\epsilon}} = o(1),$$

and hence a.e. Q_p contains no component of order larger than k_0. □

Our task for the rest of this section is to show that the above theorem is sharp. We shall be able to show that the largest component of Q_p ($p = (1-\epsilon)/(n-1)$) is essentially as given in (7), provided $\epsilon > 0$ is not too small.

Let $k \ge 1$ and $v \in Q^n$ be fixed. We shall need to know the probability that v belongs to a component of order k in $Q_p \in \mathcal{G}(Q^n, p)$. Here and in the sequel, we shall denote by $C_v = C_v(Q_p)$ the connected component of v in $Q_p \in \mathcal{G}(Q^n, p)$. Thus we shall be interested in the probability that $|C_v| = k$. Now, this probability may be estimated by running a probabilistic algorithm that, intuitively speaking, generates a spanning tree of C_v.

Let us consider Algorithm I, given in Figure 1. This algorithm generates a tree $T_v \subset Q^n$ by a certain breadth-first search. Since the n edges incident to a fixed vertex of the cube are in one-to-one correspondence with the elements of $[n]$, there is a natural ordering on these edges induced by the ordering on $[n]$. Algorithm I is based on the breadth-first search where the edges are searched in this order.

Clearly, the probability that the component $C_v \subset Q_p$ is of order k equals the probability that the tree T_v generated by Algorithm I is of order k. The proof of Lemma 10 below will be based on this simple fact.

```
begin
  $T_v := v$; $R := \emptyset$
  insert $v$ into a queue $Q$;
  while $Q$ not empty do begin
    $w :=$ 1st element of $Q$;
    delete $w$ from $Q$;
    let $w_1, \dots, w_m$ be the neighbours of $w$ in $Q^n$
      not contained in $T_v$, in their natural order;
    for $i = 1, \dots, m$ do begin
      $R := R \cup \{ww_i\}$;
      with probability $p$ do begin
        add the edge $ww_i$ and the vertex $w_i$ to $T_v$;
        put $w_i$ at the back of $Q$
        end
      end
    end;
  Output $T_v$
end
```

Figure 1. Algorithm I

So far we have considered T_n merely as a rooted tree, but let us now fix a labelling of the edges of T_n with elements of $[n]$ with the property that every vertex of T_n is incident to an edge of every label. In other words, we are now looking at T_n with a fixed proper edge-colouring with colours from $[n]$. Recalling that the edges incident to a fixed vertex of Q^n are in one-to-one correspondence with $[n]$, given a vertex v of Q^n, there is a natural homomorphism of graphs $\tau_v : T_n \to Q^n$ taking the root of T_n to v. The lemma below connects, via τ_v, the component C_v of v in Q_p to T_n^p, the component of the root of T_n in a random subgraph $T_p \in \mathcal{G}(T_n, p)$.

Lemma 10. *Let $n \geq 2$, $k \geq 1$, and $0 < p < 1$. For all vertices $v \in Q^n$ we have that*

$$\mathbb{P}(|C_v| = k) \geq \mathbb{P}(\tau_v(T_n^p) \text{ is acyclic and } |T_n^p| = k). \tag{8}$$

Proof. Let $T_0 \subset Q^n$ be a subtree of Q^n that contains v. Clearly Algorithm I can generate T_0 in a unique way, and hence there is an integer $K = K(T_0) \geq 0$ such that

$$\mathbb{P}(T_v = T_0) \geq \mathbb{P}(T_v = T_0)(1-p)^K = \mathbb{P}(T_n^p = T_0'),$$

where $T_0' \subset T_n$ is the unique subtree of T_n containing the root of T_n such that $\tau_v(T_0') = T_0$. On the other hand,

$$\mathbb{P}(|C_v| = k) = \sum \mathbb{P}(T_v = T_0),$$

where the sum is over all subtrees $T_0 \subset Q^n$ of order k that contain v. Therefore the probability that the vertex v belongs to a component of order k is at least

$$\sum \mathbb{P}(T_n^p = T_0'), \tag{9}$$

where the sum is over all subtrees $T_0' \subset T_n$ that contain the root of T_n, have order k, and are such that $\tau_v(T_0') \subset Q^n$ is acyclic. But (9) is exactly the right-hand side of (8), and hence the proof is complete. □

It should be clear that our general aim is to show that, for values of k comparable with k_0 in (7), the probability that $|C_v| = k$ is not very far off from the probability that $|T_n^p| = k$. In view of Lemma 10 above, it suffices to show that the restriction that $\tau_v(T_n^p) \subset Q^n$ should be acyclic is not a very strong one for these values of k. Thus our next task is to analyse the probability that $\tau_v(T_n^p)$ is acyclic. Since this probability clearly does not depend on $v \in Q^n$, let us assume that $v = \emptyset$ and write τ_0 for $\tau_\emptyset$.

In the sequel we shall use the fact that we may generate T_n^p by looking at $\Pi_0(p)$. Recall from Section 2 that these two random objects are very closely related. Indeed, we may clearly regard $\Pi_0(p)$ as a random rooted tree T_{bp} (the 'genealogical tree'), and hence we may generate T_n^p by first generating T_{bp} and then choosing a random rooted embedding $h : T_{\mathrm{bp}} \to T_n$. The following lemma is immediate.

Lemma 11. *Let $n \geq 2$, $k \geq 1$, $\Delta \geq 2$, and $0 < p < 1$. We have that*

$$\begin{aligned}
&\mathbb{P}(\tau_0(T_n^p) \text{ is acyclic and } |T_n^p| = k)\\
&\qquad \geq \mathbb{P}(\tau_0 h(T_{\mathrm{bp}}) \text{ is acyclic} \,\big|\, |T_{\mathrm{bp}}| = k,\, \Delta(T_{\mathrm{bp}}) \leq \Delta)\\
&\qquad\qquad \times \mathbb{P}(|T_{\mathrm{bp}}| = k \text{ and } \Delta(T_{\mathrm{bp}}) \leq \Delta). \qquad \square
\end{aligned}$$

Let a rooted tree T be given. We shall now estimate the probability that a random rooted embedding $h : T \to T_n$ is such that $\tau_0 h(T) \subset Q^n$ is acyclic. Our estimate will be in terms of the order $|T|$ of T and of the maximal degree $\Delta(T)$ of T.

Lemma 12. *Let $k = k(n) \leq n^3$ and $2 \leq \Delta = \Delta(n) \leq \surd(n/32\log n)$. Then*

$$\mathbb{P}(\tau_0 h(T_{\mathrm{bp}}) \text{ is acyclic} \,\big|\, |T_{\mathrm{bp}}| = k,\, \Delta(T_{\mathrm{bp}}) \leq \Delta) = 1 - O(k\Delta^4/n^2).$$

Proof. Fix a rooted tree T_0 of order k with $\Delta(T_0) \leq \Delta$. We shall show that a random embedding h of T_0 in T_n is such that $\tau_0 h(T_0)$ is acyclic with probability $1 - O(k\Delta^4/n^2)$.

Pick a random h, and assume that $\tau_0 h(T_0)$ is *not* acyclic. We claim that, for some $2 \leq \ell \leq k/2$, there is a cycle $C = C^{2\ell}$ in $\tau_0 h(T_0)$ and a path $P = P^{2\ell+1}$ in T_0 such that $\tau_0 h(P) = C$. In order to see this, let us first fix an ordering $e_1, \ldots, e_{k-1}$ of the edges of T_0 satisfying the the property that any initial segment $e_1, \ldots, e_i$ of it induces a subtree $T_i = T_0[\{e_1, \ldots, e_i\}]$ of T_0. Let i_0 be the minimal i for which $\tau_0 h(T_i)$ contains a cycle C. Let $v_1 \in T_0$ be the vertex added to T_{i_0-1} by the edge e_{i_0} to obtain T_{i_0}. Let $v_2 \in T_{i_0-1}$ be such that $\tau_0 h(v_2) = \tau_0 h(v_1)$. Then if we take P to be the path connecting v_1 and v_2 in $T_{i_0} \subset T_0$, we have that $\tau_0 h(P) = C$, as claimed.

Fix a cycle $C = C^{2\ell}$ and a path $P = P^{2\ell+1}$ as above, and let v be the vertex of P that is nearest to the root of T_0. Let T_1 be the subtree of T_0 induced by the descendants of v, and let its root be v. Clearly, our random embedding h is such that $\tau_0 h(T_1)$ contains a cycle that goes through $\tau_0 h(v)$. Suppose we show that

$$\mathbb{P}(\tau_0 h(T_1) \text{ contains a cycle through } v_0 = \emptyset) = O(\Delta^4/n^2), \tag{10}$$

for all rooted trees T_1 or order at most k and $\Delta(T_1) \leq \Delta$. Then it follows that the probability that a random embedding h of T_0 is acyclic is at least $1 - O(k\Delta^4/n^2)$, which is the claim of our lemma.

Let us now proceed to prove (10). Let us assume that $\tau_0 h(T_1)$ contains a cycle $C = C^{2\ell}$ of length 2ℓ. Let $P = P^{2\ell+1} \subset T_1$ be a path in T_1 that contains the root and projects to C by $\tau_0 h$. The probability that $\tau_0 h$ does map P onto C is $n^{-1}(n-1)^{-2\ell+1}$, and hence the probability in (10) is bounded from above by

$$P_0(\ell) = \alpha(T_1, \ell)\beta(n, \ell)(n-1)^{-2\ell}, \tag{11}$$

where $\alpha(T_1, \ell)$ is the number of paths $P = P^{2\ell+1}$ in T_1 that have order $2\ell + 1$ and contain the root of T_1, and $\beta(n, \ell)$ is the number of 2ℓ-cycles in Q^n containing $v_0 = \emptyset$. We shall analyse three ranges of ℓ in order to estimate $P_0(\ell)$.

Case 1. $2 \leq \ell \leq n/12\Delta^2$

It is easily checked that $\alpha(T_1, \ell) \leq 4(\ell + 1)\Delta^{2\ell}$. Hence, by (11) and Lemma 8, we have that

$$P_0(\ell) \leq 4(\ell+1)\Delta^{2\ell}\binom{2\ell}{\ell}\ell! n^\ell (n-1)^{-2\ell} \leq c\ell\left(\frac{4\Delta^2\ell}{en}\right)^\ell,$$

for some absolute constant c. Hence the probability that $\tau_0 h(T_1)$ contains a cycle of length 2ℓ ($2 \le \ell \le n/12\Delta^2$) going through $v_0 = \emptyset$ is at most

$$\sum_{\ell=2}^{\lfloor n/12\Delta^2 \rfloor} P_0(\ell) \le \sum_{\ell=2}^{\lfloor n/12\Delta^2 \rfloor} c\ell \left(\frac{4\Delta^2 \ell}{\mathrm{e}n}\right)^{\ell} \le 4c \left(\frac{8\Delta^2}{\mathrm{e}n}\right)^2 = O(\Delta^4/n^2).$$

Case 2. $n/12\Delta^2 \le \ell \le n/9$

Clearly $\alpha(T_1, \ell) \le k^2$. Hence, by (11),

$$P_0(\ell) \le k^2 \binom{2\ell}{\ell} \ell! n^{\ell} (n-1)^{2\ell} \le ck^2 \left(\frac{4\ell}{\mathrm{e}n}\right)^{\ell},$$

for some absolute constant c. Hence the probability that $\tau_0 h(T_1)$ contains a cycle of length 2ℓ ($n/12\Delta^2 \le \ell \le n/9$) going through $v_0 = \emptyset$ is at most

$$\begin{aligned}\sum_{n/12\Delta^2 \le \ell \le n/9} P_0(\ell) &\le \sum_{n/12\Delta^2 \le \ell \le n/9} ck^2 \left(\frac{4\ell}{\mathrm{e}n}\right)^{\ell} \\ &\le 2ck^2 \left(\frac{4\lceil n/12\Delta^2 \rceil}{\mathrm{e}n}\right)^{\lceil n/12\Delta^2 \rceil} = O\left[k^2 (2\mathrm{e}\Delta^2)^{-n/12\Delta^2}\right].\end{aligned}$$

Case 3. $n/9 \le \ell \le k/2$

By Lemma 8, we know that in this range of ℓ we have that $\beta(n, \ell) \le 2^{-n/11} n(n-1)^{2\ell-1}$. Then $P_0(\ell) \le k^2 2^{-n/11}$, and this completes the proof of (10).

As remarked above, relation (10) implies the assertion of the lemma. □

We now need to show that T_n^p tends to have very small maximal degree. Given a rooted tree T, let us write $d^+(v)$ for the number of direct descendants of $v \in T$, *i.e.* for the number of vertices adjacent to v which are further away from the root than v. Also, let $\Delta_{\mathrm{d}}(T) = \sup_{v \in T} d^+(v)$. Let $t(k, \Delta_0, n)$ denote the number of trees $T_0 \subset T_n$ that contain the root of T_n, have order k, and are such that $\Delta_{\mathrm{d}}(T_0) \ge \Delta_0$.

Lemma 13. *Let $k = k(n) \to \infty$ as $n \to \infty$ and $\Delta_0 = \Delta_0(n) \ge 1$. Then*

$$t(k, \Delta_0, n)/t(k, n) = O\left[k^{5/2} (\mathrm{e}/\Delta_0)^{\Delta_0}\right]. \tag{12}$$

Moreover, if $0 < p = p(n) < 1$, then

$$\mathbb{P}(\Delta_{\mathrm{d}}(T_n^p) \ge \Delta_0 \,\big|\, |T_n^p| = k) = O\left[k^{5/2} (\mathrm{e}/\Delta_0)^{\Delta_0}\right]. \tag{13}$$

Proof. Note that, for a fixed value of p, when randomly selecting T_n^p all trees $T^k \subset T_n$ of order k that contain the root of T_n are equiprobable. Therefore

$$\mathbb{P}(\Delta_{\mathrm{d}}(T_n^p) \ge \Delta_0 \,\big|\, |T_n^p| = k) = t(k, \Delta_0, n)/t(k, n),$$

and hence (12) and (13) are equivalent. Furthermore, to prove our lemma it is enough to show that (13) holds for a single value of p.

Let us fix $p_0 = 1/(n-1)$. Note that

$$\begin{aligned}\mathbb{P}(\Delta_{\mathrm{d}}(T_n^{p_0}) \ge \Delta_0 \text{ and } |T_n^{p_0}| \le k) &\le \sum_{\Delta \ge \Delta_0} k\mathbb{P}(S_{n-1,p_0} = \Delta) \\ &\le \sum_{\Delta \ge \Delta_0} k \binom{n-1}{\Delta} p_0^{\Delta} \le 2k(\mathrm{e}/\Delta_0)^{\Delta_0}.\end{aligned} \tag{14}$$

On the other hand, by Lemma 4, we have that

$$P_k(n, p_0) = \mathbb{P}(|T_n^{p_0}| = k) = (1 + o(1))(2\pi k^3)^{-1/2}, \tag{15}$$

if $k = k(n) \to \infty$. Relation (13) follows from (14) and (15). □

Let us remark that a stronger form of Lemma 13 may be proved by invoking Raney's lemma, mentioned in Section 2. For our purposes, the result above will suffice. Putting together Lemmas 10 to 13, we have the following.

Corollary 14. *Let* $0 < p = p(n) < 1$, $k = k(n) \leq n^3$, *and* $8(\log n)/\log\log n \leq \Delta = \Delta(n) \leq \sqrt{(n/32\log n)}$. *Then*

$$\mathbb{P}(|C_v| = k) \geq (1 - O(k\Delta^4/n^2))P_k(n, p)/2.$$ □

We are finally ready to prove a lower bound for the order of the largest component of $Q_p \in \mathcal{G}(Q^n, p)$, where $p = (1 - \epsilon)/n$ and $\epsilon > 0$ is not too small.

Theorem 15. *Let* $p = (1-\epsilon)/(n-1)$, *where* $(\log n)^2/(\log\log n)\sqrt{n} < \epsilon = \epsilon(n) \leq 1$. *Then a.e.* $Q_p \in \mathcal{G}(Q^n, p)$ *contains at least* $N^{1/2\log\log\log n}$ *vertices in components of order at least*

$$-\frac{n \log 2}{\epsilon + \log(1-\epsilon)}\left(1 - \frac{2}{\log\log\log n}\right).$$

Proof. Let $X = X(Q_p)$ be the number of rooted components of order

$$k_0 = -\left\lfloor \frac{n \log 2}{\epsilon + \log(1-\epsilon)}\left(1 - \frac{1}{\log\log\log n}\right)\right\rfloor.$$

Note that $k_0 \leq n^3$. By letting

$$\Delta = \left\lfloor \frac{9\log n}{\log\log n}\right\rfloor,$$

Lemma 4, Corollary 14, and some easy calculations imply that the expectation $\mathbb{E}(X)$ of X is at least $2N^{1/2\log\log\log n}$. In order to prove our lemma, we shall show that X is highly concentrated around its expectation. We shall estimate the variance $\sigma^2 = \sigma^2(X)$ of X and then apply Chebyshev's inequality.

Given a connected subgraph C of Q^n and a vertex $v \in C$, let us write C^v for the ordered pair (C, v). Note that $X = \sum X_{C^v}$, where the summation ranges over all C^v with $|C| = k_0$. Let us first estimate $\mathbb{E}[X(X-1)]$. Clearly

$$\mathbb{E}[X(X-1)] = \sum \mathbb{E}(X_{C^v} X_{D^w}),$$

where the summation ranges over all ordered pairs (C^v, D^w), $C^v \neq D^w$, $|C|$, $|D| = k_0$. Let us break the sum into three parts, according to the type of pair (C^v, D^w). Let us write $s_1 = \sum_1 \mathbb{E}(X_{C^v} X_{D^w})$, where $\sum_1$ indicates that the sum ranges over pairs (C^v, D^w) where $C = D$; let us write $s_2 = \sum_2 \mathbb{E}(X_{C^v} X_{D^w})$, where $\sum_2$ ranges over pairs (C^v, D^w) such that the distance between C and D is at least two; and finally let us write s_3 for the rest of the sum.

We easily see that

$$s_1 \leq k_0 \sum_{C^v} \mathbb{E}(X_{C^v}) \leq n^3 \mathbb{E}(X), \tag{16}$$

and that

$$s_2 = \sum\nolimits_2 \mathbb{E}(X_{C^v})\mathbb{E}(X_{D^w}) \leq (\mathbb{E}(X))^2. \tag{17}$$

Let us now look at the sum $s_3 = \sum_3 \mathbb{E}(X_{C^v} X_{D^w})$, where the sum ranges over pairs (C^v, D^w) for which the distance between C and D is at most 1. Clearly we may assume that this distance *is* 1. For each such pair, choose an arbitrary edge e of Q^n joining C to D. Note that

$$\begin{aligned} &\mathbb{P}(C \text{ and } D \text{ are both components of } Q_p) \\ &\qquad = \mathbb{P}(C \cup D \cup e \text{ is a component of } Q_p)(1-p)/p \\ &\qquad \leq n\mathbb{P}(C \cup D \cup e \text{ is a component of } Q_p). \end{aligned}$$

Let Z be the random variable counting the number of quadruples (C, v, w, e) where C is a component of Q_p of order $2k_0$, v and w are vertices in C and e is an edge in C. Let X' count the number of rooted components C^v of Q_p with $|C| = 2k_0$. Note that then

$$Z \leq 2k_0^2(\log 2k_0)X' \leq 7n^6(\log n)X'.$$

Now, by computations analogous to the ones in the proof of Theorem 9, we see that $\mathbb{E}(X') = o(\alpha^n)$ for some absolute constant $0 < \alpha < 1$. Hence

$$s_3 \leq n\mathbb{E}(Z) \leq 7n^7(\log n)\mathbb{E}(X') = o(1). \tag{18}$$

Writing $\mu = \mathbb{E}(X)$, we see from (16), (17) and (18) that

$$\sigma^2(X) \leq (1 + n^3)\mu + o(1),$$

and hence $\sigma^2/\mu^2 \leq 2n^3/\mu$. The result now follows from Chebyshev's inequality. □

Corollary 16. *Let $p = (1 - \epsilon)/n$ where $\epsilon = \epsilon(n) \geq (\log n)^2/(\log\log n)\sqrt{n}$ is bounded away from 1. Then a.e. $Q_p \in \mathcal{G}(Q^n, p)$ is such that*

$$L_1(Q_p) = -\frac{n\log 2}{\epsilon + \log(1-\epsilon)}\left(1 + O\left(\frac{1}{\log\log\log n}\right)\right). \qquad □$$

4. A gap in the sequence of components

In this section we shall prove a lemma that will be important in the proof of the fact that there is a unique giant component in $Q_p \in \mathcal{G}(Q^n, p)$ if p is a little larger than $1/n$. This lemma will concern the orders of the various components of Q_p. Roughly speaking, it will tell us that if p is a little larger than $1/n$ every component of Q_p tends to be either very large or quite small, and hence there is a 'gap' in the sequence of the orders of the components of Q_p. Let us start with a simple lemma. For brevity, let us call a component of a graph G a *k-component* if it has order k. Also, in what follows we write $S_{m,p}$ for a random variable with binomial distribution with parameters m and p.

Lemma 17. *Let $0 < p < 1$ and $k \geq 1$ be such that $(\log n)^2 pk \log k \geq 1$. Let $m_1 = \lfloor (k/2)\log_2 k \rfloor$. Then the probability that a vertex of $Q_p \in \mathcal{G}(Q^n, p)$ belongs to a k-component of size at least $k + \lfloor \mathrm{e}(\log n)pm_1 \rfloor$ is at most*

$$\exp\left\{-\frac{1}{2}(\log n)(\log\log n)pk\log k\right\}.$$

Proof. We shall make use of Algorithm I from Section 3. Let $v \in Q^n$ be fixed. Note that we can generate the component C_v of v in Q_p in the following way. We first run Algorithm I and get a tree $T_v \subset Q^n$ as output. Then we look at the edges $E(Q^n[V(T)])$ induced by $V(T_v)$ in Q^n in turn, and decide which ones should be added to T_v to form C_v. More specifically, we run the following probabilistic algorithm on T_v.

```
$C_v := T_v$;
for all $e \in E(Q^n[V(C_v)]) \setminus R$ do
    with probability $p$ do add $e$ to $C_v$;
Output $C_v$
```

The output C_v is the component of v in Q_p. Let $\ell_1 = k + \lfloor \mathrm{e}(\log n)pm_1 \rfloor$. Now, the probability that a vertex v of Q_p belongs to a k-component C_v of size $e(C_v)$ at least ℓ_1 is

$$\sum_T \mathbb{P}(T_v = T \text{ and } e(C_v) \geq \ell_1) = \sum_T \mathbb{P}(e(C_v) \geq \ell_1 \mid T_v = T)\mathbb{P}(T_v = T),$$

where the sum ranges over trees $T \subset Q^n$ that contain v and have order k. The conditional probability above is $\mathbb{P}(S_{m,p} \geq 1 + \lfloor \mathrm{e}(\log n)pm_1 \rfloor)$, where $m = |E(Q^n[V(C_v)]) \setminus R|$. Hence, by standard estimates for the tail of the binomial distribution and the edge-isoperimetric inequality in Q^n, this probability is at most

$$\begin{aligned}\mathbb{P}\left(S_{m_1,p} \geq \mathrm{e}(\log n)pm_1\right) &\leq \exp\left\{-(\log n)(\log\log n)pm_1\right\} \\ &\leq \exp\left\{-\frac{1}{2}(\log n)(\log\log n)pk\log k\right\},\end{aligned}$$

as required. □

We are now ready to prove the main result of this section.

Lemma 18. *Let $C \geq 1$ and $D \geq 125$ be fixed. Let $\epsilon_0 = 4C(\log n)^3/n$.*

(i) If $p = (1+\epsilon)/n$, where $\epsilon_0 \leq \epsilon = \epsilon(n) \leq 1$, then the following assertions hold.

(a) A.e. Q_p has no k-component with $Dn/\epsilon^2 \leq k \leq n^{C\log n}$.

(b) A.e. Q_p is such that the total number of vertices in k-components, $n/20\epsilon \leq k \leq \lceil Dn/\epsilon^2 \rceil$, is at most N/n.

(ii) For any $0 \leq t = t(n) \leq M = nN/2$, let $\epsilon_t = \epsilon(t) = 2t/N - 1$. Then a.e. cube process $(Q_t)_0^M$ is such that, for any $t_0 = \lceil (1+\epsilon_0)N/2 \rceil \leq t \leq N$, the graph Q_t has no k-component with $Dn/\epsilon_t^2 \leq k \leq n^{C\log n}$. Moreover, a.e. such process is such that for $t \geq t_0$ every component of Q_t has order smaller than n^3 or larger than $n^{C\log n}$.

Proof. We shall consider the two spaces $\mathcal{G}(Q^n, p)$ and $\mathcal{G}(Q^n, t)$ for certain p and t. Let $\mathbb{E}_p$ denote expectation in the first space and $\mathbb{E}_t$ in the second. Let us write $X_k = X_k(G)$ for the number of rooted k-components of the graph G. Clearly X_k counts the total number of vertices of G that belong to k-components. Let us now start the proof of (i).

(i) We shall consider two cases according to the size of ϵ.

Case 1. $\epsilon \geq 2(\log n)(7C/n)^{1/2}$

By Corollary 3 and the edge-isoperimetric inequality in Q^n, for $k \geq 3$,

$$\begin{aligned}\mathbb{E}_p(X_k) &\leq Nt(k,n)p^{k-1}(1-p)^{kn-k\log_2 k} \\ &\leq \frac{3N}{2kn}(\mathrm{e}n)^k\left(\frac{1+\epsilon}{n}\right)^{k-1}\left(1-\frac{1+\epsilon}{n}\right)^{kn-k\log_2 k} \\ &\leq N\mathrm{e}^k(1+\epsilon)^k\exp\left\{-k(1+\epsilon)\left(1-\frac{\log_2 k}{n}\right)\right\}. \qquad (19)\end{aligned}$$

Therefore

$$\begin{aligned}\log\mathbb{E}_p(X_k) &\leq n(\log 2) + k + k\log(1+\epsilon) - k(1+\epsilon)\left(1-\frac{\log_2 k}{n}\right) \\ &\leq n(\log 2) + k + k\left(\epsilon - \frac{\epsilon^2}{2} + \frac{\epsilon^3}{3}\right) - k(1+\epsilon)\left(1-\frac{\log_2 k}{n}\right) \\ &\leq n(\log 2) - \frac{k\epsilon^2}{2} + \frac{k\epsilon^3}{3} + \frac{2k\log_2 k}{n}.\end{aligned}$$

Let us now assume that $k \leq n^{C\log n}$. Then

$$\log\mathbb{E}_p(X_k) \leq n(\log 2) - k\epsilon^2/42. \qquad (20)$$

We can now prove (a) using (20). Indeed, if $k \geq Dn\epsilon^{-2} > 112(\log 2)n\epsilon^{-2}$, then

$$\log\mathbb{E}_p(X_k) \leq -5(\log 2)n/3,$$

and hence

$$\sum_{\lceil Dn\epsilon^{-2}\rceil}^{\lfloor n^{C\log n}\rfloor} \mathbb{E}_p(X_k) \le n^{C\log n}2^{-5n/3}, \tag{21}$$

which proves (a).

To prove (b), we simply note that (20) tells us that

$$\sum_{\lceil n/20\epsilon\rceil}^{\lceil Dn\epsilon^{-2}\rceil} \mathbb{E}_p(X_k) \le \exp\left\{-n^{1/2}\right\} N.$$

Case 2. $\epsilon \le 2(\log n)(7C/n)^{1/2}$

Let $X_{k,\ell,m} = X_{k,\ell,m}(G)$ denote the number of rooted k-components C of G that have size $e(C) = \ell$ and are such that the number $e(Q^n[V(C)])$ of edges induced by $V(C)$ in Q^n is m. Let $m_1 = m_1(k) = \lfloor (k/2)\log_2 k\rfloor$. Then, by the edge-isoperimetric inequality in Q^n,

$$X_k = \sum_{m=k-1}^{m_1}\sum_{\ell=k-1}^{m} X_{k,\ell,m}.$$

We shall assume throughout the proof that $n/20\epsilon \le k \le n^{C\log n}$. Write $X_{k,m}^{(1)}$ for the number of rooted k-components of size less than $\ell_1 = k + \lfloor \mathrm{e}(\log n)pm_1\rfloor$ that span m edges in Q^n, and $X_{k,m}^{(2)}$ for the rest of $X_{k,m} = \sum_\ell X_{k,\ell,m}$. Since, by Lemma 1,

$$\begin{aligned}\mathbb{E}_p(X_{k,\ell,m}) &= Nc(k,\ell,m)p^\ell(1-p)^{nk-m-\ell}\\ &\le 3N(nk)^{1/2}\binom{nk-m}{\ell}p^\ell(1-p)^{nk-m-\ell},\end{aligned}$$

the expected number $\mathbb{E}_p(X_{k,m}^{(1)})$ of vertices in k-components spanning m edges in Q^n with size less than ℓ_1 is at most $3N(nk)^{1/2}\mathbb{P}(S_{nk-m,p} < \ell_1)$. Let us write $\mu = \mathbb{E}(S_{nk-m,p}) = p(nk-m)$. Note that

$$\mu - (k + \mathrm{e}(\log n)pm_1) \ge k\left(\epsilon - \frac{p\log_2 k}{2} - \frac{\mathrm{e}}{2}(\log n)p\log_2 k\right) \ge k\epsilon/2,$$

and also that $2k \ge \mu \ge k$. Therefore

$$\mathbb{P}(S_{nk-m,p} < \ell_1) \le \mathbb{P}(|S_{nk-m,p} - \mu| \ge \epsilon\mu/4) \le \exp\left\{-\frac{1}{3}\cdot\frac{\epsilon^2}{16}\cdot k\right\} = \exp(-\epsilon^2 k/48),$$

and hence

$$\mathbb{E}_p(X_{k,m}^{(1)}) \le 3N(nk)^{1/2}\exp(-\epsilon^2 k/48). \tag{22}$$

On the other hand, Lemma 17 tells us that

$$\sum_m \mathbb{E}_p(X_{k,m}^{(2)}) \le N\exp\left\{-\frac{1}{2}(\log n)(\log\log n)pk\log k\right\}. \tag{23}$$

We are now ready to conclude the proof of our lemma in this case.

Let us deal with (a) first. Let k be such that $Dn/\epsilon^2 \le k \le n^{C\log n}$. By (22),

$$\mathbb{E}_p(X_{k,m}^{(1)}) \le 3N(nk)^{1/2}\exp\{-(D/48)n\} \le \mathrm{e}^{-5n/2}N \tag{24}$$

and, by (23),

$$\sum_m \mathbb{E}_p(X_{k,m}^{(2)}) \le N\exp\left\{-\frac{1}{2}(\log n)^2(\log\log n)\cdot\frac{1}{n}\cdot\frac{Dn}{\epsilon^2}\right\} \le \mathrm{e}^{-3n}N. \tag{25}$$

Thus

$$\sum_{k=\lceil Dn/\epsilon^2\rceil}^{\lfloor n^{C\log n}\rfloor} \sum_{m=k-1}^{m_1} \left\{\mathbb{E}_p(X^{(1)}_{m,k}) + \mathbb{E}_p(X^{(2)}_{m,k})\right\} = o(1),$$

and hence a.s. Q_p has no k-components with $Dn/\epsilon^2 \le k \le n^{C\log n}$.

Let us now prove (b). Let k be such that $n/20\epsilon \le k \le \lceil Dn/\epsilon^2\rceil$. By (22),

$$\mathbb{E}_p(X^{(1)}_{k,m}) \le 3N(nk)^{1/2}\exp\left\{-(\log n)^3/240\right\} \le Nn^{-(\log n)^2/241}, \tag{26}$$

and also, by (23),

$$\sum_m \mathbb{E}_p(X^{(2)}_{k,m}) \le N\exp\left\{-\frac{1}{2}(\log n)^2(\log\log n)\cdot\frac{1}{n}\cdot\frac{n}{20\epsilon}\right\} \le \exp\left\{-n^{1/2}\right\} N. \tag{27}$$

Therefore, by (26) and (27),

$$\sum_{k=\lceil n/20\epsilon\rceil}^{\lceil Dn/\epsilon^2\rceil} \sum_{m=k-1}^{m_1} \left\{\mathbb{E}_p(X^{(1)}_{k,m}) + \mathbb{E}_p(X^{(2)}_{k,m})\right\} \le Nn^{-\log n},$$

which implies (b). This completes the proof of Case 2, and hence the proof of (i) is complete. Let us now turn to the proof of (ii).

(ii) For $0 \le t \le M = nN/2$ let $p = p(t) = t/M$. Clearly, if $\epsilon = \epsilon_t$ is as defined in assertion (ii) of our lemma then $p = (1+\epsilon)/n$. Let us first note that, for any r.v. X,

$$\begin{aligned}\mathbb{E}_p(X) &= \sum_{m=0}^{M} \binom{M}{m} p^m(1-p)^{M-m}\mathbb{E}_m(X)\\ &\ge \binom{M}{t} p^t(1-p)^{M-t}\mathbb{E}_t(X)\\ &\ge (1/3)M^{-1/2}\mathbb{E}_t(X).\end{aligned} \tag{28}$$

We shall now estimate the sum $\sum_{t_0\le t\le N}\sum_k \mathbb{E}_p(X_k)$ by breaking it into two parts. Let $\epsilon_1 = 2(\log n)(7C/n)^{1/2}$, and set $t_1 = \lfloor(1+\epsilon_1)N/2\rfloor$. Now, by (21),

$$\sum_{t=t_1}^{N} \sum_{k=\lceil Dn\epsilon^{-2}\rceil}^{\lfloor n^{C\log n}\rfloor} \mathbb{E}_p(X_k) \le Nn^{C\log n}2^{-5n/3} \le n^{C\log n}2^{-2n/3} = o(M^{-1/2}).$$

On the other hand, we recall that

$$X_k = \sum_{m=k-1}^{m_1} \left\{\mathbb{E}_p(X^{(1)}_{k,m}) + \mathbb{E}_p(X^{(2)}_{k,m})\right\},$$

and hence, by using (24) and (25), we see that

$$\sum_{t=t_0}^{t_1} \sum_{k=\lceil Dn\epsilon^{-2}\rceil}^{\lfloor n^{C\log n}\rfloor} \mathbb{E}_p(X_k) \le Nn^{C\log n}\left\{n^{C\log n+3}N\mathrm{e}^{-5n/2} + N\mathrm{e}^{-3n}\right\} = o(M^{-1/2}).$$

Thus, by (28), we have that

$$\sum_{t=t_0}^{N} \sum_{k=\lceil Dn\epsilon^2\rceil}^{\lfloor n^{C\log n}\rfloor} \mathbb{E}_t(X_k) = o(1),$$

```
begin
  $H_p(v) := v$;
  insert $v$ into a queue $Q$;
  while $Q$ not empty do begin
    $w :=$ 1st element of $Q$;
    delete $w$ from $Q$;
    let $w_1, \dots, w_{n-1-u}$ be the first $n-1-u$ neighbours
      of $w$ in $Q^n$ at distance at least 5 from $w$ in $H_p(v)$;
    for $i = 1, \dots, n-1-u$ do
      with probability $p$ do begin
        add the edge $ww_i$ and the vertex $w_i$ to $H_p(v)$;
        if $H_p(v)$ is not acyclic
            or $\Delta(H_p(v)) > \log n$ or $|H_p(v)| = n_0$
          then Output $H_p(v)$ and Halt
          else put $w_i$ at the back of $Q$
        end
    end;
  Output $H_p(v)$
end
```

Figure 2. Algorithm II

as required.

Let us turn to the second statement in (*ii*). It suffices to consider the range $N < t \le M = nN/2$, since we know that the statement holds for $t_0 \le t \le N$. Now, in the range $N \le t \le M$, relation (19) applies, and we have that

$$\mathbb{E}_p(X_k) \le N \left\{ (1+\epsilon) \exp\left(-\epsilon + (1+\epsilon)\frac{\log_2 k}{n} \right) \right\}^k ,$$

and hence, if $k \le n^{C \log n}$ and $\epsilon \ge 1$, we have that

$$\mathbb{E}_p(X_k) \le N \left[(1+\epsilon) \mathrm{e}^{-(1+o(1))\epsilon} \right]^k ,$$

and so $\mathbb{E}_p(X_k) \le N(3/4)^k$, say. We now simply note that

$$\sum_{t=N}^{M} \sum_{k=n^3}^{\lfloor n^{C \log n} \rfloor} \mathbb{E}_p(X_k) \le \sum_{N}^{M} n^{C \log n} N (3/4)^{n^3} \le n^{1 + C \log n} N^2 (3/4)^{n^3} = o(M^{-1/2}),$$

and hence, by (28),

$$\sum_{N}^{M} \sum_{n^3}^{\lfloor n^{C_2} \rfloor} \mathbb{E}_t(X_k) = o(1). \qquad \square$$

5. The vertices in the large components

Let a constant $C \ge 1$ be fixed. Let us call a component of $Q_p \in \mathcal{G}(Q^n, p)$ *large* if it has order at least $n^{C \log n}$, and *small* if it has order less than n^3. In most of this section, we shall be dealing with the problem of estimating the total number of vertices in large components of Q_p. In order to carry out this estimation, we shall consider the binomial branching processes that we introduced in Section 2.

Let $n_0 \ge 1$ be fixed and let $u = \lfloor \log n \rfloor^2$. Recall that we consider the edges incident to a fixed vertex v of Q^n ordered by the ordering induced by the natural ordering on $[n]$. Thus the neighbours of v are also naturally ordered. Let us now consider Algorithm II, given in Figure 2.

Note that this probabilistic algorithm simulates, up to a certain point, the binomial branching process $\Pi_{n-1-u}(p)$. Moreover, roughly speaking, it may be thought of as an algorithm for generating the component C_v of v in Q_p. It must be noted however that certain edges of Q^n are examined twice in this algorithm, and hence have probability $p + (1-p)p$ of belonging to $H_p(v)$. In the next lemma, we analyse the behaviour of Algorithm II.

Lemma 19. *Let v be any vertex of Q^n, $1/n \le \epsilon = \epsilon(n) \le 1$, $p = (1+\epsilon)/n$, and $n_0 = \lceil n/20\epsilon \rceil$.*
(i) The probability that $H_p(v)$ is acyclic, is such that $\Delta(H_p(v)) \le \log n$, and has order strictly smaller than n_0 is at most $1 - \pi_{n-1-u}(p)$.
(ii) The probability that $H_p(v)$ contains a cycle or contains a vertex of degree greater than $\log n$ is at most $O(1/n)$.
(iii) The probability that $H_p(v)$ contains n_0 vertices and is acyclic is at least

$$\pi_{n-1-u}(p) + O(1/n).$$

Proof. (i) We only output an $H_p(v)$ such that $|H_p(v)| < n_0$, $\Delta(H_p(v)) \le \log n$ and furthermore is acyclic if the queue became empty in our generation of $H_p(v)$. This corresponds to our simulation of the branching process $\Pi_{n-1-u}(p)$ having died out, which happens with probability $1 - \pi_{n-1-u}(p)$.

(ii) Let us first estimate the probability that $H_p(v)$ contains a vertex of degree greater than $\log n$ and is acyclic. Clearly this happens only if, when running Algorithm II, for some vertex $w \in H_p(v)$ exactly $\lfloor \log n \rfloor$ neighbours of w are added to $H_p(v)$ in the **for** loop. But this happens with probability at most

$$\mathbb{P}(S_{n-1-u,p} \ge \lfloor \log n \rfloor) \le \mathbb{P}(S_{n-1-u,p} \ge (2/3)\log n) \le n^{-(\log\log n)/3\mathrm{e}}.$$

Let us now estimate the probability that $H_p(v)$ contains a cycle. The probability that $H_p(v)$ contains a cycle of length 2ℓ is at most

$$\left\lceil \frac{n}{20\epsilon} \right\rceil \binom{2\ell}{\ell} n^\ell \ell! \frac{4}{n} \left(\frac{2}{n}\right)^{2\ell-1} \le \frac{1}{6}\cdot\frac{n}{\epsilon}\left(\frac{16\ell}{\mathrm{e}n}\right)^\ell \le n^2 \left(\frac{16\ell}{\mathrm{e}n}\right)^\ell,$$

where the unexpected term $4/n$ comes from the fact that the last edge added to form the cycle might have been checked twice by Algorithm II. Thus the probability that $H_p(v)$ contains a cycle of length 2ℓ with $3 \le \ell \le n/32$ is at most

$$\sum_{3 \le \ell \le n/32} n^2 \left(\frac{16\ell}{\mathrm{e}n}\right)^\ell \le 2n^2 \left(\frac{48}{\mathrm{e}n}\right)^3 = O(1/n).$$

The probability that $H_p(v)$ contains a cycle of length 2ℓ, $n/32 < \ell \le n/40\epsilon$, is at most

$$\left\lceil \frac{n}{20\epsilon} \right\rceil 2^{-n/11} n^{2\ell} 2 \left(\frac{1+\epsilon}{n}\right)^{2\ell} \le \frac{n}{9\epsilon} 2^{-n/11} \mathrm{e}^{2\ell\epsilon} \le n^2 2^{-n/11} \mathrm{e}^{n/20}.$$

Hence the probability that $H_p(v)$ contains a 2ℓ-cycle with $n/32 < \ell \le n/40\epsilon$ is exponentially small.

(iii) This follows from (i) and (ii). □

It is intuitively clear that the probability that Algorithm II generates an $H_p(v)$ of order n_0 is a lower bound for the probability that C_v has order at least n_0. Indeed, as remarked above, Algorithm II essentially generates the component C_v of v in Q_p, except for the fact that certain edges of Q^n are examined twice in the **while** loop, and hence these edges are more likely to be in $H_p(v)$ than they are likely to be in Q_p. However, these edges are always edges whose deletion does not disconnect C_v, and hence they are irrelevant as far as the order $|C_v|$ of C_v is concerned. We make this statement precise in the proof of the corollary below.

```
for $i = 1, \dots, n-1-u$ do
  if $w_i \notin H_p(v)$ then
    with probability $p$ do begin
      add the edge $ww_i$ and the vertex $w_i$ to $H_p(v)$;
      if $\Delta(H_p(v)) > \log n$ or $|H_p(v)| = n_0$
        then Output $H_p(v)$ and Halt
        else put $w_i$ at the back of $Q$
      end
```

Figure 3. The **for** loop for Algorithm III

Corollary 20. *Let $p = (1+\epsilon)/n$ where $1/n \le \epsilon = \epsilon(n) \le 1$. Let v be a fixed vertex of Q^n and denote by C_v the connected component of $Q_p \in \mathcal{G}(Q^n, p)$ that contains v. The probability that $|C_v| \ge n_0 = \lceil n/20\epsilon \rceil$ is at least $\pi_{n-1-u}(p) + O(1/n)$.*

Proof. Let us consider a third probabilistic algorithm, Algorithm III, which is very similar to Algorithm II, except that it never examines an edge more than once. In fact, Algorithm III may be obtained from Algorithm II by replacing the **for** loop in Algorithm II by the **for** loop in Figure 3.

For clarity, let us denote the output of Algorithm III by $H'_p(v)$. Clearly

$$\mathbb{P}(|C_v| \ge n_0) \ge \mathbb{P}(|H'_p(v)| = n_0).$$

On the other hand, for any tree $T_0 \subset Q^n$ containing v, with $\Delta(T_0) \le \log n$ and $|T_0| = n_0$,

$$\mathbb{P}(H'_p(v) = T_0) \ge \mathbb{P}(H'_p(v) = T_0)(1-p)^K = \mathbb{P}(H_p(v) = T_0),$$

where $K = K(T_0) \ge 0$ depends only on the tree T_0 (cf. the proof of Lemma 10). Thus, by summing over all such trees T_0, we see that

$$\mathbb{P}(|H'_p(v)| = n_0) \ge \mathbb{P}(|H_p(v)| = n_0),$$

and the result follows from Lemma 19(iii). □

We can now estimate from below the number of vertices in large components. We know from Lemma 18(i) that, provided p is a little larger than $1/n$, a.s. every component of $Q_p \in \mathcal{G}(Q^n, p)$ is either small or large. We shall first estimate from above the number of vertices in components of the former type.

Lemma 21. *Let $p = (1+\epsilon)/n$ where $1/n \le \epsilon = \epsilon(n) \le 1$. Then a.e. $Q_p \in \mathcal{G}(Q^n, p)$ is such that the number of vertices in components of order smaller than $n_0 = \lceil n/20\epsilon \rceil$ is at most*

$$\left\{1 - \pi_{n-1-u}(p) + O\left(\frac{1}{n}\right)\right\} N.$$

Proof. By Corollary 20, if $X = X(Q_p)$ is the number of vertices in components of order smaller than $n_0 = \lceil n/20\epsilon \rceil$, then its expectation $\mathbb{E}(X)$ is at most $(1 - \pi_{n-1-u}(p) + O(1/n))N$. To prove our lemma, we shall show that X is highly concentrated around its expectation. We shall estimate the variance $\sigma^2 = \sigma^2(X)$ of X and then apply Chebyshev's inequality. We shall estimate the variance by the same method we used in the proof of Theorem 15.

Given a connected subgraph C of Q^n and a vertex $v \in C$, let us write C^v for the ordered pair (C, v). Note that $X = \sum X_{C^v}$, where the summation ranges over all C^v with $|C| < n_0$. Let us first estimate $\mathbb{E}[X(X-1)]$. Clearly

$$\mathbb{E}[X(X-1)] = \sum \mathbb{E}(X_{C^v} X_{D^w}),$$

where the summation ranges over all ordered pairs (C^v, D^w), $C^v \ne D^v$, $|C|$, $|D| < n_0$. Let us break the sum into three parts, according to the type of pair (C^v, D^w). Let us write $s_1 = \sum_1 \mathbb{E}(X_{C^v} X_{D^w})$, where $\sum_1$

indicates that the sum ranges over pairs (C^v, D^w) where $C = D$; let us write $s_2 = \sum_2 \mathbb{E}(X_{C^v} X_{D^w})$, where $\sum_2$ ranges over pairs (C^v, D^w) such that the distance between C and D is at least two; and finally let us write s_3 for the rest of the sum.

Now, we easily see that

$$s_1 \leq n_0 \sum_{C^v} \mathbb{E}(X_{C^v}) \leq n^2 \mathbb{E}(X), \tag{29}$$

and that

$$s_2 = \sum\nolimits_2 \mathbb{E}(X_{C^v}) \mathbb{E}(X_{D^w}) \leq (\mathbb{E}(X))^2 . \tag{30}$$

Let us now look at the sum $s_3 = \sum_3 \mathbb{E}(X_{C^v} X_{D^w})$, where the sum ranges over pairs (C^v, D^w) for which the distance between C and D is at most 1. Clearly we may assume that this distance *is* 1. For each such pair, choose an arbitrary edge e of Q^n joining C to D. Note that

$$\begin{aligned} \mathbb{P}(&C \text{ and } D \text{ are both components of } Q_p) \\ &= \mathbb{P}(C \cup D \cup e \text{ is a component of } Q_p)(1-p)/p \\ &\leq n\mathbb{P}(C \cup D \cup e \text{ is a component of } Q_p). \end{aligned}$$

Let Z be the random variable counting the number of quadruples (C, v, w, e) where C is a component of Q_p of order at most $2n_0 - 2$, v and w are vertices in C and e is an edge in C. Let X' count the number of rooted components C^v of Q_p with $|C| \leq 2n_0 - 2 < 2n_0$. Note that then

$$Z \leq 2n_0^2 (\log 2n_0) X' \leq n^4 (\log n) X'.$$

Hence

$$s_3 \leq n\mathbb{E}(Z) \leq n^5 (\log n) \mathbb{E}(X') \leq n^5 (\log n) N. \tag{31}$$

To finish the proof, we shall need a lower bound for $\mu = \mathbb{E}(X)$. It turns out that the following extremely crude bound suffices. If X_1 counts the number of isolated vertices of Q_p, then

$$\mu = \mathbb{E}(X) \geq \mathbb{E}(X_1) = N(1-p)^n \geq N\mathrm{e}^{-pn/2} \geq N/\mathrm{e}^{1/2}.$$

We now note that (29), (30) and (31) imply that

$$\frac{\sigma^2}{\mu^2} \leq \frac{(1+n^2)\mu + n^5(\log n)N}{\mu^2} \leq \mathrm{e}^{1/2}(1+n^2)/N + \mathrm{e}n^5(\log n)/N = O(n^5(\log n)/N),$$

and the lemma follows from Chebyshev's inequality. □

In the lemma below we show that the upper bound for the total number of vertices in small components given in Lemma 21 is sharp.

Lemma 22. *Let $p = (1+\epsilon)/(n-1)$, where $0 < \epsilon = \epsilon(n) \leq 1$. Then a.e. $Q_p \in \mathcal{G}(Q^n, p)$ contains at least $(1 - \pi_0(p) + O(N^{-1/3}))N$ vertices in components of order at most $4n/\epsilon^2$.*

Proof. We shall again use the fact that the probability that a fixed vertex of the cube belongs to a component of order at most k is at least as large as the probability that $|T_n^p|$ is at most k (cf. the proof of Theorem 9). Let $k_0 = \lfloor 4n/\epsilon^2 \rfloor$.

We claim that

$$\mathbb{P}(|T_n^p| \leq k_0) = \mathbb{P}(|T_n^p| < \infty) + o(\mathrm{e}^{-n}) = 1 - \pi_0(p) + o(\mathrm{e}^{-n}). \tag{32}$$

Indeed, by Lemma 4 we have that

$$\begin{aligned} \mathbb{P}(k_0 \leq &|T_n^p| < \infty) = \sum_{k \geq k_0} P_k(n,p) \\ &\leq \sum_{k \geq k_0} \left[(1+\epsilon)\mathrm{e}^{-\epsilon}\right]^k \leq \sum_{k \geq k_0} \exp\{-k\epsilon^2/3\} = o(\mathrm{e}^{-n}), \end{aligned}$$

which proves the claim. From (32), we see that the expected number of vertices in components of order at most k_0 is at least $\{1 - \pi_0(p) + o(\mathrm{e}^{-n})\}N$. The method used in the proof of Lemma 21 can now be used to complete the proof of this lemma. □

We can now give our estimate for the total number of vertices in large components of $Q_p \in \mathcal{G}(Q^n, p)$.

Corollary 23. *Let $C \geq 1$ be fixed and $\epsilon_0 = 4C(\log n)^3/n$.*

(i) If $p = (1+\epsilon)/n$, where $\epsilon_0 \leq \epsilon = \epsilon(n) \leq 1$, then a.e. $Q_p \in \mathcal{G}(Q^n, p)$ is such that the total number of vertices in components of order at least $n^{C\log n}$ is $(1+o(1))\pi_0(p)N$.

(ii) If $t = \lfloor (1+\epsilon)N/2 \rfloor$, where $\epsilon_0 \leq \epsilon = \epsilon(n) \leq 1$, then a.e. $Q_t \in \mathcal{G}(Q^n, t)$ is such that the total number of vertices in components of order at least $n^{C\log n}$ is $(1+o(1))\pi_0(p)N$, where $p = (1+\epsilon)/n$.

Proof. (i) This follows from Lemmas $18(i)$, 21, and 22. Indeed, by Lemma 21 we know that the total number of vertices in k-components ($k < n/20\epsilon$) is at most $\{1 - \pi_{n-1-u}(p) + O(1/n)\}N$, and by Lemmas $18(i)(b)$ and 22 the number of such vertices is at least $\{1 - \pi_0(p) + O(1/n)\}N$. Now, by Lemmas $18(i)(a)$ and (b), we conclude that the total number of vertices in k-components ($k \geq n^{C\log n}$) is $(1+o(1))\pi_0(p)N$.

(ii) This follows from (i) by convexity. □

In Corollary 23 above, we established that if $\epsilon > 0$ is not too small then fairly many vertices of $Q_p \in \mathcal{G}(Q^n, p)$, where $p = (1+\epsilon)/n$, belong to large components. In the final lemma in this section, we shall turn our attention to the distribution of such vertices in Q^n. We shall see that Corollary 23 easily implies that these vertices are in a sense extremely well distributed in the cube.

Lemma 24. *Let $C \geq 2$ be fixed and set $p = (1+\epsilon_0)/n$ where $\epsilon_0 = 4C(\log n)^3/n$. Then a.e. $Q_p \in \mathcal{G}(Q^n, p)$ is such that every vertex of Q^n is at Hamming distance at most $\log_2 n$ from a component of order at least $n^{C\log n}$.*

Proof. Let v_0 be a fixed vertex of Q^n. Let us estimate from above the probability that v_0 fails to be at distance at most $s = \lfloor \log_2 n \rfloor$ from a component of order at least $n^{C\log n}$. Without loss of generality we may assume that v_0 is the empty set. Let $k = 2s$ and let Q_i ($1 \leq i \leq \ell$) be $\ell = \binom{k}{s}$ disjoint subcubes of dimension $m = n - k$ which are at Hamming distance s from v_0. For instance, we may let Q_i be the subcube of Q^n induced by

$$\{v \in Q^n : v \cap [k] = S_i\},$$

where S_i is the ith s-subset of $[k] = \{1, \ldots, k\}$ in some fixed ordering on $[k]^{(s)}$. Let $v_i \in Q_i$ be a vertex at distance s from v_0. Note that

$$\frac{1+\epsilon_0}{n} = \frac{1+\epsilon_0}{m}\left(1 - \frac{k}{n}\right) \geq \frac{1+(1+o(1))\epsilon_0}{m}.$$

Therefore, by Corollary 23, the probability that v_i belongs to a component of Q_p of order at least $m^{C\log m} \geq n^{(C/2)\log n}$ is $(2+o(1))\epsilon_0$. Hence the probability that v_0 fails to be at distance at most s from a vertex in a component of order at least $n^{(C/2)\log n}$ is, rather crudely, at most

$$(1-\epsilon_0)^\ell \leq \exp\left\{-\frac{\log n}{n}\binom{k}{s}\right\} = o(2^{-n}).$$

Our lemma now follows from Markov's inequality, and the fact that in a.e. Q_p a component of order at least $n^{(C/2)\log n}$ has in fact order at least as large as $n^{C\log n}$, by Lemma $18(i)(a)$. □

6. The emergence of the giant component

We are now ready to prove the main results of this note. We start by showing that a.e. $Q_t \in \mathcal{G}(Q^n, t)$ has one large component only, provided t is a little larger than $N/2$.

Theorem 25. *Let $C \geq 12$ be fixed. Then a.e. cube process $\widetilde{Q} = (Q_t)_0^M$ is such that, for every*

$$t \geq t_1 = \left\lfloor \left\{1 + \frac{5C(\log n)^3}{n}\right\} \frac{N}{2} \right\rfloor,$$

the graph Q_t has a unique component of order at least $n^{C\log n}$ and all the others have order smaller than n^3.

Proof. Let us say that a component of Q_t is large if it has order at least $n^{C\log n}$ and small if it has order smaller than n^3. Let $\epsilon_0 = 4C(\log n)^3/n$ and $t_0 = \lceil (1+\epsilon_0)N/2 \rceil$. Lemma 18(*ii*) states that a.e. $\widetilde{Q} = (Q_t)_0^M$ is such that if $t \geq t_0$ then every component of Q_t is either small or large. Therefore our theorem will follow if we prove that for a.e. such $\widetilde{Q}$, for some $t_0 \leq t \leq t_1$, all large components of Q_{t_0} belong to the same component in Q_t.

By Lemma 24, we may assume that every vertex of Q^n is at distance at most $s = \lfloor \log_2 n \rfloor$ from a large component of Q_{t_0}. Let us fix such a graph $H = Q_{t_0}$, and show that with the addition of very few edges almost surely all large components of H merge into a single component. More precisely, let $p_0 = n^{-2}$, and randomly pick $Q_0 \in \mathcal{G}(Q^n, p_0)$. Let us consider $G = H \cup Q_0$. Note that G has a.s. at most $t_0 + 2N/3n < t_1$ edges, and hence our theorem follows if we show that a.s. all large components of H belong to a single component in G.

Let the number of large components in H be ℓ. Assume that we can split the large components of H into two classes such that there are no paths in G, and hence in Q_0, between vertices that belong to components in different classes. Let the number of components in one of the classes be k, and assume that $k \leq \ell/2$. Let the set of the vertices in the components in one class be S, and the corresponding set for the other class be T. Clearly we have that $|S|$ and $|T|$ are at least as large as $kn^{C\log n}$.

Recall that all vertices of Q^n are at distance at most s from $S \cup T$. Let S' and T' form a partition of the vertex-set of the cube such that $S \subset S'$, $T \subset T'$, and every vertex of S' (resp. T') is at distance at most s from S (resp. T). We shall now make use of the following result related to vertex-isoperimetric inequalities in Q^n (see [8]). For any subset $A \subset Q^n$ of the vertices of the cube, let $\delta^-(A)$ denote the elements in A that are adjacent to elements in $A^{\rm c} = Q^n \setminus A$, the complement of A. Similarly, let $\delta^+(A) = \delta^-(A^{\rm c})$. For $1 \leq a \leq N$, let $\delta^-(a) = \min \delta^-(A)$ and $\delta^+(a) = \min \delta^+(A)$ where we take the minimum over all subsets $A \subset Q^n$ with $|A| = a$. Then for any $A \subset Q^n$ there exists a matching between A and $A^{\rm c}$, and hence between $\delta^-(A)$ and $\delta^+(A)$, of size at least $\min\{\delta^-(a), \delta^+(a)\}$. In particular, the size of the matching is at least $(1/\sqrt{n}) \min\{|A|, |A^{\rm c}|\}$. Thus Q^n has a matching between S' and T' of size at least $kn^{C\log n - 1/2}$. From such a matching, we can obtain a collection of at least $u = kn^{C\log n - 1/2 - 2s}/2s$ edge-disjoint paths of Q^n between S and T, all of them of length at most $2s+1$. By our assumption on G, none of these paths are paths in Q_0. Note that this happens with probability at most $P_0 = (1 - p_0^{2s+1})^u$. However, the number of partitions of the large components of H into two classes with one of them having k members is clearly at most

$$\binom{N/n^{C\log n}}{k}.$$

Therefore

$$\sum_{k=1}^{\lfloor \ell/2 \rfloor} P_0 \binom{N/n^{C\log n}}{k} \leq \sum_{1}^{\lfloor \ell/2 \rfloor} \left[N \exp(-p_0^{2s+1} u) \right]^k \leq \sum_{1}^{\lfloor \ell/2 \rfloor} (N{\rm e}^{-n})^k = o(1),$$

completing the proof that a.e. Q_0 is such that all large components of H belong to a single component in $G = H \cup Q_0$. □

Corollary 26. *Let $C \geq 12$ be fixed and $t = \lfloor (1+\epsilon)N/2 \rfloor$, where $5C(\log n)^3/n \leq \epsilon = \epsilon(n) \leq 1$. Then a.e. $Q_t \in \mathcal{G}(Q^n, t)$ has a unique component of order at least $n^{C\log n}$ and all other components of Q_t are of order at most $125\epsilon^{-2}n$.*

Proof. This is immediate from Lemma 18(*ii*) and Theorem 25. □

Corollary 27. *Let $C \geq 12$ be fixed and $p = (1+\epsilon)/n$, where $\epsilon = \epsilon(n) \geq 5C(\log n)^3/n$ is bounded away from 1. Then a.e. $Q_p \in \mathcal{G}(Q^n, p)$ has a unique component of order at least $n^{C\log n}$ and all other components of Q_p are of order at most $125\epsilon^{-2}n$.*

Proof. Let us say that a subgraph of Q^n satisfies property A if it has a unique component of order at least $n^{C\log n}$ and all other of components are of order at most $125\epsilon^{-2}n$. Corollary 26 states that there is a function $\psi(n) = o(1)$ such that for all $5C(\log n)^3/n \le \epsilon = \epsilon(n) \le 1$

$$\mathbb{P}(Q_t \text{ fails } A) \le \psi(n).$$

Thus if $s = \omega(n)N^{1/2}$ and $\omega(n) \to \infty$, then

$$\begin{aligned}\mathbb{P}(Q_p \text{ fail } A) &= \sum_{t=0}^{M} \binom{M}{t} p^t (1-p)^{M-t} \mathbb{P}(Q_t \text{ fails } A) \\ &\le \sum_{|t-pM|\le s} \binom{M}{t} p^t (1-p)^{M-t} \mathbb{P}(Q_t \text{ fails } A) \\ &\qquad\qquad + \sum_{|t-pM|>s} \binom{M}{t} p^t (1-p)^{M-t} \\ &\le \psi(n) + o(1) = o(1)\end{aligned}$$

□

Theorem 25 tells us that a unique large component emerges sometime soon after time $N/2$. Let us now invoke the lemmas in Section 5 to give an estimate for the order of this giant component.

Theorem 28. *Let $t = N/2 + s$. If $30N(\log n)^3/n \le s = o(N)$ then a.e. $Q_t \in \mathcal{G}(Q^n, t)$ is such that*

$$L_1(Q_t) = (4 + o(1))s$$

and

$$L_2(Q_t) \le 125\epsilon^{-2} n,$$

where $\epsilon = \epsilon(n) = 2s/N$.

Proof. This follows from Lemma 6(*iii*), Corollary 23(*ii*), and Corollary 26. □

Theorem 29. *Let $1 < c \le 2$ be a constant and set $t = \lfloor cN/2 \rfloor$. Let $\eta = \eta(c)$ be the unique solution of $x + \mathrm{e}^{-cx} = 1$ in the interval $0 < x < 1$. Then a.e. $Q_t \in \mathcal{G}(Q^n, t)$ is such that*

$$L_1(Q_t) = (\eta + o(1))N,$$

and furthermore

$$L_2(Q_t) \le \left[\frac{\log 2}{c - 1 - \log c} + o(1)\right] n.$$

Proof. The statement concerning $L_1(Q_t)$ follows Lemma 5(*ii*), Corollary 23(*ii*), and Corollary 26. The inequality concerning $L_2(Q_t)$ can be proved by calculations similar to the ones in the proof of Lemma 18.

□

7. The second largest component

This section will be devoted to prove a lower bound for the order of the second largest component. To simplify the computations, we shall deal with $Q_p \in \mathcal{G}(Q^n, p)$ instead of $Q_t \in \mathcal{G}(Q^n, t)$. Let us start with the following simple lemma.

Lemma 30. *Let p_1 and p_2 be such that $p_2 n < 1 < p_1 n$ and $p_1(1-p_1)^n = p_2(1-p_2)^n$. Let v be a fixed vertex of Q^n and let $\rho(p, k)$ denote the probability that v is contained in a component of order k in $Q_p \in \mathcal{G}(Q^n, p)$. Then*

$$\frac{p_2}{p_1} \le \frac{\rho(p_1, k)}{\rho(p_2, k)} \le \left(\frac{1-p_2}{1-p_1}\right)^{k \log_2 k}.$$

Proof. We are interested in estimating $\rho(p, k) = \mathbb{P}(|C_v| = k)$, where $C_v = C_v(Q_p)$ is the component of Q_p containing v. We shall again use Algorithm I from Section 3, which randomly generates a tree $T_v \subset Q^n$. Let $\rho_T(p, k)$ be the probability that we generate a fixed tree $T \subset Q^n$ by Algorithm I. Then

$$\rho(p, k) = \mathbb{P}(|C_v| = k) = \sum \rho_T(p, k),$$

where the sum ranges over all trees $T \subset Q^n$ which contain v and have order k. Also,

$$\rho_T(p, k) = p^{k-1}(1-p)^{kn-2m+\alpha},$$

where $m = m(T)$ is the number of edges induced by $V(T)$ in Q^n, and $\alpha = \alpha(T)$ is the number of such edges checked by the Algorithm I in the generation of T. By the edge-isoperimetric inequality in Q^n we have that $m \le (k/2)\log_2 k$. Thus $2m - \alpha \le k \log_2 k$.

Hence

$$\begin{aligned}\frac{\rho_T(p_1, k)}{\rho_T(p_2, k)} &= \frac{p_1^{k-1}(1-p_1)^{nk-2m+\alpha}}{p_2^{k-1}(1-p_2)^{nk-2m+\alpha}}\\ &= \frac{p_2}{p_1}\left(\frac{1-p_2}{1-p_1}\right)^{2m-\alpha} \le \left(\frac{1-p_2}{1-p_1}\right)^{k\log_2 k}.\end{aligned}$$

Also, as clearly $2m - \alpha \ge m \ge k - 1 \ge 0$, we have that

$$\frac{\rho_T(p_1, k)}{\rho_T(p_2, k)} \ge \frac{p_2}{p_1}.$$

The result now follows by summing over all trees $T \subset Q^n$ of order k containing v. □

We are now ready to prove a precise bound for the order of the second largest component of $Q_p \in \mathcal{G}(Q^n, p)$, where p is a little larger than $1/n$.

Theorem 31. *Let $p = (1+\epsilon)/n$ where $(\log n)^2/(\log\log n)\sqrt{n} \le \epsilon = \epsilon(n) \le 1$. Then a.e. $Q_p \in \mathcal{G}(Q^n, p)$ has at least $N^{1/2 \log\log\log n}$ components of order at least*

$$\frac{n \log 2}{\epsilon - \log(1+\epsilon)}\left(1 + O\left(\frac{1}{\log\log\log n}\right)\right). \tag{33}$$

Moreover, the second largest component of a.e. Q_p is of order at most (33).

Proof. Let $X(p, k) = X_{p,k}(Q_p)$ be the number of rooted k-components of Q_p. Let p' be such that $p'(1-p')^n = p(1-p)^n$ and $p'n < 1$. Note that if $p' = (1-\epsilon')/n$, then

$$\epsilon' + \log(1-\epsilon') = \log(1+\epsilon) - \epsilon + O(1/n).$$

Moreover, since

$$\left(\frac{1-p'}{1-p}\right)^{k\log_2 k} \leq \left(1+\frac{3\epsilon}{n}\right)^{k\log_2 k} \leq \exp\left(\frac{3\epsilon k\log_2 k}{n}\right),$$

we have that for some absolute constant $c>0$

$$c\,\mathbb{E}\left[X(p',k)\right] \leq \mathbb{E}\left[X(p,k)\right] \leq \exp\left(\frac{3\epsilon k\log_2 k}{n}\right)\mathbb{E}\left[X(p',k)\right]. \tag{34}$$

Now, for some integer k^+ satisfying

$$\begin{aligned} k^+ &= -\frac{n\log 2}{\epsilon' + \log(1-\epsilon')}\left(1+O\left(\frac{1}{\log\log\log n}\right)\right) \\ &= \frac{n\log 2}{\epsilon - \log(1+\epsilon)}\left(1+O\left(\frac{1}{\log\log\log n}\right)\right), \end{aligned}$$

computations analogous to the ones in te proof of Theorem 9 show that

$$\sum_{k\geq k^+} \mathbb{E}[X(p',k)] = o(N^{-1/\log\log\log n}). \tag{35}$$

By (34) and (35) we have that

$$\sum_{k=k^+}^{\lfloor 125n/\epsilon^2\rfloor} \mathbb{E}[X(p,k)] \leq \exp\left\{\frac{1125\log_2 n}{\epsilon}\right\} N^{-1/\log\log\log n} = o(1),$$

and hence a.e. Q_p has no k-component with $k^+ \leq k \leq 125n/\epsilon^2$. On the other hand, by Corollary 27 a.e. Q_p has no component of order larger than $125n/\epsilon^2$ besides the giant, and hence $L_2(Q_p) < k^+$.

Furthermore, as in the proof of Theorem 15, for some integer k^- satisfying

$$\begin{aligned} k^- &= -\frac{n\log 2}{\epsilon' + \log(1-\epsilon')}\left(1+O\left(\frac{1}{\log\log\log n}\right)\right) \\ &= \frac{n\log 2}{\epsilon - \log(1+\epsilon)}\left(1+O\left(\frac{1}{\log\log\log n}\right)\right), \end{aligned}$$

we have that $\mathbb{E}\left[X(p',k)\right] \geq (1-o(1))N^{1/2\log\log\log n}$. We can now complete the proof by applying Chebyshev's inequality, after estimating the variance with the method used in the proof of Lemma 21. □

Let us close this section with the following theorem, in which we have compiled our main results concerning the component structure of Q_p.

Theorem 32. *Let $p=(1+\epsilon)/n$ where $60(\log n)^3/n \leq \epsilon = \epsilon(n) \leq 1$. Then a.e. $Q_p \in \mathcal{G}(Q^n,p)$ is such that*

$$L_1(Q_p) = (1+o(1))\pi_0(p)N.$$

Moreover, if $\epsilon \geq (\log n)^2/(\log\log n)\sqrt{n}$ and $k\geq 2$ is a fixed integer then a.e. Q_p is such that

$$L_k(Q_p) = \frac{n\log 2}{\epsilon - \log(1+\epsilon)}\left(1+O\left(\frac{1}{\log\log\log n}\right)\right).$$

Proof. This follows from Corollary 23(i), Theorem 25, and Theorem 31. □

8. Concluding remarks and open problems

From Theorems 9 and 25, we know that the critical time for Q^n-processes is somewhere in the interval $\{1+1/(n-1)\}N/2 \leq t \leq \{1+O((\log n)^3/n)\}N/2$. As remarked in the introduction, it is a little surprising that the critical point should not be $N/2$, when the average degree is 1. Indeed, that is the case for ordinary random graph processes. However, this is duly explained by the fact that the 'reason' why the value $t = n/2$ is important for ordinary processes is that the corresponding value of p is $1/(n-1)$. The proofs of the results given in this note indicate that $p = 1/(n-1)$ might again be the critical probability for $Q_p \in \mathcal{G}(Q^n, p)$.

On the other hand, the proof of Theorem 9 is rather crude since the only facts we used about Q^n were that it is n-regular and that it has order $N = 2^n$. One can probably do better and show that a similar result to Theorem 9 holds for values of p slightly larger than $1/(n-1)$. In any case, it would be most interesting to determine the 'critical probability' and to investigate the order of the largest component of Q_p at that point.

The proof of Theorem 24 is rather simple-minded, and probably it is there where one might want to put in some work to improve our results. Indeed, we believe that by improving both Theorem 24 and the argument in Theorem 25 one might be able to prove the following.

Conjecture 33. *There is a constant $C > 0$ for which the following holds. Let $t = N/2+s$. If $(C\log n)N/2n \leq s = o(N)$ then a.e. $Q_t \in \mathcal{G}(Q^n, t)$ is such that*

$$L_1(Q_t) = (4+o(1))s$$

and

$$L_2(Q_t) = (c+o(1))\epsilon^{-2}n,$$

where $\epsilon = \epsilon(s) = 2s/N$ and $c > 0$ is an absolute constant. □

In [6], we investigate a problem closely related to the one studied in this note. We prove the existence of a phase transition in the component structure of the subgraph of Q^n induced by a random set of vertices. Let a random induced subgraph Q^p of the cube Q^n be chosen by letting $\mathbb{P}(v \in V(G^p)) = p$, all such events being independent. Let $p = (1+\epsilon)/n$. Weber [21] observed that if $\epsilon < 0$ is independent of n then a.s. all the components of Q^p have order $O(n)$.

We prove the result analogous to the Ajtai, Komlós and Szemerédi theorem: if $\epsilon > 0$ is fixed then a.s. the largest component of Q^p has order about ηpN, where $\eta = \eta(\epsilon) > 0$ is computed explicitly. We also show that a.s. the second largest component has order $O(n^{10})$. It is very likely that there is a unique giant component even when ϵ depends on n in such a way that $\epsilon n \to \infty$ at a reasonable rate. It would be very interesting to determine whether results similar to the main results of this note could be proved for Q^p.

Acknowledgements. The first author was partially supported by the NSF under grant DMS-8806097. The second author was supported by FAPESP, University of São Paulo, and by a grant from the Louisiana Education Quality Support Fund through the Board of Regents. The third author was partially supported by a SERC grant while visiting the University of Cambridge.

References

[1] Ajtai, M., Komlós, J., Szemerédi, E., Largest random component of a k-cube, *Combinatorica* **2** (1982), 1–7.

[2] Bollobás, B., The evolution of the cube, in *Combinatorial Mathematics* (Berge, C., Bresson, D., Camion, P., Maurras, J.F. and Sterboul, F., eds), North-Holland, Amsterdam 1983, pp. 91–97.

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

[4] Bollobás, B., *Random Graphs*, Academic Press, London, 1985, *xvi* + 447pp.

[5] Bollobás, B., Complete matchings in random subgraphs of the cube, *Random Structures and Algorithms* **1** (1990), 95–104.

[6] Bollobás, B., Kohayakawa, Y., Łuczak, T., On the evolution of random Boolean functions, to appear.

[7] Bollobás, B., Leader, I., Exact face-isoperimetric inequalities, *Europ. J. Combinatorics* **11** (1990), 335–340.

[8] Bollobás, B., Leader, I., Matchings in the cube, to appear.

[9] Burtin, Yu.D., On the probability of the connectedness of a random subgraph of the n-cube (in Russian), *Problemy Pered. Inf.* **13** (1977), 90–95.

[10] Dyer, M.E., Frieze, A.M., Foulds, L.R., On the strength of connectivity of random subgraphs of the n-cube, in *Random Graphs '85*, Annals of Discrete Mathematics 33 (Karoński, M., Palka, Z., eds), North-Holland, Amsterdam 1987, pp. 17–40.

[11] Erdős, P., Rényi, A., On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* **5** (1960), 17–61.

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

[13] Erdős, P., Spencer, J.H., Evolution of the n-cube, *Comput. Math. Applics* **5** (1979), 33-39.

[14] Füredi, Z., On the connectedness of a random graph with a small number of edges, *Studia Sci. Math. Hung.* **14** (1979), 419–425.

[15] Harris, T.E., *The Theory of Branching Processes*, Springer–Verlag, Berlin, 1963, $xiv + 230$pp.

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

[17] Kolchin, V.F., *Random Mappings*, Optimization Software, New York, 1986, $xiv + 206$pp.

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

[19] Raney, G.N., Functional composition patters and power series reversion, *Trans. Amer. Math. Soc.* **94** (1960), 441–451.

[20] Toman, E., On the probability of connectedness of random subgraphs of the n-cube (in Russian), *Math. Slovaca* **30** (1980), 251–265.

[21] Weber, K., On the evolution of random graphs in the n-cube, in *Graphs, Hypergraphs and Applications* (Sachs, H., ed), Leipzig, DDR 1985, pp. 203–206.