

EXPANSION IN SUPERCRITICAL RANDOM SUBGRAPHS OF THE HYPERCUBE AND ITS CONSEQUENCES

BY JOSHUA ERDE^{1,*}, MIHYUN KANG^{1,†} AND MICHAEL KRIVELEVICH²

¹*Institute of Discrete Mathematics, Graz University of Technology, Steyrergasse 30, 8010 Graz, Austria,*
^{*}erde@math.tugraz.at; [†]kang@math.tugraz.at

²*School of Mathematical Sciences, Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 6997801, Israel,*
krivelev@tauex.tau.ac.il

It is well-known that the behaviour of a random subgraph of a d -dimensional hypercube, where we include each edge independently with probability p , undergoes a phase transition when p is around $\frac{1}{d}$. More precisely, standard arguments show that just below this value of p all components of this graph have order $O(d)$ with probability tending to one as $d \rightarrow \infty$ (whp for short), whereas Ajtai, Komlós and Szemerédi [Largest random component of a k -cube, *Combinatorica* 2 (1982), no. 1, 1–7; MR0671140] showed that just above this value, in the *supercritical regime*, whp there is a unique ‘giant’ component of order $\Theta(2^d)$. We show that whp the vertex-expansion of the giant component is inverse polynomial in d . As a consequence we obtain polynomial in d bounds on the diameter of the giant component and the mixing time of the lazy random walk on the giant component, answering questions of Bollobás, Kohayakawa and Łuczak [On the diameter and radius of random subgraphs of the cube, *Random Structures and Algorithms* 5 (1994), no. 5, 627–648; MR1300592] and of Pete [A note on percolation on \mathbb{Z}^d : isoperimetric profile via exponential cluster repulsion, *Electron. Commun. Probab.* 13 (2008), 377–392; MR2415145]. Furthermore, our results imply lower bounds on the circumference and Hadwiger number of a random subgraph of the hypercube in this regime of p which are tight up to polynomial factors in d .

1. Introduction Percolation is a mathematical process, initially studied by Broadbent and Hammersley [17] to model the flow of a fluid through a porous medium whose channels may be randomly blocked. The underlying mathematical model is simple: given a graph G , usually some sort of lattice-like graph, the *percolated subgraph* G_p is the random subgraph of G obtained by retaining each edge of G independently with probability p . For a more detailed introduction to percolation theory see, e.g., [44, 33, 15].

In this paper we are concerned with percolation on the hypercube. The d -dimensional hypercube Q^d is the graph with vertex set $V(Q^d) = \{0, 1\}^d$ and in which two vertices are adjacent if they differ in exactly one coordinate. Throughout this paper we will write $n := 2^d$ for the order of the hypercube and we note that $|E(Q^d)| = \frac{nd}{2}$. The hypercube is a ubiquitous object in graph theory and combinatorics, arising naturally in many contexts, in particular due to its interpretation as the Hasse diagram of the partial order on $[d] := \{1, 2, \dots, d\}$ given by the subset relation.

The random subgraph Q_p^d of the hypercube was first studied by Sapoženko [60] and by Burtin [6], who showed that Q_p^d has a threshold for connectivity at $p = \frac{1}{2}$; for a fixed constant $p < \frac{1}{2}$, whp¹ Q_p^d is disconnected, whereas for $p > \frac{1}{2}$, whp Q_p^d is connected. This result was

MSC2020 subject classifications: Primary 05C80, 82B43, 05C48.

Keywords and phrases: Hypercube percolation, Graph expansion.

¹Throughout the paper, all asymptotics will be considered as $d \rightarrow \infty$ and so, in particular, whp (with high probability) means with probability tending to one as $d \rightarrow \infty$.

strengthened by Erdős and Spencer [25] and by Bollobás [8], who determined the probability of connectivity of Q_p^d when p is close to $\frac{1}{2}$. Bollobás [10] also showed that $p = \frac{1}{2}$ is the threshold for the existence of a perfect matching in Q_p^d . Very recently, answering a long-standing open problem, Condon, Espuny Díaz, Girão, Kühn and Osthus [19] showed that $p = \frac{1}{2}$ is also the threshold for the existence of a Hamilton cycle in Q_p^d .

Motivated by results from the *binomial random graph model*, it was conjectured by Erdős and Spencer [25] that the component structure of Q_p^d should undergo a *phase transition* at $p = \frac{1}{d}$: it is relatively easy to see, by a coupling with a branching process, that when $p = \frac{1-\epsilon}{d}$ for $\epsilon > 0$, whp all components of Q_p^d have order $O(d)$, but they conjectured that when $p = \frac{1+\epsilon}{d}$, whp Q_p^d contains a unique ‘giant’ component $L_1(Q^d)$, whose order is linear in $n = 2^d$. This conjecture was confirmed by Ajtai, Komlós and Szemerédi [2].

THEOREM 1.1 (Ajtai, Komlós and Szemerédi [2]). *Let $\epsilon > 0$ and let $p = \frac{1+\epsilon}{d}$. Then there is a constant $\gamma > 0$ such that whp Q_p^d contains a component of order at least γn .*

These results were later extended to a wider range of p , describing more precisely the component structure of Q_p^d when $p = \frac{1+\epsilon}{d}$ with $\epsilon = o(1)$ by Bollobás, Kohayakawa and Łuczak [12], by Borgs, Chayes, van der Hofstad, Slade and Spencer [16] and by Hulshof and Nachmias [41], with the correct width of the critical window in this model being only recently identified by van der Hofstad and Nachmias [39]. McDiarmid, Scott and Withers [56] also give a description of the component structure of Q_p^d for fixed $p \in (0, \frac{1}{2})$, when p is quite far from the critical window, but still below the connectivity threshold. For a more detailed background on the phase transition in this model, see the survey of van der Hofstad and Nachmias [38].

In this paper we are interested in the typical structural properties of the giant component L_1 of Q_p^d in the *supercritical regime*, where $p = \frac{1+\epsilon}{d}$ for some fixed $\epsilon > 0$. Our main result, from which we will be able to deduce a lot of structural information about L_1 , concerns its *expansion* properties, in particular vertex-expansion. Informally, a graph has good vertex-expansion if all sufficiently small vertex sets have a large vertex boundary, expressing a kind of discrete isoperimetric inequality. The notion of graph expansion seems to be quite a fundamental one to the study of graphs, demonstrating a deep link between the geometric and structural properties of a graph, its algebraic spectrum and also the mixing time of the random walk on the graph. For these reasons and more, graph expansion has turned out to have fundamental importance in many diverse areas of discrete mathematics and computer science. For a comprehensive introduction to expander graphs, see the survey of Hoory, Linial and Wigderson [40]. In particular, notions of expansion have turned out to be a powerful tool in the study of random structures. See, for example, the survey paper of Krivelevich [47].

In order to motivate our results and methods, let us discuss briefly what is known in a simpler model of percolation, the binomial random graph. The binomial random graph $G(d+1, p)$, introduced by Gilbert [32], is a percolated subgraph of the complete graph K_{d+1} , where we retain each edge with probability p . As a particularly simple model of percolation, where the underlying graph G lacks the geometric structure of \mathbb{Z}^d or Q^d , the binomial random graph has been extensively studied. A particularly striking feature of this model is the phase transition that it undergoes at $p = \frac{1}{d}$, exhibiting vastly different behaviour when $p = \frac{1-\epsilon}{d}$ to when $p = \frac{1+\epsilon}{d}$ (where ϵ is a positive constant).

More precisely, it follows from results of Erdős and Rényi [24] that when $p = \frac{1-\epsilon}{d}$ for a fixed $\epsilon > 0$, whp every component of $G(d+1, p)$ has order at most $O(\log d)$ and is either a tree or unicyclic, whereas when $p = \frac{1+\epsilon}{d}$, whp $G(d+1, p)$ contains a unique ‘giant component’ $L_1(G(d+1, p))$ of order $\Omega(d)$, whose structure is quite complex. We note that this

is only a very broad picture of the phase transition in this model, and much more precise results are known, in particular extending these results into the *weakly supercritical regime* where $\epsilon = o(1)$ and $\epsilon^3 d \rightarrow \infty$, see, for example, the works of Bollobás [9] and Łuczak [55]. However this is not the focus of our paper.

Much subsequent work has focused on the structural properties of the giant component $L_1 = L_1(G(d+1, p))$ in the supercritical regime. For example, a well-known result of Ajtai, Komlós and Szemerédi [1] shows that in this regime whp L_1 contains a path of length $\Omega(d)$, from which it is easy to deduce that the *circumference*, the length of the largest cycle, of L_1 is of order $\Omega(d)$. Fountoulakis, Kühn and Osthus [27] showed that whp the *Hadwiger number*, the size of the largest complete minor, of L_1 is of order $\Omega(\sqrt{d})$. Chung and Lu [18] showed that whp L_1 has diameter $O(\log d)$ and this result was later strengthened, in particular determining the correct leading constant, by work of Fernholz and Ramachandran [26] and of Riordan and Wormald [59]. Benjamini, Kozma and Wormald [4] and Fountoulakis and Reed [29] showed that whp the mixing time of the lazy random walk on L_1 is $O((\log d)^2)$ and Berestycki, Lubetzky, Peres and Sly [5] showed that if we start the lazy random walk from a uniformly chosen vertex of L_1 , then whp the mixing time is $O(\log d)$ (for definitions related to mixing time see Section 4.1). We note that this is just a small subset of the results known about the supercritical random graph, chosen judiciously for comparison to our results later. For more general background on the theory of random graphs, see [11, 42, 30].

A lot of structural information about the giant component of $G(d+1, p)$, including many of the results mentioned above, can be deduced as consequences of its expansion properties. Indeed, it is known that the expansion properties of a graph can be linked to various structural graph properties, for example its diameter and circumference, and furthermore there are well-known links between expansion and mixing times of Markov chains. See Sections 2 and 4.1 for more precise statements.

Given a graph G and a subset $S \subseteq V(G)$ we write $N_G(S)$ for the *external neighbourhood* of S in G , that is, the set of vertices in $V(G) \setminus S$ which have a neighbour in S and we write $S^c := V(G) \setminus S$ and $e_G(S, S^c)$ for the number of edges between S and S^c in G .

DEFINITION 1. We say a graph G is an α -*expander* if $|N_G(S)| \geq \alpha|S|$ for every $S \subseteq V(G)$ such that $|S| \leq \frac{|V(G)|}{2}$, where α is the *expansion ratio*.

Similarly we say a graph G is an α -*edge-expander* if $e_G(S, S^c) \geq \alpha|S|$ for all S such that $\sum_{v \in S} d_G(v) \leq |E(G)|$.

It can easily be seen that $G(d+1, p)$ is whp not an expander in the supercritical regime, since the graph is likely not connected. In fact, standard results imply that even the *2-core*, the largest subgraph of minimum degree at least two, of the giant component of $G(d+1, p)$ does not have constant vertex- or edge-expansion, since it typically contains logarithmically long *bare paths*, paths in which each internal vertex has degree two. However, it was shown by Benjamini, Kozma and Wormald [4] that in the supercritical regime whp the giant component of $G(d+1, p)$ is a *decorated expander*, which roughly means that it has a linear sized subgraph which is an α -edge-expander for some constant $\alpha > 0$, and the deletion of this subgraph splits the giant component into logarithmically small pieces. It is not hard to show that in the supercritical regime the existence of a linear sized subgraph which is an α -edge-expander implies the existence of one which is an α' -expander for some $\alpha' > 0$, the existence of which was also shown by Krivelevich [46] using general considerations of expansion in locally sparse graphs. We also note that the work of Ding, Lubetzky and Peres [20] gives a particularly simple model contiguous to the giant component in the supercritical regime, which implies that whp the *kernel* of the giant component, the graph obtained by contracting all bare paths in the 2-core, is an α -expander for some fixed $\alpha > 0$, and from which it is

possible to determine the likely expansion properties of the giant component and its 2-core. Then, using these connections between the structural properties of a graph and its expansion mentioned previously, many properties of the giant component of $G(d+1, p)$ can be deduced as consequences of the results of Benjamini, Kozma and Wormald [4], of Krivelevich [46] and of Ding, Kim, Lubetzky and Peres [20].

More recently, random subgraphs G_p of an arbitrary graph G of large minimum degree $\delta(G) \geq d$ have been studied. It has been observed that some of the complex behaviour which occurs whp in $G(d+1, p)$ once we pass the critical point of $p = \frac{1}{d}$ also occurs whp in G_p in the same regime of p . For example, when $p = \frac{1+\epsilon}{d}$ for $\epsilon > 0$, it has been shown that whp G_p contains a path or cycle of length at least linear in d , see [51, 50, 22]. Furthermore, for this range of p it has been shown by Frieze and Krivelevich [31] that whp G_p is non-planar and by Erde, Kang and Krivelevich [23] that in fact, whp G_p has Hadwiger number $\tilde{\Omega}(\sqrt{d})$.²

Whilst the model $G(d+1, p)$ shows that these results are optimal when G can be an arbitrary graph, for specific graphs, and in particular for $G = Q^d$, they may be far from the truth. Indeed, it is plausible that circumference and Hadwiger number of Q_p^d could be exponentially large in d in the supercritical regime. Furthermore, since the host graph G can be chosen arbitrarily, we cannot hope to prove much about global properties of Q_p^d , such as the diameter or mixing time, by considering the far more general model of arbitrary random subgraphs G_p .

The main aim of this paper is to show that whp the giant component of Q_p^d in the supercritical regime has good expansion properties. Given constants $\alpha, \beta > 0$ and a statement A , we will write ‘Let $\alpha \ll \beta$. Then A holds’ to indicate that there is some fixed, implicit function f such that A holds for all $\alpha \leq f(\beta)$.

THEOREM 1.2. *Let $\epsilon > 0$, let $p = \frac{1+\epsilon}{d}$ and let L_1 be the largest component of Q_p^d . Then there exists a constant $\beta > 0$ such that whp L_1 is a βd^{-5} -expander.*

Furthermore, we show that whp the giant component of Q_p^d contains an almost spanning subgraph with much better expansion.

THEOREM 1.3. *Let $0 < \alpha \ll \epsilon$, let $p = \frac{1+\epsilon}{d}$ and let L_1 be the largest component of Q_p^d . Then there exist a constant $\beta > 0$ and a subgraph H of L_1 of size at least $(1 - \alpha)|V(L_1)|$ such that whp H is a $\beta d^{-2}(\log d)^{-1}$ -expander.*

We will be able to deduce certain structural consequences from the vertex-expansion of the giant component which are almost optimal, up to polynomial factors in d .

The first consequence will be a bound on the mixing time of the lazy random walk on the giant component L_1 of Q_p^d . Answering a well-known question (see, e.g., Pete [57] and van der Hofstad and Nachmias [39]), we show that whp the mixing time of this random walk is polynomial in d .

THEOREM 1.4. *Let $\epsilon > 0$, let $p = \frac{1+\epsilon}{d}$ and let L_1 be the largest component in Q_p^d . Then whp the mixing time of the lazy random walk on L_1 is $O(d^{11})$.*

Next, we consider the diameter of the giant component in Q_p^d . Previously, Bollobás, Kohayakawa and Łuczak [13] considered the diameter of a random Q^d -process. Although their results mainly concern the structure of Q_p^d close to the connectivity threshold, they asked

²The notation $\tilde{\Omega}(\cdot)$ here is hiding a polylogarithmic factor in d .

whether the diameter of any component in a typical Q_p^d -process is ever superpolynomial in d .

In fact, since then it has been shown that the diameter of components in the regime close to the critical window can grow even exponentially large in d . Hulshof and Nachmias [41] show that in the weakly subcritical regime, when $n^{-\frac{1}{3}} \ll \epsilon = o(1)$ and $p = \frac{1-\epsilon}{d}$, whp the maximal diameter of a component in Q_p^d is $(1 + o(1))\epsilon^{-1} \log(\epsilon^3 n)$, although they mention that this is not achieved by the largest component, which they conjecture to have diameter $\Theta\left(\epsilon^{-1} \sqrt{\log(\epsilon^3 n)}\right)$. Heydenreich and van der Hofstad [37] mention that their methods also show that the diameter of the largest component of the critical percolated hypercube is $\Theta_p\left(n^{\frac{1}{3}}\right)$ and it is also stated by van der Hofstad and Nachmias [39] that in the weakly supercritical regime, when $n^{-\frac{1}{3}} \ll \epsilon = o(1)$ and $p = \frac{1+\epsilon}{d}$, whp the diameter of the giant component in Q_p^d is $(1 + o(1))\epsilon^{-1} \log(\epsilon^3 n)$.

Hence, the question of Bollobás, Kohayakawa and Łuczak [13] perhaps only makes sense once we are quite far from the critical window. In this range, we give a polynomial bound on the likely diameter of the giant component.

THEOREM 1.5. *Let $\epsilon > 0$ and let $p = \frac{1+\epsilon}{d}$. Then whp the largest component L_1 of Q_p^d has diameter $O(d^3)$.*

Finally, we also consider the circumference and the Hadwiger number of Q_p^d .

THEOREM 1.6. *Let $\epsilon > 0$ and let $p = \frac{1+\epsilon}{d}$. Then whp the circumference of Q_p^d is*

$$\Omega(nd^{-2}(\log d)^{-1}).$$

THEOREM 1.7. *Let $\epsilon > 0$ and let $p = \frac{1+\epsilon}{d}$. Then whp the Hadwiger number of Q_p^d is*

$$\Omega(\sqrt{nd}^{-2}(\log d)^{-1}).$$

We note that, since whp $|E(Q_p^d)| = O(n)$ when $p = O\left(\frac{1}{d}\right)$ and the Hadwiger number $h(G)$ of a graph G satisfies $\binom{h(G)}{2} \leq |E(G)|$, it follows that both of these results are optimal up to polynomial terms in d . We note further that a bound on the likely circumference of Q_p^d was also obtained in concurrent work by Haslegrave, Hu, Kim, Liu, Luan and Wang [36], who showed the likely existence of a cycle of length $\Omega(nd^{-32})$ using different methods.

The paper is structured as follows. In Section 2 we collect some lemmas which will be useful in the rest of the paper. Then in Section 3 we prove our main results (Theorems 1.2 and 1.3) on the likely expansion of the giant component of Q_p^d . In Section 4 we use our main result to prove Theorems 1.4–1.7, and then finally in Section 5 we mention some open problems and avenues for further research.

2. Preliminaries For real numbers x, y, z we will write $x = y \pm z$ to mean that

$$y - z \leq x \leq y + z.$$

If G is a graph and X is either a subgraph of G , or a subset of $V(G)$, then we will write $G - X$ for the induced subgraph of G on $V(G) \setminus V(X)$ or $V(G) \setminus X$, respectively. Given a $k \in \mathbb{N}$ and a subset $S \subseteq V(G)$, we will write $N_G^k(S)$ for the k th external neighbourhood of S in G , that is, the set of vertices in $S^c := V(G) \setminus S$ which are at distance at most k from S in G . When $k = 1$, or when the graph is clear from context, we will omit the superscript or

subscript, respectively. Given (not necessarily disjoint) subsets $A, B \subseteq V(G)$ we will write $e_G(A, B)$ for the number of edges in G with one endpoint in A and the other in B and we write $e_G(S) = e_G(S, S)$. We say a subset $S \subseteq V(G)$ is *connected (in G)* if $G[S]$ is connected.

Throughout the paper, unless the base is explicitly mentioned, all logarithms will be the natural logarithm. We will also omit floor and ceiling signs for ease of presentation.

We will want to use a more explicit form of Theorem 1.1. It is stated in [2] that a careful treatment of their proof gives the following result, which also appears explicitly in the work of Bollobás, Kohayakawa and Łuczak [12, Theorem 32].

THEOREM 2.1 ([2]). *Let $c, \delta > 0$, let $p = \frac{1+\delta}{d}$ and let $\gamma := \gamma(\delta)$ be the survival probability of the $Po(1 + \delta)$ branching process. Then whp the largest component L_1 of Q_p^d is such that $|V(L_1)| = (\gamma \pm c)n$.*

The following simple lemma, which is a slight adaptation of a result in [49], allows us to decompose a tree into roughly equal sized parts.

LEMMA 2.2. *Let T be a tree such that $\Delta(T) \leq C_1$, all but r vertices of T have degree at most $C_2 \leq C_1$ and $|V(T)| \geq \ell$, for some $C_1, C_2, \ell, r > 0$. Then there exist disjoint vertex sets $A_1, \dots, A_m \subseteq V(T)$ such that*

- $V(T) = \bigcup_{i=1}^m A_i$;
- $T[A_i]$ is connected for each $1 \leq i \leq m$;
- $T[A_i]$ has diameter at most 2ℓ ;
- $\ell \leq |A_i| \leq C_1 \ell$ for each $1 \leq i \leq r$; and
- $\ell \leq |A_i| \leq C_2 \ell$ for each $r < i \leq m$.

PROOF. We choose an arbitrary root w for T . For a vertex v in a rooted tree S , let us write S_v for the subtree of S rooted at v .

We construct the vertex sets A_i inductively. Let us start by setting $T(0) = T$. Given a tree $T(i)$ rooted at w such that $|V(T(i))| \geq \ell$, let v_i be a vertex of maximal distance from w such that $|V(T(i)_{v_i})| \geq \ell$. We take $A_{i+1} = V(T(i)_{v_i})$ and let $T(i+1) = T(i) - T(i)_{v_i}$. We stop when $|V(T(i))| < \ell$, and in that case we add $V(T(i))$ to the final A_i . Finally, let us re-order the sets A_i so that they are non-increasing in size.

We claim that the sets A_1, A_2, \dots, A_m satisfy the conclusion of the lemma. Indeed, the first two properties are clear by construction. Note that each A_i is the union of the vertices of $T(i)_x$ over all children x of v_i , together with a connected set of size at most ℓ which contains v_i , and by our choice of v_i , $|T(i)_x| < \ell$ for every child x of v_i .

In particular, it follows that every vertex in A_i is at distance at most ℓ from v_i , and so $T[A_i]$ has diameter at most 2ℓ , and the third property holds. Furthermore, if $v_i \neq w$, then v_i has $d(v_i) - 1$ children and so, it follows that $|A_{i+1}| \leq (d(v_i) - 1)(\ell - 1) + \ell \leq d(v_i)\ell$. Similarly, if $v_i = w$, then we note that $A_{i+1} = V(T(i))$ and so $|A_{i+1}| \leq d(w)(\ell - 1) + 1 \leq d(w)\ell$. Therefore, since all but r vertices of T have degree at most C_2 , the fourth and fifth properties also hold. \square

We will also need the following bound on the number of subtrees of a graph.

LEMMA 2.3 ([7, Lemma 2]). *Let G be a graph with maximum degree Δ , let $v \in V(G)$ and let $t(v, k)$ be the number of rooted trees in G which have root v and k vertices. Then*

$$t(v, k) \leq \frac{k^{k-2} \Delta^{k-1}}{(k-1)!} \leq (e\Delta)^{k-1}.$$

The following theorem allows us to deduce the existence of a long cycle from vertex-expansion properties of a graph. For wider context on properties of expanding graphs, see the survey of Krivelevich [47].

THEOREM 2.4 ([48, Theorem 1]). *Let $k \geq 1, t \geq 2$ be integers. Let G be a graph on more than k vertices satisfying*

$$|N(W)| \geq t, \quad \text{for every } W \subseteq V(G) \text{ with } \frac{k}{2} \leq |W| \leq k.$$

Then G contains a cycle of length at least $t + 1$.

Furthermore, the next theorem allows us to deduce the existence of a large complete minor in a graph without any small separators. It is easy to see that graphs with good vertex-expansion do not contain any small separators, and in fact it is known (see [47, Section 5]) that the converse is true, in the sense that graphs without small separators must contain large induced subgraphs with good vertex-expansion.

THEOREM 2.5 ([43, Theorem 1.2]). *Let G be a graph with N vertices and with no K_t -minor. Then $V(G)$ contains a subset S of size $O(t\sqrt{N})$ such that each connected component of $G - S$ has at most $\frac{2}{3}N$ vertices.*

Given a discrete random variable X taking values in \mathcal{X} the *entropy* of X is given by

$$H(X) := \sum_{x \in \mathcal{X}} -p(x) \log_2(p(x)),$$

where $p(x) = \mathbb{P}(X = x)$. We will need only two basic facts about the entropy function.

LEMMA 2.6.

- (i) $H(X) \leq \log_2(|\mathcal{X}|)$ with equality iff X is uniformly distributed,
- (ii) $H(X_1, X_2, \dots, X_d) \leq \sum_{i=1}^d H(X_i)$,

where the joint entropy $H(X_1, X_2, \dots, X_d)$ is the entropy of the random vector (X_1, X_2, \dots, X_d) .

For proofs of these facts, and more background on discrete entropy, see, e.g., [3, Chapter 15].

We will use the following Chernoff type bounds on the tail probabilities of the binomial distribution, see, e.g., [3, Appendix A].

LEMMA 2.7. *Let $N \in \mathbb{N}$, let $p \in [0, 1]$ and let $X \sim \text{Bin}(N, p)$.*

- (i) *For every positive a with $a \leq \frac{Np}{2}$,*

$$\mathbb{P}(|X - Np| > a) < 2 \exp\left(-\frac{a^2}{4Np}\right).$$

- (ii) *For every positive b ,*

$$\mathbb{P}(X > bNp) \leq \left(\frac{e}{b}\right)^{bNp}.$$

In particular, the following two simple consequences of Lemma 2.7 in our setting will be useful. The first bounds the number of high degree vertices in Q_p^d for small p .

LEMMA 2.8. *Let $c > 0$ be a constant and let $p = \frac{c}{d}$. Then whp Q_p^d contains at most nd^{-4} vertices of degree at least $\log d$.*

PROOF. For any fixed vertex $v \in V(Q^d)$, the degree of v in Q_p^d is distributed as $\text{Bin}(d, p)$, and so by Lemma 2.7 (ii) we have that

$$\mathbb{P}\left(d_{Q_p^d}(v) \geq \log d\right) \leq \left(\frac{ec}{\log d}\right)^{\log d} \leq d^{-\frac{\log \log d}{2}}.$$

It follows that the expected number of vertices in Q_p^d with degree at least $\log d$ is at most $nd^{-\frac{\log \log d}{2}}$. Hence, by Markov's inequality, whp there are at most nd^{-4} vertices with degree at least $\log d$. \square

We note that the above argument is suboptimal and with a little more care, the bound on the degree of the exceptional vertices could be improved from $\log d$ to $\frac{C \log d}{\log \log d}$ for some suitably large constant C . However, for ease of presentation we have not attempted to optimise any logarithmic factors in our proofs.

The second consequence of Lemma 2.7 allows us to find large matchings in random subsets of edges in Q_p^d .

LEMMA 2.9. *Let $\delta > 0$ be a constant, let $p = \frac{\delta}{d}$ and let $F \subseteq E(Q^d)$ be such that $|F| \geq t$. Then there exists a constant $c = c(\delta) > 0$ such that F_p contains a matching of size at least ctd^{-1} with probability at least $1 - \exp(-ctd^{-1})$.*

PROOF. We note that we can assume that $|F| = t$. Let us consider the number of maximal matchings in F_p of size ℓ .

There are clearly at most $\binom{|F|}{\ell}$ potential maximal matchings of size ℓ , and given a matching M of size ℓ in F , in order for it to be a maximal matching in F_p its edges have to appear in F_p , which happens with probability p^ℓ , and also there can be no other edges in F_p which are disjoint from M . Since there are at most $2\ell d$ edges which share a vertex with edges in M , there is a set of $|F| - 2\ell d$ edges which do not appear in F_p , which happens with probability at most $(1-p)^{|F|-2\ell d}$. Hence, by the union bound, the probability that F_p contains a maximal matching of size ℓ is at most

$$\binom{|F|}{\ell} \left(\frac{\delta}{d}\right)^\ell \left(1 - \frac{\delta}{d}\right)^{|F|-2\ell d}.$$

In particular, as long as $c \ll \delta$, we can bound the probability q that F_p contains a maximal matching of size $\ell \leq ctd^{-1}$ from above by

$$\begin{aligned} q &\leq \sum_{\ell=1}^{ctd^{-1}} \binom{t}{\ell} \left(\frac{\delta}{d}\right)^\ell \left(1 - \frac{\delta}{d}\right)^{t-2\ell d} \leq \sum_{\ell=1}^{ctd^{-1}} \left(\frac{et}{\ell}\right)^\ell \left(\frac{\delta}{d}\right)^\ell \left(1 - \frac{\delta}{d}\right)^{\frac{t}{2}} \\ &\leq \exp\left(-\frac{\delta t}{2d}\right) \sum_{\ell=1}^{ctd^{-1}} \left(\frac{e\delta t}{\ell d}\right)^\ell, \end{aligned}$$

since in this range of ℓ we have that $t - 2\ell d \geq \frac{t}{2}$. However, since $c \ll \delta$, it can be seen that the ratio of the consecutive terms $\left(\frac{e\delta t}{\ell d}\right)^\ell$ is at least 2, and so the sum is dominated by the final term. Hence, since $c \ll \delta$, we can bound

$$q \leq 2 \exp\left(-\frac{\delta t}{2d}\right) \left(\frac{e\delta}{c}\right)^{ctd^{-1}} \leq 2 \exp\left(\frac{ct}{d} \log\left(\frac{e\delta}{c}\right) - \frac{\delta t}{2d}\right) \leq \exp(-ctd^{-1}). \quad \square$$

We note that the conclusion of Lemma 2.9 is optimal up to a constant factor. Indeed, whp, for example by Lemma 2.7, there will only be $O(td^{-1})$ edges in F_p .

We will use the following well-known result on edge-isoperimetry in the hypercube, originally due to Harper [34], see also Lindsey [54], Bernstein [6], and Hart [35].

THEOREM 2.10 ([34, 54, 6, 35]). *For any $A \subseteq V(Q^d)$ with $|A| \leq 2^{d-1}$,*

$$e(A, A^c) \geq |A|(d - \log_2 |A|).$$

Finally we will use the following lemma which bounds the likely number of edges spanned by connected subsets in Q_p^d .

LEMMA 2.11. *Let $\delta > 0$ be a constant and let $p = \frac{\delta}{d}$. Then there exists a constant $C := C(\delta)$ such that whp every subset $S \subseteq V(Q^d)$ such that $|S| \geq d$ and $Q_p^d[S]$ is connected satisfies $e_{Q_p^d}(S) \leq C|S|$.*

PROOF. Note that, if $S \subseteq V(Q^d)$ has size $|S| =: k$, then, since Q^d is d -regular, it follows from Theorem 2.10 that $e_{Q^d}(S) \leq \frac{k \log_2 k}{2}$.

Let us bound from above the probability that there exists a subset of $V(Q_p^d)$ of size k which is connected and spans at least Ck edges. Such a subset must span a tree, which we can specify by choosing a vertex and one of the at most $(ed)^{k-1}$ trees of size k containing that vertex, using Lemma 2.3 to bound this quantity. This tree is contained in Q_p^d with probability p^{k-1} .

If we let S be the vertex set of this tree, then by the above comment $e_{Q^d}(S) \leq \frac{k \log_2 k}{2}$. In order for $e_{Q_p^d}(S) \geq Ck$ there must be a set of $Ck - (k-1) \geq (C-1)k$ further edges of Q^d in S which appear in Q_p^d , which happens with probability at most $p^{(C-1)k}$.

Hence, writing $C' = C - 1$ for ease of **presentation**, by the union bound, the probability that such a set of size $k \geq d$ exists is at most

$$\begin{aligned} \sum_{k=d}^n n(ed)^{k-1} p^{k-1} \left(\frac{k \log_2 k}{C'k} \right) p^{C'k} &\leq \sum_{k=d}^n 2^k (ed)^{k-1} \left(\frac{e\delta \log_2 k}{2C'd} \right)^{C'k} \\ &\leq \sum_{k=d}^n \left(\frac{2^{\frac{1}{C'}} (ed)^{1+\frac{k-1}{C'}}}{2C'} \right)^{C'k} = o(1), \end{aligned}$$

as long as $C' = C - 1$ is sufficiently large in terms of δ . □

3. Expansion in the giant component We begin by establishing some likely properties of the giant component of Q_p^d which will be useful in our proof.

The first says that whp the second largest component of Q_p^d in the supercritical regime is only of linear size in d . We note that it is mentioned already in [2] that such a result can be shown using methods of Komlós, Sulyok and Szemerédi from [45], however a proof can be found in [12, Theorem 31].

LEMMA 3.1 ([12, Theorem 31]). *Let $0 < \delta < 1$ and let $p = \frac{1+\delta}{d}$. Then there exists a constant $K_1 := K_1(\delta) > 0$ such that the second largest component in Q_p^d has order at most $K_1 d$.*

We will also use the following consequence of Lemma 3.1.

LEMMA 3.2. *Let $0 < \delta_1 < 1$ and $\delta_2 \ll \delta_1$, let $q_1 = \frac{1+\delta_1}{d}$ and $q_2 = \frac{\delta_2}{d}$ and let L'_1 and L_1 be the largest components in $Q_1 := Q_{q_1}^d$ and $Q_2 := Q_1 \cup Q_{q_2}^d$, respectively. Given a vertex $v \in V(L'_1)$, let C_v be the set of vertices which are contained in a component of $L_1 - L'_1$ which is adjacent to v in Q_2 . Then there exists a constant $K_2 := K_2(\delta_1) > 0$ such that whp $|C_v| \leq K_2 d$ for every $v \in V(L'_1)$.*

PROOF. We first note that by Lemma 3.1, there exists a constant $K_1 := K_1(\delta_1)$ such that whp every component of Q_1 except L'_1 has order at most $K_1 d$. Let $K_2 \gg \delta_1^{-1}$.

Suppose that there is some vertex $v \in V(L'_1)$ such that $|C_v| \geq K_2 d$. We note that $C_v \cup \{v\}$ is connected in Q_2 , and C_v is the disjoint union of some set $\{C_1, \dots, C_r\}$ where each C_i is the vertex set of some component of Q_1 , each of which has size at most $K_1 d$. It follows that there must be some subset $\hat{C} \subseteq C_v$ such that $\hat{C} \cup \{v\}$ is connected in Q_2 , $K_2 d \leq |\hat{C}| \leq (K_1 + K_2)d$ and \hat{C} is the union of some subset of $\{C_1, C_2, \dots, C_r\}$.

In particular, there is some spanning tree T of $\hat{C} \cup \{v\}$, all of whose edges are present in Q_2 , such that no edge in the edge-boundary of $V(T) \setminus \{v\}$ is present in Q_1 .

Let us bound the probability that such a tree of size k exists in Q_2 for each

$$K_2 d + 1 \leq k \leq (K_1 + K_2)d + 1.$$

We can fix such a tree T by choosing a root vertex v and choosing one of the at most $(ed)^{k-1}$ possible rooted trees of size k with root v in Q^d , where we have bounded the number of possible trees by Lemma 2.3.

Now, T has $k - 1$ edges and furthermore, since $|V(T) \setminus \{v\}| = k - 1$, by Theorem 2.10 there are at least $(k - 1)(d - \log_2(k - 1))$ edges in the edge-boundary of $V(T) \setminus \{v\}$. Note that each edge is in Q_2 with probability at most $q_1 + q_2$ and each edge is not in Q_1 with probability $1 - q_1$, and that, whilst these two events are not necessarily independent, they are clearly negatively correlated.

It follows by the union bound that the probability that such a tree of size k exists in Q_2 is at most

$$n(ed)^{k-1} (q_1 + q_2)^{k-1} (1 - q_1)^{(k-1)(d - \log_2(k-1))}.$$

In particular, the probability that such a tree exists for $k \in I := [K_2 d + 1, (K_1 + K_2)d + 1]$ is at most

$$\begin{aligned} & n \sum_{k \in I} (e(1 + \delta_1 + \delta_2))^{k-1} \left(1 - \frac{1 + \delta_1}{d}\right)^{(1-o(1))(k-1)d} \\ & \leq n \sum_{k \in I} \exp((k-1)(1 + \log(1 + \delta_1 + \delta_2) - (1 - o(1))(1 + \delta_1))) \\ & \leq n \sum_{k \in I} \exp\left(-\frac{(k-1)\delta_1^2}{5}\right) = o(1), \end{aligned}$$

where we used that $\log(1 + \delta_1 + \delta_2) \leq \delta_1 - \frac{\delta_1^2}{4}$ for all $\delta_1 \in (0, 1)$, since $\delta_2 \ll \delta_1$ and that $K_2 d \delta_1^2 \gg d$, since $K_2 \gg \delta_1^{-1}$. \square

The next lemma says that whp the giant component of Q_p^d is in some sense ‘dense’ in the hypercube Q^d .

LEMMA 3.3. *Let $\delta > 0$ and let $p = \frac{1+\delta}{d}$. Then there exists a constant $c > 0$ such that whp every vertex in Q^d is at distance at most two from at least cd^2 vertices in the largest component L_1 of Q_p^d .*

PROOF. Let us choose some constant $c' \ll \delta$. Fix an arbitrary vertex $v \in V(Q^d)$, which without loss of generality we may assume to be the origin. Let $k = c'd$ and let us write $d' = d - k = (1 - c')d$. We define $s = \binom{k}{2}$ pairwise disjoint subcubes of dimension d' at distance at most two from v given by fixing some pair of 1s in the first k coordinates and varying the last d' coordinates. Let these cubes be $Q(1), \dots, Q(s)$ and let v_i be the vertex in $Q(i)$ at distance two from v .

Now, since $c' \ll \delta$, p is still supercritical in $Q^{d'}$ and so, by Theorem 2.1 and the fact that Q^d is transitive, there is some constant $\alpha > 0$ such that the probability that v_i is contained in the largest component of $Q(i)_p$ is at least α , and these events are independent for different i . Hence, by Lemma 2.7, with probability at least $1 - \exp(-c's)$ at least $\frac{\alpha s}{2}$ of the v_i are contained in the largest component of $Q(i)_p$.

Furthermore, again by Theorem 2.1 and Lemma 2.7, whp each $Q(i)_p$ contains a component whose order is $\Omega(2^{d'})$ and again these events are independent for different i , and hence with probability at least $1 - \exp(-c's)$ at least $(1 - \frac{\alpha}{4})s$ of the $Q(i)_p$ contain a component whose order is $\Omega(2^{d'})$.

It follows that with probability at least $1 - 2\exp(-c's)$ at least $\frac{\alpha s}{4} := cd^2$ of the v_i are contained in a component in $Q(i)_p$ whose order is $\Omega(2^{d'})$. Hence v is within distance two of at least cd^2 vertices lying in components in Q_p^d whose order is $\Omega(2^{d'})$ with probability at least $1 - 2\exp(-c's) = 1 - o(n^{-1})$.

Then, by the union bound, whp every vertex in Q^d is within distance two of at least cd^2 vertices lying in components in Q_p^d whose order is $\Omega(2^{d'})$. However, by Lemma 3.1, whp there is a unique component L_1 in Q_p^d whose order is superlinear in d , and so whp every vertex in Q^d is within distance two of at least cd^2 vertices in L_1 . \square

With these results in hand, let us briefly sketch the strategy to prove Theorem 1.2 about the vertex-expansion properties of the giant component of Q_p^d , where $p = \frac{1+\epsilon}{d}$. We will use a sprinkling argument, viewing Q_p^d as the union of two independent random subgraphs $Q_{q_1}^d$ and $Q_{q_2}^d$ where we have chosen $q_1 = \frac{1+\delta_1}{d}$ and $q_2 = \frac{\delta_2}{d}$ such that $(1 - q_1)(1 - q_2) = 1 - p$ and $\delta_2 \ll \delta_1$. Note that, in particular, δ_1 is approximately ϵ , and so q_1 still lies in the supercritical regime. Let us denote by L'_1 and L_1 the largest components in $Q_1 := Q_{q_1}^d$ and $Q_2 := Q_1 \cup Q_{q_2}^d$, respectively, noting that $Q_2 \sim Q_p^d$.

Given a partition of $V(L'_1)$ into two disjoint subsets A, B such that $|A|, |B| \geq t$, it is relatively easy to show that whp there is a large family of vertex-disjoint A - B -paths of length at most 5 in $Q_{q_2}^d$. Indeed, by Lemma 3.3 every vertex in Q^d is within distance two of L'_1 and so we can extend A, B to a partition of $V(Q^d)$ into two pieces $A' \supseteq A$ and $B' \supseteq B$ such that every vertex in A' is within distance two of A and every vertex in B' is within distance two of B . By Theorem 2.10 there are many edges in Q^d between A' and B' and each such edge can be extended to an A - B -path in Q^d of length at most 5. Very naively, we can thin this family of paths out to a vertex-disjoint family using the fact that $\Delta(Q^d) = d$ whilst retaining an $\Omega(d^{-6})$ proportion of them. Furthermore, after sprinkling we expect about an $\Omega(d^{-5})$ proportion of these paths to be contained in $Q_{q_2}^d$. Then, as long as t is large enough, the Chernoff bound will imply that whp there is a large vertex-disjoint family of A - B -paths of length at most 5 in $Q_{q_2}^d$.

In fact, our actual argument will be a bit more precise, to enable us to find a larger family of paths. However, the probability of failure in these arguments will not be small enough to deduce from a union bound that this holds for *all* such partitions of $V(L'_1)$.

Instead, we can use Lemma 2.2, with ℓ being some small power of d , to split L'_1 into a collection \mathcal{C} of connected pieces each having polynomial size in d . If these pieces are large enough, then there will be sufficiently few partitions of \mathcal{C} into two pieces that the probability bound from the argument above will be effective, and we can deduce that whp whenever we partition \mathcal{C} into two parts there will be a large family of vertex-disjoint paths between them in $Q_{q_2}^d$.

Furthermore, by Lemma 3.2 we may assume that for any vertex $v \in V(L'_1)$ there are only a small number of vertices contained in the components of $R := L_1 - L'_1$ which are adjacent to v in Q_2 .

Suppose then that S is some subset of $V(L_1)$. We split S into three pieces:

- S_1 is the set of vertices which lie in components of R ;
- S_2 is the set of vertices which are contained in pieces $C \in \mathcal{C}$ such that $C \cap S \neq \emptyset$ and $C \setminus S \neq \emptyset$;
- S_3 is the set of vertices which are contained in pieces C such that $C \subseteq S$.

If S_1 is large, then since each vertex in L'_1 is only adjacent in Q_2 to components in R with a small total volume, we can greedily choose a large disjoint family $\{C(x) : x \in X\}$ of components of R which all meet S , each of which is adjacent in Q_2 to a unique vertex $x \in L'_1$.

For each $x \in X$, either $x \in S$, or there is some vertex in the neighbourhood of S in $C(x) \cup \{x\}$. In particular, either S has a large neighbourhood, or $S \cap V(L'_1) = S_2 \cup S_3$ is large.

Similarly, if S_2 is large then, since each piece in \mathcal{C} is small, S_2 contains vertices in many pieces of \mathcal{C} , and for each such piece C we have that $C \setminus S \neq \emptyset$. However, since each piece $C \in \mathcal{C}$ is connected in Q_1 , each piece such that $C \cap S_2 \neq \emptyset$ contains some vertex in the neighbourhood of S , and so the neighbourhood of S is large.

Hence, we may assume that S_3 is large and S_2 is small. In this case, we look at the partition of \mathcal{C} given by pitting the pieces contained in S_3 against the rest. By the above argument whp there is a large family of vertex-disjoint paths between these two partition classes and, since S_2 is small, not many of these meet S_2 . Every path which does not meet S_2 starts in $S_3 \subseteq S$ and ends in S^c , and so contains some vertex in the neighbourhood of S . Hence, in every case we can conclude that S has a large neighbourhood.

Let us begin by proving the following result, which guarantees the likely existence of a large family of vertex-disjoint paths between the two parts of any fixed non-trivial partition of $V(L'_1)$. Note that, since there are subsets A of $V(Q^d)$ whose edge-boundary in Q^d is as small as $|A|(d - \log_2 |A|)$, for example subcubes, we cannot hope to guarantee the likely existence of a family of paths from A to A^c in the random subgraph Q_p^d with $p = O(\frac{1}{d})$ of size larger than $O\left(|A| \left(1 - \frac{\log_2 |A|}{d}\right)\right)$. Hence, the following lemma is optimal up to a multiplicative constant.

LEMMA 3.4. *Let $\delta, c > 0$, let $q = \frac{\delta}{d}$, let $L \subseteq Q^d$ be such that every vertex in Q^d is at distance at most two from at least cd^2 vertices in L and let $A \cup B = V(L)$ be a partition of $V(L)$ with $\min\{|A|, |B|\} = t$. Then there exists a constant $c' > 0$ such that there exists a family of $c't \left(1 - \frac{\log_2 t}{d}\right)$ vertex-disjoint A - B -paths of length at most five in Q_q^d with probability at least $1 - \exp\left(-c't \left(1 - \frac{\log_2 t}{d}\right)\right)$.*

PROOF. Throughout the proof we will introduce a sequence of constants c_1, c_2, c_3, \dots under the assumption that each c_i is sufficiently small in terms of the preceding c_j, δ and c .

Let $s := t(d - \log_2 t)$ and let us define

$$\hat{N}(A) = \left\{ v \in V(Q^d) : v \notin A \text{ and } |N_{Q^d}(v) \cap A| \geq c_1 d \right\},$$

and similarly

$$\hat{N}(B) = \left\{ v \in V(Q^d) : v \notin B \text{ and } |N_{Q^d}(v) \cap B| \geq c_1 d \right\}.$$

We first note that we may assume that there are at most $c_6 s$ edges between A and B in Q^d , since otherwise, by Lemma 2.9, with probability at least $1 - \exp(-c_7 s d^{-1})$ there will be a matching of size at least $c_7 s d^{-1}$ between A and B in Q^d . In particular, we can assume that

$$(1) \quad |B \cap \hat{N}(A)| \leq c_5 s d^{-1} \quad \text{and} \quad |A \cap \hat{N}(B)| \leq c_5 s d^{-1}.$$

By assumption every vertex in Q^d is at distance at most two from at least cd^2 vertices in L . Hence, we can partition the vertices of Q^d into two disjoint subsets A' and B' such that $A \subseteq A'$ and $B \subseteq B'$, each vertex in $A' \setminus A$ is within distance two of at least $\frac{cd^2}{2}$ vertices in A and each vertex in $B' \setminus B$ is within distance two of at least $\frac{cd^2}{2}$ vertices in B .

Since $|A'|, |B'| \geq \min\{|A|, |B|\} = t$, it follows from Theorem 2.10 that there is a set F of s edges between A' and B' . Then, at least one of the following four cases happens:

- i) At least $\frac{s}{4}$ edges of F have an endpoint in A ;
- ii) At least $\frac{s}{4}$ edges of F have an endpoint in $A' \cap \hat{N}(A)$;
- iii) At least $\frac{s}{4}$ edges of F have an endpoint in $A' \cap \hat{N}(B)$;
- iv) At least $\frac{s}{4}$ edges of F have an endpoint in A_0 , where $A_0 = A' \setminus (A \cup \hat{N}(A) \cup \hat{N}(B))$.

We will see that case iv) is the most complicated case, so let us assume for now that cases i)–iii) do not hold. We will indicate briefly how to deal with the other cases at the end. If we let $F' \subseteq F$ be a set of $\frac{s}{4}$ edges, each of which has an endpoint in A_0 , then again at least one of the following four cases happens:

- I) At least $\frac{s}{16}$ edges of F' have an endpoint in B ;
- II) At least $\frac{s}{16}$ edges of F' have an endpoint in $B' \cap \hat{N}(B)$;
- III) At least $\frac{s}{16}$ edges of F' have an endpoint in $B' \cap \hat{N}(A)$;
- IV) At least $\frac{s}{16}$ edges of F' have an endpoint in B_0 , where $B_0 = B' \setminus (B \cup \hat{N}(A) \cup \hat{N}(B))$.

Again we will assume for now that that cases I)–III) do not hold.

We will construct our family of paths using a sequence of matchings: the first M_1 between A' and B' ; then M_2 and M_3 joining some subset of the endpoints of M_1 to $\hat{N}(A)$ and $\hat{N}(B)$, respectively; and then finally M_4 and M_5 joining some subset of the endpoints of these matchings to A and B , respectively. See Figure 1.

Let $F_1 \subseteq F'$ be a set of $\frac{s}{16}$ edges whose endpoints lie in A_0 and B_0 . Then, by Lemma 2.9, with probability at least $1 - \exp(-c_2 s d^{-1})$ there is a matching M_1 contained in $(F_1)_q$ of size at least $c_2 s d^{-1}$. Let $A_1 \subseteq A_0$ and $B_1 \subseteq B_0$ be the endpoints of this matching.

Since each vertex $u_i \in A_1$ is within distance two in Q^d of at least $\frac{cd^2}{2}$ vertices in A and $\Delta(Q^d) = d$, u_i must have at least $2c_1 d$ neighbours which are each adjacent to at least $c_1 d$ vertices in A . Furthermore, since $u_i \notin \hat{N}(A)$, u_i has less than $c_1 d$ neighbours in A . It follows that we can fix, for each vertex $u_i \in A_1$, a star T_i in Q^d rooted at u_i with $c_1 d$ leaves, such that each leaf is adjacent in Q^d to at least $c_1 d$ vertices in A and no leaf is in A , and so each leaf is in $\hat{N}(A)$.

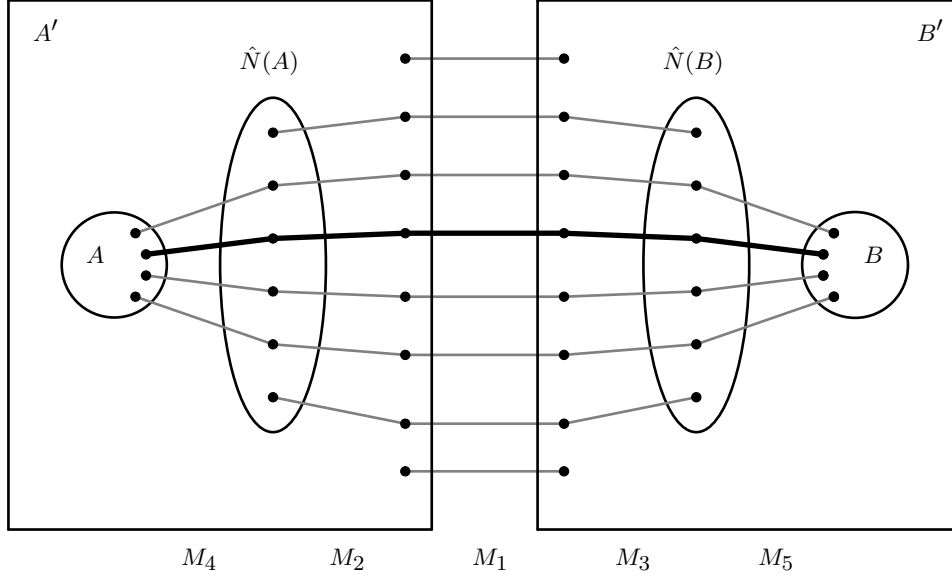


FIG 1. The sequence of matchings M_1 – M_5 , with one of the paths resulting from the concatenation highlighted in bold. Note that, unlike in the diagram, it may be the case that $\hat{N}(A)$ meets B' , or $\hat{N}(B)$ meets A' , or $\hat{N}(A)$ meets $\hat{N}(B)$.

Note that, since each edge in these stars is from $A_1 \subseteq A'$ to $\hat{N}(A)$ and $B_0 \subseteq B'$ is disjoint from $A' \cup \hat{N}(A)$, it follows that none of the edges in these stars lie in F_1 , each edge of which meets B_0 .

Let C_1 be the set of vertices which are leaves in some star T_i and let F_2 be the set of edges between A_1 and C_1 contained in these stars. Then $F_2 \cap F_1 = \emptyset$ and $|F_2| = |A_1|c_1d = c_1c_2s$.

Then, again by Lemma 2.9, with probability at least $1 - \exp(-c_3sd^{-1})$ there is a matching M_2 contained in $(F_2)_q$ of size at least c_3sd^{-1} . Let $A_2 \subseteq A_1$ be the endpoints of this matching in A_1 and let B_2 be the set of vertices in B_1 joined to a vertex in A_2 via the matching M_1 . We now make a similar argument for the vertices in B_2 .

Namely, since each vertex in B_2 is within distance two in Q^d of at least $\frac{cd^2}{2}$ vertices in B , $\Delta(Q^d) = d$, and no vertex in B_2 is in B or adjacent to B , it follows that we can fix, for each vertex $v_i \in B_2$ a star T'_i in Q^d rooted at v_i with c_1d leaves, such that each leaf is adjacent in Q^d to at least c_1d vertices in B and no leaf is in B , and so each leaf is in $\hat{N}(B)$. Note that, as before, since each edge in these stars is from $B_2 \subseteq B'$ to $\hat{N}(B)$ and $A_0 \subseteq A'$ is disjoint from $B' \cup \hat{N}(B)$, none of the edges in these stars lie in F_1 or F_2 , each edge of which has an endpoint in A_0 .

Hence, if we let D_1 be the set of vertices which are leaves in some star T'_i and consider the set of edges F_3 between B_2 and D_1 contained in these stars, then $F_3 \cap (F_1 \cup F_2) = \emptyset$ and $|F_3| = |B_2|c_1d = c_1c_3s$. Again, by Lemma 2.9, we can conclude that with probability at least $1 - \exp(-c_4sd^{-1})$ there is a matching M_3 contained in $(F_3)_q$ of size at least $2c_4sd^{-1}$.

By combining the matchings M_1, M_2 and M_3 , we obtain a family of vertex-disjoint paths $\mathcal{P}' = \{P_1, \dots, P_r\}$ of size $2c_4sd^{-1}$, where each P_i has one endpoint $x_i \in C_1$, which by construction of C_1 is adjacent in Q^d to c_1d vertices in A , and a second endpoint $y_i \in D_1$, which similarly is adjacent in Q^d to c_1d vertices in B . Let $C_2 \subseteq C_1$ and $D_2 \subseteq D_1$ be the sets of endpoints of \mathcal{P}' .

We note that the set of edges between C_2 and A and the set of edges between D_2 and B do not intersect with the set of edges $F_1 \cup F_2 \cup F_3$ which we already exposed. Indeed, every edge in $F_1 \cup F_2$ has an endpoint in A_0 , which by construction is disjoint from $C_2 \subseteq \hat{N}(A)$,

$D_2 \subseteq \hat{N}(B)$, A and B . Similarly every edge in $F_1 \cup F_3$ has an endpoint in B_0 , which is again disjoint from C_2, D_2, A and B . However, the set of edges between C_2 and A might intersect with the set of edges between D_2 and B .

To deal with this, let I_1 be the set of i such that some edge from $x_i \in C_2$ to A coincides with an edge from D_2 to B . Then, since $A \cap B = \emptyset$, it follows that $x_i \in B$. However, since each $x_i \in C_2 \subseteq \hat{N}(A)$, it follows that $\{x_i : i \in I_1\} \subseteq \hat{N}(A) \cap B$ and hence $|I_1| \leq |\hat{N}(A) \cap B| \leq c_5 s d^{-1}$ by (1).

Similarly, if we let I_2 be the set of i such that some edge from $y_i \in D_2$ to B coincides with an edge from C_2 to A , then we can conclude that $|I_2| \leq c_5 s d^{-1}$.

Hence, if we let $\mathcal{P} = \{P_i : i \in [r] \setminus (I_1 \cup I_2)\}$ and let C_3 and D_3 be the endpoints of these paths, then $|\mathcal{P}| \geq c_4 s d^{-1}$. Then, there is a set F_4 of at least $|\mathcal{P}| c_1 d = c_8 s$ edges between C_3 and A and by construction $F_4 \cap (F_1 \cup F_2 \cup F_3) = \emptyset$. Hence, by Lemma 2.9, with probability at least $1 - \exp(-c_9 s d^{-1})$ there is a matching M_4 of size $c_9 s d^{-1}$ in $(F_4)_q$ between C_3 and A . Let $D_4 \subseteq D_3$ be the endpoints of the paths in \mathcal{P} whose other endpoint is an endpoint of an edge in M_4 .

As before, there is a set F_5 of at least $c_{10} s$ edges between D_4 and B , and by construction $F_5 \cap (F_1 \cup F_2 \cup F_3 \cup F_4) = \emptyset$. Therefore, again by Lemma 2.9, with probability at least $1 - \exp(-c_{11} s d^{-1})$ there is a matching M_5 of size $c_{11} s d^{-1}$ in $(F_5)_q$ between D_4 and B .

In particular, by combining the matchings M_1, M_2, M_3, M_4 and M_5 we can construct a family of $c_{11} s d^{-1}$ A - B -paths in Q_q^d . Whilst these paths are not necessarily vertex-disjoint, since they are formed via the concatenation of matchings, it is clear that no two paths use the same vertex in the same position. In particular, each path meets at most 25 other paths, and so we can greedily chose a subfamily of at least $\frac{c_{11}}{25} s d^{-1}$ paths which are vertex-disjoint.

Note that, throughout the argument we assumed a finite number of whp events occurred, and in each case the probability of failure was at most $\exp(-c_{11} s d^{-1})$, and so the conclusion holds with probability at least $1 - \exp(-c_{12} s d^{-1})$. In particular, it follows that the claim holds with $c' = c_{12}$.

If one of the cases i)–iii) or I)–III) holds, then we can avoid building some of the matchings M_i . For example, if case ii) holds, then instead of building M_2 and M_4 we can instead build a large matching directly from A_0 to A before building M_1, M_3 and M_5 . Similarly, if case i) holds, then we only need to build M_1, M_3 and M_5 . If neither case i) nor ii) hold, but case iii) holds, then A' must contain at least $\frac{s}{4d}$ vertices of $\hat{N}(B)$, and so there is a set of at least $\frac{c_4 s}{4}$ edges from A' to B , using which we can build a matching of size $c_5 s$ using Lemma 2.9. We can then extend this matching to a family of A - B -paths using matchings M_2 and M_4 as before. Similar arguments work when one of the cases I)–III) holds.

In fact, some possibilities are already excluded by our earlier assumptions, for example we can assume i) and I) do not simultaneously hold since we are assuming there are at most $c_6 s$ edges between A and B . \square

At this point we have all the necessary tools to prove the following theorem, which clearly implies Theorem 1.2.

THEOREM 3.5. *Let $0 < \alpha \ll \epsilon$, let $p = \frac{1+\epsilon}{d}$ and let L_1 be the largest component of Q_p^d . Then there exists a constant $\beta > 0$ such that whp:*

(a) *every subset $S \subseteq V(L_1)$ satisfying $|S| \leq \frac{|V(L_1)|}{2}$ is such that*

$$\left| N_{Q_p^d}(S) \right| \geq \beta |S| \left(1 - \frac{\log_2 |S|}{d} \right)^2 d^{-3} \geq \beta |S| d^{-5};$$

(b) every subset $S \subseteq V(L_1)$ satisfying $\alpha n \leq |S| \leq \frac{|V(L_1)|}{2}$ is such that

$$\left| N_{Q_p^d}(S) \right| \geq \beta |S| d^{-2} (\log d)^{-1}.$$

PROOF. Throughout the proof we will introduce a sequence of constants c_1, c_2, c_3, \dots under the assumption that each c_i is sufficiently small in terms of the preceding c_j , ϵ and α .

We will argue using a sprinkling argument. Let $c_2 \ll c_1 \ll \alpha$, let $q_2 = \frac{c_2}{d}$ and let $q_1 = \frac{p-q_2}{1-q_2}$. Note that, since $c_2 \ll \epsilon$ it follows that q_1 is still supercritical. Furthermore, if we let γ and γ_1 be the survival probabilities of the $\text{Po}(1+\epsilon)$ and $\text{Po}(dq_1)$ branching processes, respectively, then since $c_2 \ll c_1$ we may assume that

$$(2) \quad \gamma - \gamma_1 \leq c_1.$$

We will generate independently two random subgraphs $Q_{q_1}^d$ and $Q_{q_2}^d$ and let $Q_1 := Q_{q_1}^d$ and $Q_2 := Q_1 \cup Q_{q_2}^d$, so that $Q_2 \sim Q_p^d$.

Let us first note a few likely properties of the graphs Q_1 and Q_2 . Firstly, it follows from Lemma 2.8 that whp

$$(3) \quad Q_1 \text{ contains at most } nd^{-4} \text{ vertices of degree at least } \log d.$$

Furthermore, if we let L'_1 and L_1 be the largest components in Q_1 and Q_2 , respectively, then by Theorem 2.1 whp

$$(4) \quad |V(L_1)| = (\gamma \pm c_3)n \text{ and } |V(L'_1)| = (\gamma_1 \pm c_3)n.$$

Note that (2) and (4) imply that $|V(L'_1)| \geq (\gamma - c_1 - c_3)n \geq \frac{3}{4}|V(L_1)|$.

Given a vertex $v \in V(L'_1)$, let C_v be the set of vertices which are contained in some component of $L_1 - L'_1$ which is adjacent to v in Q_2 . Then, by Lemma 3.2 there exists a constant $K_2 := K_2(dq_1 - 1) > 0$ such that whp

$$(5) \quad |C_v| \leq K_2 d \text{ for every vertex in } v \in V(L'_1),$$

where we may assume that $K_2^{-1} := c_4 \ll c_3$. Note, in particular, that (5) implies that every component in $L_1 - L'_1$ has size at most $K_2 d$.

Moreover, by Lemma 3.3, whp

$$(6) \quad \text{every vertex in } Q^d \text{ is at distance at most two from at least } c_5 d^2 \text{ vertices in } L'_1.$$

In what follows, we will assume that (2)–(6) hold.

We want to split L'_1 into a disjoint family of relatively small, connected pieces, however in order to treat the cases (a) and (b) we will need to use slightly different families. Given $s \in \mathbb{N}$, let us write

$$b(s) := 1 - \frac{\log_2 s}{d}.$$

- For case (a), for each $1 \leq s \leq \frac{|V(L_1)|}{2}$ we use Lemma 2.2 with $C_1 = C_2 = d$, $\ell = c_8^{-1} b(s)^{-1} d$ and $r = 0$ to split L'_1 into a family $\mathcal{C}(s)$ of vertex-disjoint connected subgraphs, which we will refer to as *pieces*, such that all pieces have size between $c_8^{-1} b(s)^{-1} d$ and $c_8^{-1} b(s)^{-1} d^2$.
- Similarly, for case (b) we use (3) and Lemma 2.2 with $C_1 = d$, $C_2 = \log d$, $\ell = c_8^{-1} d$ and $r = nd^{-4}$ to split L'_1 into a family \mathcal{C}' of connected pieces such that at most nd^{-4} of the pieces in \mathcal{C}' have size between $c_8^{-1} d$ and $c_8^{-1} d^2$ and the rest have size between $c_8^{-1} d$ and $c_8^{-1} d \log d$.

We now state two very similar claims about the existence of certain path families, whose proofs we defer to the end of this proof. The first will be useful for case (a).

CLAIM 3.6. *Whp for any $1 \leq t \leq 2s \leq |V(L_1)|$ and any partition of $\mathcal{C}(s)$ into two sets $\{\mathcal{C}_A, \mathcal{C}_B\}$, where $A := \bigcup \mathcal{C}_A$ and $B := \bigcup \mathcal{C}_B$, with $\min\{|A|, |B|\} = t$, there is a family of at least $c_7 t b(t)$ vertex-disjoint A - B -paths of length at most five in $Q_{q_2}^d$.*

The second, which follows by a similar argument, will be useful for case (b).

CLAIM 3.7. *Whp for any partition of \mathcal{C}' into two sets $\{\mathcal{C}'_A, \mathcal{C}'_B\}$, where $A := \bigcup \mathcal{C}'_A$ and $B := \bigcup \mathcal{C}'_B$, with $\min\{|A|, |B|\} = t \geq \frac{\alpha n}{4}$ there is a family of at least $c_7 t d^{-1}$ vertex-disjoint A - B -paths of length at most five in $Q_{q_2}^d$.*

Let us further assume that $Q_{q_2}^d$ satisfies the conclusions of Claims 3.6 and 3.7. We will subsequently be able to deduce the claimed expansion properties deterministically.

Let $S \subseteq V(L_1)$ be an arbitrary subset of size $s \leq \frac{|V(L_1)|}{2}$ and let $S_1 := S \cap V(R)$ be the vertices of S which lie in $R := L_1 - L'_1$.

In order to deal with case (a) let us further split $S \cap V(L'_1)$ into two parts as follows:

- S_2 is the set of vertices which are contained in pieces $C \in \mathcal{C}(s)$ such that $S \cap C \neq \emptyset$ and $S \setminus C \neq \emptyset$;
- S_3 is the set of vertices which are contained in pieces $C \in \mathcal{C}(s)$ such that $C \subseteq S$.

Similarly, to deal with case (b) we further split $S \cap V(L'_1)$ into two parts as follows:

- S'_2 is the set of vertices which are contained in pieces $C \in \mathcal{C}'$ such that $S \cap C \neq \emptyset$ and $S \setminus C \neq \emptyset$;
- S'_3 is the set of vertices which are contained in pieces $C \in \mathcal{C}'$ such that $C \subseteq S$.

Case (a) :

Suppose first that $|S_1| \geq \frac{s}{2}$. Then, by (5) we can choose some subset $X \subseteq V(L'_1)$ of size at least $\frac{c_4 s}{2d}$ together with a disjoint family $\{C(x) : x \in X\}$ of components of R such that x is adjacent to $C(x)$ in Q_2 for each $x \in X$ and each $C(x)$ meets S .

For each $x \in X$ either $x \in S$, or there is some vertex in $C(x) \cup \{x\}$ which lies in the neighbourhood of S . In particular, it follows that either $|N_{Q_2}(S)| \geq \frac{c_4 s}{4d}$ or

$$|S_2 \cup S_3| = |S \cap V(L'_1)| \geq \min \left\{ \frac{s}{2}, \frac{c_4 s}{4d} \right\} = \frac{c_4 s}{4d}.$$

Suppose then that $|S_2| \geq \frac{c_4 c_7 s b(s)}{16d}$. Since each piece in $\mathcal{C}(s)$ has size at most $c_8^{-1} b(s)^{-1} d^2$ it follows that S_2 contains vertices in at least $\frac{c_4 c_7 c_8 s b(s)^2}{16d^3}$ pieces of $\mathcal{C}(s)$. Since each such piece contains at least one vertex in the neighbourhood of S , it follows that $|N_{Q_2}(S)| \geq \frac{c_4 c_7 c_8 s b(s)^2}{16d^3}$.

On the other hand, if $|S_2 \cup S_3| \geq \frac{c_4 s}{4d}$ and $|S_2| \leq \frac{c_4 c_7 s b(s)}{16d}$ then $|S_3| \geq \frac{c_4 s}{8d}$. Let

$$\mathcal{C}_A = \{C \in \mathcal{C}(s) : C \subseteq S\} \quad \text{and} \quad \mathcal{C}_B = \mathcal{C}(s) \setminus \mathcal{C}_A,$$

and let $A := \bigcup \mathcal{C}_A$ and $B := \bigcup \mathcal{C}_B$. Then $|A| = |S_3| \geq \frac{c_4 s}{8d}$, and by (2) and (4)

$$|B| \geq |V(L'_1)| - |S| \geq |V(L'_1)| - \frac{|V(L_1)|}{2} \geq \frac{|V(L_1)|}{4} \geq \frac{c_4 s}{8d}.$$

Hence, by Claim 3.6 there is family of $\frac{c_4 c_7 s b(\frac{c_4 s}{8d})}{8d} \geq \frac{c_4 c_7 s b(s)}{8d}$ vertex-disjoint A - B -paths in $Q_{q_2}^d$. Since at most $|S_2| \leq \frac{c_4 c_7 s b(s)}{16d}$ of these paths can meet S_2 , the rest of these paths go from $A \subseteq S$ to $B \setminus S_2 \subseteq S^c$, and so each path contributes a vertex to the neighbourhood of S . It follows that $|N_{Q_2}(S)| \geq \frac{c_4 c_7 s b(s)}{16d}$. In particular, in all cases

$$|N_{Q_2}(S)| = |N_{Q_2}(S)| \geq c_9 s b(s)^2 d^{-3}.$$

Case (b) :

We first note that, by (4) and (2)

$$|S_1| \leq |V(R)| \leq |V(L_1)| - |V(L'_1)| \leq (\gamma + c_3 - \gamma_1 + c_3)n \leq (c_1 + 2c_3)n$$

and hence $|S'_2 \cup S'_3| \geq |S| - |S_1| \geq \frac{\alpha n}{2}$.

If $|S'_2| \geq \frac{c_7 \alpha n}{8d}$ then, since at most $nd^{-4}c_8^{-1}d^2 = c_8^{-1}nd^{-2}$ vertices lie in pieces of \mathcal{C}' of size larger than $c_8^{-1}d \log d$, it follows that S'_2 contains vertices in at least $\frac{c_8 c_7 \alpha n}{16d^2 \log d}$ pieces of \mathcal{C}' . Since each piece contains at least one vertex in the neighbourhood of S it follows that $|N_{Q_2}(S)| \geq \frac{c_8 c_7 \alpha n}{16d^2 \log d}$.

On the other hand, if $|S'_2| \leq \frac{c_7 \alpha n}{8d}$ and $|S'_2 \cup S'_3| \geq \frac{\alpha n}{2}$, then $|S'_3| \geq \frac{\alpha n}{4}$. Let

$$\mathcal{C}_A = \{C \in \mathcal{C}' : C \subseteq S\} \quad \text{and} \quad \mathcal{C}_B = \mathcal{C}' \setminus \mathcal{C}_A,$$

and let $A := \bigcup \mathcal{C}'_A$ and $B := \bigcup \mathcal{C}'_B$. Then $|A| = |S'_3| \geq \frac{\alpha n}{4}$ and by (4)

$$|B| \geq |V(L'_1)| - |S| \geq |V(L'_1)| - \frac{|V(L_1)|}{2} \geq \frac{|V(L_1)|}{4} \geq \frac{\alpha n}{4}.$$

Hence, by Claim 3.7 there is a family of $\frac{c_7 \alpha n}{4d}$ vertex-disjoint A - B -paths in $Q_{q_2}^d$. Since at most $|S'_2| \leq \frac{c_7 \alpha n}{8d}$ of these paths can meet S'_2 , the rest of these paths go from $A \subseteq S$ to $B \setminus S'_2 \subseteq S^c$, and so each path contributes a vertex to the neighbourhood of S . It follows that $|N_{Q_2}(S)| \geq \frac{c_7 \alpha n}{8d}$. In particular, in both cases

$$|N_{Q_2}(S)| = |N_{Q_2}(S)| \geq c_9 nd^{-2}(\log d)^{-1}.$$

The conclusion of the theorem then follows with $\beta = c_9$.

PROOF OF CLAIM 3.6. Let us fix an s and such a partition $\{\mathcal{C}_A, \mathcal{C}_B\}$, where $\min\{|A|, |B|\} = t \leq 2s$. Note that, since each piece in $\mathcal{C}(s)$ has size at least $c_8^{-1}b(s)^{-1}d$, it follows that

$$t \geq c_8^{-1}b(s)^{-1}d =: t_{\min} \quad \text{and} \quad s \geq \frac{t_{\min}}{2} =: s_{\min}.$$

Let k be the number of pieces of \mathcal{C} contained in \mathcal{C}_A , where

$$k_1 := c_8 tb(s)d^{-2} \leq k \leq c_8 tb(s)d^{-1} := k_2.$$

Note that there are at most

$$\binom{|\mathcal{C}|}{k} \leq n^k$$

partitions of this form. Since $A \cup B = V(L'_1)$, by Lemma 3.4 the probability that such a partition does not satisfy the conclusion of the claim is at most $\exp(-c_7 tb(t))$, and so by the union bound the probability that some partition does not satisfy the conclusion of the lemma is at most

$$\begin{aligned} \sum_{s=s_{\min}}^{\frac{|V(L_1)|}{2}} \sum_{t=t_{\min}}^{2s} \sum_{k=k_1}^{k_2} \exp(-c_7 tb(t)) n^k &\leq \sum_{s=s_{\min}}^{\frac{|V(L_1)|}{2}} \sum_{t=t_{\min}}^{2s} (k_2 - k_1) \exp(-c_7 tb(t)) n^{k_2} \\ &\leq \sum_{s=s_{\min}}^{\frac{|V(L_1)|}{2}} \sum_{t=t_{\min}}^{2s} c_8 tb(s)d^{-1} \exp(-c_7 tb(t)) n^{c_8 tb(s)d^{-1}} \\ &\leq \sum_{s=s_{\min}}^{\frac{|V(L_1)|}{2}} \sum_{t=t_{\min}}^{2s} \exp\left(-\frac{c_7}{2} tb(t)\right) \end{aligned}$$

$$\leq \sum_{s=s_{\min}}^{\frac{|V(L_1)|}{2}} 2s \exp\left(-\frac{c_7}{2} t_{\min} b(t_{\min})\right) = o(1),$$

where in the above we used that, since $t \leq \frac{n}{2}$, $b(s) \leq 2b(t)$ and so

$$c_8 t b(s) d^{-1} n^{c_8 t b(s) d^{-1}} \leq \exp(c_8 t b(s)) \leq \exp\left(\frac{c_7 t b(t)}{2}\right),$$

and that $\frac{c_7}{2} t_{\min} b(t_{\min}) \geq \frac{c_7 t_{\min}}{8} \geq d$. \square

PROOF OF CLAIM 3.7. Let us fix such a partition $\{C'_A, C'_B\}$, where $\min\{|A|, |B|\} = t \geq \frac{\alpha n}{4}$. Since there are at most $c_8 n d^{-1}$ pieces in C' , there are at most

$$2^{c_8 n d^{-1}}$$

partitions of C' . Since $A \cup B = V(L'_1)$ and $1 - \frac{\log_2 t}{d} \geq d^{-1}$, by Lemma 3.4 the probability that such a partition does not satisfy the conclusion of the claim is at most $\exp(-c_7 t d^{-1})$, and so by the union bound the probability that any partition does not satisfy the conclusion of the lemma is at most

$$\sum_{t \geq \frac{\alpha n}{4}} \exp(-c_7 t d^{-1}) 2^{c_8 n d^{-1}} \leq \sum_{t \geq \frac{\alpha n}{4}} \exp\left(-\frac{c_7 t d^{-1}}{2}\right) = o(1).$$

\square

\square

REMARK 1. The restriction to subsets of size at most $\frac{|V(L_1)|}{2}$ in Theorem 3.5 is mostly for ease of presentation. It is relatively simple to see that subsets of $V(L_1)$ of size up to $(1 - 2\alpha)|V(L_1)|$ will have similar expansion properties.

Indeed, let S be any subset of $V(L_1)$ such that $\frac{|V(L_1)|}{2} \leq |S| \leq (1 - 2\alpha)|V(L_1)|$ and whose boundary satisfies $|N(S)| \leq \beta |S| d^{-2} (\log d)^{-1} \leq \alpha n$. It follows that the set $W := V(L_1) \setminus (S \cup N(S))$ is such that $\alpha n \leq |W| \leq \frac{|V(L_1)|}{2}$ and $N(W) \subseteq N(S)$.

In particular, under the conclusion of Theorem 3.5 (b),

$$(7) \quad |N(S)| \geq |N(W)| \geq \beta |W| d^{-2} (\log d)^{-1} \geq \frac{\alpha \beta}{1 - 2\alpha} |S| d^{-2} (\log d)^{-1} \geq \beta |S| d^{-5}.$$

PROOF OF THEOREM 1.2. The claim follows immediately from Theorem 3.5 (a). \square

We note that it is relatively easy to see with a similar argument that Claims 3.6–3.7 imply that whp the subset $V(L'_1) \subseteq V(L_1)$ has good expansion properties, in fact slightly better than the expansion we have for the whole giant component L_1 , although this expansion may happen ‘outside’ of $V(L'_1)$.

LEMMA 3.8. *Let $\delta, \epsilon > 0$, let $q_1 = \frac{1+\epsilon}{d}$ and let $q_2 = \frac{\delta}{d}$. Let L'_1 be the largest component of $Q_{q_1}^d$ and let $Q_2 := Q_{q_1}^d \cup Q_{q_2}^d$. Then whp every subset $S \subseteq V(L'_1)$ of size $|S| \leq \frac{|V(L'_1)|}{2}$ is such that*

$$|N_{Q_2}^5(S) \cap V(L'_1)| \geq \beta |S| \left(1 - \frac{\log_2 |S|}{d}\right)^2 d^{-2}.$$

REMARK 2. As with Theorem 3.5, there will be somewhat stronger expansion if we additionally assume that $|S| \geq \alpha n$.

PROOF OF LEMMA 3.8. We sketch the argument below, without keeping careful track of the constants.

Given a subset $S \subseteq V(L'_1)$, we can consider the partition $S = S_2 \cup S_3$ as in the proof of case (a) of Theorem 3.5. Each piece $C \in \mathcal{C}(s)$ containing a vertex of S_2 also contains a vertex in $N_{Q_2}(S) \cap V(L'_1)$, and so, since each piece in $\mathcal{C}(s)$ has size $O(b(s)^{-1}d^2)$, we are done if $|S_2| = \Omega(sb(s))$. Hence, we may assume that $|S_2| = o(sb(s))$ and so $|S_3| = \Omega(s)$.

If we let \mathcal{C}_A be the set of pieces of $\mathcal{C}(s)$ contained in S and \mathcal{C}_B be the rest, where $A := \bigcup \mathcal{C}_A$ and $B := \bigcup \mathcal{C}_B$, then, as before, $|A| = |S_3| = \Omega(s)$ and $|B| = \Omega(s)$ and so by Claim 3.6 whp there is a family of $\Omega(sb(s))$ vertex-disjoint A - B -paths of length at most five in $Q_{q_2}^d$. Since at most $|S_2| = o(sb(s))$ of them have an endpoint in S'_2 , there are $\Omega(sb(s))$ of them with an endpoint in $V(L'_1) \setminus S$. In particular, in both cases

$$|N_{Q_2}^5(S) \cap V(L'_1)| = \Omega(sb(s)^2 d^{-2}).$$

□

In a supercritical random graph $G(d+1, p)$ with $p = \frac{1+\epsilon}{d}$ the giant component itself will likely not be an α -expander for any constant $\alpha > 0$. Indeed, let the 2-core of a graph G be the maximal subgraph of G of minimum degree at least two. Standard results imply that it is likely that there are logarithmically sized pendant trees attached to the 2-core of $G(d+1, p)$, and also that the 2-core of $G(d+1, p)$ typically contains logarithmically long bare paths, both of which lead to logarithmically large sets whose neighbourhoods have constant size.

However, the results of Benjamini, Kozma and Wormald [4] and Krivelevich [46] imply that whp the giant component contains a linear sized subgraph which is an α -expander. In the case of Q_p^d is it a simple consequence of Theorem 3.5, using some ideas of Krivelevich [47], that we can also pass to a linear sized subset of the giant component with a significantly better expansion ratio than that guaranteed for the whole of L_1 by Theorem 3.5.

PROOF OF THEOREM 1.3. Let $\alpha' \ll \epsilon$ and let us assume that the conclusion of Theorem 3.5 and (7) holds with this α' for some constant β' . Note that, by Theorem 2.1 there is some constant $\gamma = \gamma(\epsilon)$ such that whp $|V(L_1)| = (1 + o(1))\gamma n$. By (7) whp every set $S \subseteq V(L_1)$ such that $\alpha' n \leq |S| \leq (1 - 2\alpha')|V(L_1)|$ is such that

$$|N_{L_1}(S)| \geq \frac{\alpha' \beta'}{1 - 2\alpha'} |S| d^{-2} (\log d)^{-1} =: t|S|.$$

Let $U \subseteq V(L_1)$ be a set of maximal cardinality such $|U| < \alpha' n$ and $|N_{L_1}(U)| < t|U|$ and let $H := L_1 - U$. Note that $|V(H)| \geq (1 - \frac{2\alpha'}{\gamma})|V(L_1)|$. Suppose that there is some subset $W \subseteq V(H)$ with $|W| \leq \frac{|V(H)|}{2}$ and $|N_H(W)| < t|W|$. Then, in particular,

$$|N_{L_1}(U \cup W)| \leq |N_{L_1}(U)| + |N_H(W)| \leq t|U \cup W|.$$

By construction, $|U \cup W| \leq \alpha' n + \frac{|V(H)|}{2} \leq (1 - 2\alpha')|V(L_1)|$, since $|V(L_1)| = (1 + o(1))\gamma n$ and $\alpha' \ll \epsilon$. However, by maximality of U it follows that $|U \cup W| \geq \alpha' n$. Hence $U \cup W$ contradicts our assumption on the expansion of L_1 .

It follows that H is a t -expander, and so the conclusion of the theorem follows with $\alpha = \frac{2\alpha'}{\gamma}$ and $\beta = \frac{\alpha' \beta'}{1 - 2\alpha'}$. □

4. Consequences of expansion in the giant component

4.1. *Mixing time of the lazy random walk* Given a graph $G = (V, E)$, the *lazy simple random walk* on G is a random walk on V which remains at the same vertex with probability $\frac{1}{2}$ in each time step, and otherwise moves to a uniformly chosen random neighbour of its current position. The *stationary distribution* π of the lazy random walk is given by $\pi(v) = \frac{d_G(v)}{2|E|}$ for each $v \in V$. For a subset $S \subseteq V$ let us write $\pi(S) = \sum_{v \in S} \pi(v)$ and let

$$\pi_{\min} = \min\{\pi(v) : v \in V\}.$$

The *edge measure* Q of the random walk is given by

$$Q(x, y) := \pi(x)P(x, y) \quad \text{and} \quad Q(A, B) = \sum_{x \in A, y \in B} Q(x, y),$$

where P is the transition matrix of the lazy random walk, so that

$$Q(x, y) = \begin{cases} \frac{1}{4|E|} & \text{if } xy \in E; \\ \frac{d_G(x)}{4|E|} & \text{if } x = y; \\ 0 & \text{otherwise.} \end{cases}$$

The *bottleneck ratio* of a subset $S \subseteq V(G)$ is defined to be

$$\Phi(S) = \frac{Q(S, S^c)}{\pi(S)} = \frac{e_G(S, S^c)}{2d_G(S)},$$

where $d_G(S) = \sum_{v \in S} d_G(v)$ is the *total degree* of S . The bottleneck ratio of the random walk, sometimes known as the *Cheeger constant* of G , is given by

$$\Phi(G) := \min_{S: \pi(S) \leq \frac{1}{2}} \Phi(S).$$

Note that, for a k -regular graph G , $\Phi(G) \geq \alpha$ is equivalent to G being an $f(\alpha, k)$ -edge-expander for some function $f(\alpha, k)$.

Let $P^t(v, \cdot)$ denote the distribution on V given by starting the lazy random walk at $v \in V$ and running for t steps, and let us define

$$d(t) := \max_{v \in V} d_{TV}(P^t(v, \cdot), \pi)$$

to be the maximal distance (over $v \in V$) between $P^t(v, \cdot)$ and the stationary distribution π , where we measure this distance in terms of the *total variation distance*. That is, given two random variables X and Y distributed on the same finite set Z we have

$$d_{TV}(X, Y) = \frac{1}{2} \sum_{z \in Z} |\mathbb{P}(X = z) - \mathbb{P}(Y = z)|.$$

The *mixing time* of the lazy random walk is then defined as

$$t_{\text{mix}} := \min \left\{ t : d(t) \leq \frac{1}{4} \right\}.$$

See [53] for more background on mixing time of Markov chains.

It is relatively easy to use our result on the vertex-expansion of the giant component L_1 of Q_p^d to show that the likely value of the Cheeger constant of L_1 is $\Omega(d^{-6})$. Again, with a more careful argument we can improve this result somewhat.

LEMMA 4.1. *Let $\epsilon > 0$, let $p = \frac{1+\epsilon}{d}$ and let L_1 be the largest component in Q_p^d . Then whp $\Phi(L_1) = \Omega(d^{-5})$.*

PROOF. Let $\alpha \ll \epsilon$ be a positive constant and let γ be the survival probability of the $\text{Po}(1 + \epsilon)$ branching process. Note that, since L_1 is connected, by Theorem 2.1 whp

$$d_{L_1}(V(L_1)) \geq 2(|V(L_1)| - 1) \geq (2 + o(1))\gamma n.$$

Let us further assume that the conclusions of Theorem 3.5 and (7) hold with this α for some constant β . Note that, for every subset $S \subseteq V(L_1)$ we have that

$$e_{Q_p^d}(S) = e_{L_1}(S) \quad \text{and} \quad e_{Q_p^d}(S, S^c) = e_{L_1}(S, S^c).$$

We first note that it is unlikely that any very large subset of $V(L_1)$ has small total degree. Indeed, suppose $S \subseteq V(L_1)$ is such that $|S| \geq (1 - 2\alpha)|V(L_1)|$ and $d_{L_1}(S) \leq \frac{d_{L_1}(V(L_1))}{2}$. Then it follows that S^c is such that $|S^c| \leq 2\alpha|V(L_1)| \leq 2\alpha n$ and

$$d_{L_1}(S^c) \geq \frac{d_{L_1}(V(L_1))}{2} \geq (1 + o(1))\gamma n.$$

However, $d_{Q_p^d}(X) = d|X| = e_{Q_p^d}(X, X^c) + 2e_{Q_p^d}(X)$ for any $X \subseteq V(Q^d)$. Hence, if $|X| \leq 2\alpha n$, then $\max\{e_{Q_p^d}(X, X^c), 2e_{Q_p^d}(X)\} \leq 2\alpha dn$ and so $e_{Q_p^d}(X, X^c)$ and $e_{Q_p^d}(X)$ are both stochastically dominated by $\text{Bin}(2\alpha dn, p)$. Therefore, it follows by Lemma 2.7 (ii) that

$$\mathbb{P}\left(d_{Q_p^d}(X) \geq \frac{\gamma}{2}n\right) \leq 2\left(\frac{4e\alpha(1+\epsilon)}{\gamma}\right)^{\frac{\gamma n}{2}}.$$

Hence, by the union bound, the probability that there exists any set S as above is at most

$$2\binom{n}{2\alpha n}\left(\frac{4e\alpha(1+\epsilon)}{\gamma}\right)^{\frac{\gamma n}{2}} \leq 2\left(\frac{e}{2\alpha}\right)^{\alpha n}\left(\frac{4e\alpha(1+\epsilon)}{\gamma}\right)^{\frac{\gamma n}{2}} = o(1),$$

since $\alpha \ll \epsilon$.

In particular, whp

$$\begin{aligned} \Phi(L_1) &= \min\left\{\Phi(S) : S \subseteq V(L_1) \text{ and } d_{L_1}(S) \leq \frac{d_{L_1}(V(L_1))}{2}\right\} \\ &\geq \min\{\Phi(S) : S \subseteq V(L_1) \text{ and } |S| \leq (1 - 2\alpha)n\}. \end{aligned}$$

Next, we note that it is sufficient to consider the bottleneck ratio of connected subsets of $V(L_1)$. Indeed, let S be an arbitrary subset of $V(L_1)$ with $|S| \leq (1 - 2\alpha)|V(L_1)|$ and let C_1, C_2, \dots, C_t be the connected components of $L_1[S]$. Then, for each i , $e_{L_1}(C_i, C_i^c) = e_{L_1}(C_i, S^c)$ and hence

$$e_{L_1}(S, S^c) = \sum_{i=1}^t e_{L_1}(C_i, C_i^c) \quad \text{and} \quad d_{L_1}(S) = \sum_{i=1}^t d_{L_1}(C_i).$$

It follows that

$$\Phi(S) = \frac{e_{L_1}(S, S^c)}{2d_{L_1}(S)} = \frac{\sum_{i=1}^t e_{L_1}(C_i, C_i^c)}{2\sum_{i=1}^t d_{L_1}(C_i)} \geq \min_i \left\{ \frac{e_{L_1}(C_i, C_i^c)}{2d_{L_1}(C_i)} \right\} = \min_i \{\Phi(C_i)\}.$$

Finally, let us bound the bottleneck ratio of connected subsets of $V(L_1)$. We note that, by Lemma 2.11, there is a constant C such that whp every subset $S \subseteq V(Q^d)$ which is connected in Q_p^d satisfies $|S| \leq d$, or $e_{Q_p^d}(S) \leq C|S|$. Let S be a connected subset of $V(L_1)$ with $|S| \leq (1 - 2\alpha)|V(L_1)|$.

Suppose first that $|S| \leq d$. Then, since L_1 is connected, $e_{Q_p^d}(S, S^c) \geq 1$ and hence

$$\Phi(S) = \frac{e_{L_1}(S, S^c)}{2d_{L_1}(S)} \geq \frac{1}{2d|S|} = \Omega(d^{-2}).$$

Suppose then that $d \leq |S| \leq (1 - 2\alpha)|V(L_1)|$, but $d_{Q_p^d}(S) \leq 4C|S|$. By Theorem 3.5 and (7), we have that $e_{Q_p^d}(S, S^c) \geq \beta|S|d^{-5}$ and hence

$$\Phi(S) = \frac{e_{L_1}(S, S^c)}{2d_{L_1}(S)} \geq \frac{\beta|S|d^{-5}}{8C|S|} = \Omega(d^{-5}).$$

Finally, if $d \leq |S| \leq (1 - 2\alpha)|V(L_1)|$, and hence $e_{Q_p^d}(S) \leq C|S|$, and $d_{Q_p^d}(S) \geq 4C|S|$, then it follows that

$$e_{Q_p^d}(S, S^c) = d_{Q_p^d}(S) - 2e_{Q_p^d}(S) \geq \frac{d_{Q_p^d}(S)}{2}$$

and hence

$$\Phi(S) = \frac{e_{L_1}(S, S^c)}{2d_{L_1}(S)} \geq \frac{1}{4}.$$

Hence, we can conclude that

$$\Phi(L_1) = \min\{\Phi(S) : |S| \leq (1 - 2\alpha)|V(L_1)| \text{ and } S \text{ connected}\} = \Omega(d^{-5}).$$

□

We can relate the mixing time of the lazy random walk on L_1 to its Cheeger constant using the following theorem from Levin, Peres and Wilmer [53]. This theorem is a consequence of an important theorem of Sinclair and Jerrum [61] and of Lawler and Sokal [52], which bounds the Cheeger constant in terms of the spectral gap.

THEOREM 4.2 ([53, Theorem 17.10]). *The mixing time of a lazy random walk on a graph G satisfies the inequality*

$$t_{\text{mix}} \leq \frac{2}{\Phi(G)^2} \log\left(\frac{4}{\pi_{\min}}\right).$$

Theorem 1.4 is then an immediate consequence of Lemma 4.1 and Theorem 4.2

PROOF OF THEOREM 1.4. We note that whp $|E(Q_p^d)| \leq n$ and so $\pi_{\min} \geq \frac{1}{2n}$. Hence, by Lemma 4.1 and Theorem 4.2, whp

$$t_{\text{mix}} \leq \frac{2}{\Phi(L_1)^2} \log\left(\frac{4}{\pi_{\min}}\right) = O(d^{11}).$$

□

4.2. Diameter It is immediate from Theorem 3.5 that whp the giant component of Q_p^d in the supercritical regime has diameter $O(d^6)$. However, with a more careful argument we can improve this crude estimate.

PROOF OF THEOREM 1.5. We argue as in the proof of Theorem 3.5, using the same terminology. In particular we take $q_1, q_2, L'_1, L_1, b(s), \mathcal{C}(s)$ and constants β, c_1, c_2, \dots as in the proof.

Note, by Lemma 2.2, for each $1 \leq s \leq \frac{|V(L_1)|}{2}$ each piece in $\mathcal{C}(s)$ has diameter at most $r(s) := 2c_8^{-1}b(s)^{-1}d$. Also, as before it follows from Lemma 3.2 that

$$(8) \quad \text{whp every component of } R := L_1 - L'_1 \text{ has order at most } c_4^{-1}d.$$

Furthermore, by Claim 3.6, whp for any $1 \leq t \leq 2s \leq |V(L_1)|$ and any partition of $\mathcal{C}(s)$ into two sets $\{\mathcal{C}_A, \mathcal{C}_B\}$, where $A := \bigcup \mathcal{C}_A$ and $B := \bigcup \mathcal{C}_B$, with $\min\{|A|, |B|\} = t$ there is a family of at least $c_7b(t)t$ vertex-disjoint A - B -paths of length at most five in $Q_{q_2}^d$. Let us assume that both of these likely events hold.

Let v be an arbitrary vertex in L'_1 and let $S(0) = \{v\}$. We recursively define a sequence of vertex sets $S(i)$ as follows: given $S(i)$, let $t_i := |S(i)|$ and let $S'(i)$ be the union of all pieces of $\mathcal{C}(t_i)$ which contain vertices in $S(i)$. If $t'_i := |S'(i)| \geq \min\left\{\frac{|V(L'_1)|}{2}, 2t_i\right\}$, then we let $S(i+1) = S'(i)$, otherwise by the above assumption there is a family of at least $c_7t'_ib(t'_i)$ paths of length at most five in $Q_{q_2}^d$ between $S'(i)$ and its complement in $V(L'_1)$. In this case, we let $S(i+1)$ be $S'(i)$ together with the endpoints of these paths.

We note that, since the diameter of each piece of $\mathcal{C}(t_i)$ is at most $r(t_i)$, we have that

$$(9) \quad S(i+1) \subseteq N_{Q_2}^{r(t_i)+5}(S_i).$$

Since $f(t) = tb(t)$ is an increasing function of t for $t \leq \frac{n}{2}$, it follows that, for each $i \geq 0$ with $t_i \leq \frac{|V(L'_1)|}{2}$, we have

$$(10) \quad t_{i+1} \geq \min\{t'_i + c_7t'_ib(t'_i), 2t_i\} \geq t_i(1 + c_7b(t_i)).$$

So, let us analyse the growth rate of the sequence (x_i) defined recursively as $x_0 = 1$ and $x_{i+1} = x_i \cdot (1 + c_7b(x_i))$.

We claim that

$$\text{for any } 0 < \epsilon \leq \frac{1}{2}, \text{ if } 2^{(1-2\epsilon)d} \leq x_i \leq 2^{(1-\epsilon)d}, \text{ then } x_{i+2c_7^{-1}d} \geq 2^{(1-\epsilon)d}.$$

Indeed, for any $x_j \leq 2^{(1-\epsilon)d}$ we have that $b(x_j) \geq \epsilon$ and hence

$$x_{i+2c_7^{-1}d} \geq \min\left\{2^{(1-\epsilon)d}, x_i(1 + c_7\epsilon)^{2c_7^{-1}d}\right\} \geq \min\left\{2^{(1-\epsilon)d}, 2^{(1-2\epsilon)d}e^{\epsilon d}\right\} \geq 2^{(1-\epsilon)d},$$

using that $(1 + \alpha) \geq e^{\frac{\alpha}{2}}$ for any $0 \leq \alpha \leq \frac{1}{2}$.

It follows that for any $\epsilon > 0$ there are at most $2c_7^{-1}d$ different i such that $2^{(1-2\epsilon)d} \leq t_i \leq 2^{(1-\epsilon)d}$. Note that, if $2^{(1-2\epsilon)d} \leq t_i \leq 2^{(1-\epsilon)d}$, then $r(t_i) \leq 2c_8^{-1}\epsilon^{-1}d$.

Hence, if we let $I_j = \left\{i : 2^{(1-2^{-j})d} \leq t_i \leq 2^{(1-2^{-(j+1)})d}\right\}$ for each $j = 0, 1, \dots, j_{\max}$, where j_{\max} is the minimal value such that $2^{(1-2^{-(j_{\max}+1)})d} > \frac{|V(L'_1)|}{2}$, then $|I_j| \leq 2c_7^{-1}d$ for each j and so

$$\begin{aligned} \sum_{i: t_i \leq \frac{|V(L'_1)|}{2}} (r(t_i) + 5) &\leq \sum_{j=0}^{j_{\max}} \sum_{i \in I_j} (r(t_i) + 5) \leq \sum_{j=0}^{j_{\max}} \sum_{i \in I_j} (2c_8^{-1}2^{j+1}d + 5) \\ &= O\left(d^2 \sum_{j=0}^{j_{\max}} 2^{j+1}\right) = O(d^3), \end{aligned}$$

since whp $|V(L_1)| = \Omega(n)$ and so $2^{j_{\max}} = O(d)$.

In particular, by (9) there is some constant C such that for each $v \in V(L'_1)$

$$\left|N_{Q_2}^{Cd^3}(v) \cap V(L'_1)\right| > \frac{|V(L'_1)|}{2},$$

and so the distance between any two vertices in L'_1 in Q_2 is at most $2Cd^3$.

Finally, by (8) every vertex in L_1 is at distance at most $c_4^{-1}d$ from a vertex in L'_1 and hence the diameter of L_1 is at most

$$2Cd^3 + 2c_4^{-1}d = O(d^3).$$

□

4.3. Long cycles and large minors

PROOF OF THEOREM 1.6. Let $\alpha \ll \epsilon$ and let L_1 be the largest component of Q_p^d . Then by Theorem 3.5 (b) whp every $S \subseteq V(L_1)$ such that $\alpha n \leq |S| \leq \frac{|V(L_1)|}{2}$ satisfies $|N_{Q_p^d}(S)| \geq \beta nd^{-2}(\log d)^{-1}$.

Hence, applying Theorem 2.4 with $k = \frac{|V(L_1)|}{2}$ and $t = \beta nd^{-2}(\log d)^{-1}$, we can conclude that L_1 contains a cycle of length $\Omega(nd^{-2}(\log d)^{-1})$.

□

PROOF OF THEOREM 1.7. Let $\alpha \ll \epsilon$ and let L_1 be the largest component of Q_p^d . Again by Theorem 3.5 (b) whp every $S \subseteq V(L_1)$ such that $\alpha n \leq |S| \leq \frac{|V(L_1)|}{2}$ satisfies $|N_{Q_p^d}(S)| \geq \beta nd^{-2}(\log d)^{-1}$.

If L_1 does not contain a K_t -minor, then by Theorem 2.5 there is some constant $C > 0$ such that $V(L_1)$ contains a subset X of size at most $Ct\sqrt{|V(L_1)|} \leq Ct\sqrt{n}$, such that each component of $G - X$ has order at most $\frac{2|V(L_1)|}{3}$. It follows that there is some subset $S \subseteq V(L_1)$, which is the union of some components of $G - X$, such that $\frac{|V(L_1)|}{3} \leq |S| \leq \frac{|V(L_1)|}{2}$ and $N_{Q_p^d}(S) \subseteq X$, and so by Theorem 3.5 (b)

$$\beta nd^{-2}(\log d)^{-1} \leq |N_{Q_p^d}(S)| \leq |X| \leq Ct\sqrt{n}.$$

It follows that $t \geq \frac{\beta\sqrt{n}}{Cd^2 \log d} = \Omega(\sqrt{n}d^{-2}(\log d)^{-1})$.

□

5. Discussion Theorem 1.2 gives a good bound on the likely expansion of the giant component of Q_p^d in the supercritical regime, although it is unlikely to be optimal in terms of its dependence on d . It would be interesting to determine the optimal expansion ratio.

However, perhaps this is not quite the right question to ask. As explained in the introduction, in the case of $G(d+1, p)$, whp the giant component itself is not an α -expander for any constant $\alpha > 0$, but whp contains a linear sized subgraph which is. Analogously, it seems likely that there should be a large subgraph of the giant component of Q_p^d which will have a significantly better expansion ratio than that of the giant component itself.

QUESTION 5.1. *For what $\alpha = \alpha(d)$ is it true that whp Q_p^d contains a subgraph of size $\Omega(n)$ which is an α -expander?*

As we saw in Theorem 1.3, we can take $\alpha(d) = d^{-2}(\log d)^{-1}$. We note that some inverse polynomial power of d is still necessary in the expansion ratio here.

CLAIM 5.2. *Let $\delta > 0$, let $\alpha \in (0, 1)$ and let $p = \frac{\delta}{d}$. Then there exists a $C = C(\delta, \alpha)$ such that whp there are no subsets $W \subseteq V(Q^d)$ of size at least αn such that $Q_p^d[W]$ is a $\frac{C}{d}$ -expander.*

PROOF. We first note that for any $i \in [d]$ there are $\frac{n}{2}$ edges ‘in direction i ’ in Q^d , that is, between two vertices which differ in the i th coordinate. Hence, it is a simple consequence of Lemma 2.7 that there is some $C' = C'(\delta) > 0$ such that whp for any $i \in [d]$ there are less than $\frac{C'n}{d}$ edges in direction i .

We will show that there is some $\beta > 0$ such that for any subset $W \subseteq V(Q^d)$ of size at least αn there is some $i \in [d]$ such that

$$W_i^0 := \{v \in W : v_i = 0\} \quad \text{and} \quad W_i^1 := \{v \in W : v_i = 1\}$$

both have size at least $\alpha\beta n$. By our assumption on the number of edges in direction i ,

$$\left| N(W_i^j) \right| < \frac{C'n}{d} < \frac{C' |W_i^j|}{\alpha\beta d}, \quad \text{for } j \in \{0, 1\}.$$

However, since there is some $j \in \{0, 1\}$ such that $|W_i^j| \leq \frac{|W|}{2}$, it follows that $Q_p^d[W]$ is not a $\frac{C'}{\alpha\beta d}$ -expander, and so the claim holds with $C = \frac{C'}{\alpha\beta}$.

A simple way to show the existence of β is using the language of discrete entropy. Let X be a uniformly random chosen element of W and let X_i be the projection of X to the i th coordinate. Note that $X_i \sim \text{Ber}(p_i)$ where p_i is the proportion of $v \in W$ with $v_i = 1$, i.e., $p_i = \frac{|W_i^1|}{|W|}$.

Then, by Lemma 2.6 (i), $H(X) = \log_2(|W|) \geq d + \log \alpha \geq \frac{d}{2}$ and, by Lemma 2.6 (ii),

$$H(X) \leq \sum_{i=1}^d H(X_i) = \sum_{i=1}^d h(p_i) \leq d \max_i h(p_i)$$

where $h(x) = -x \log_2 x - (1-x) \log_2(1-x)$ is the binary entropy function. In particular, it follows that $\max_i h(p_i) \geq \frac{1}{2}$. However, since h is symmetric around $\frac{1}{2}$ and increasing on $[0, \frac{1}{2}]$, we can conclude that there is some β such that $\min\{p_i, 1-p_i\} \geq \beta$. Hence

$$\left| W_i^j \right| = |W| \cdot \mathbb{P}(X_i = j) \geq \beta |W| \geq \alpha\beta n, \quad \text{for } j \in \{0, 1\},$$

as claimed. \square

As consequences of the expansion properties of the giant component, we deduced bounds on its mixing time, diameter, circumference and Hadwiger number which are almost optimal, up to some polynomial factors in d . However, it seems unlikely that any of these results are optimal in terms of their dependence in d .

For the diameter and the mixing time it is not clear what the ‘correct’ answer should be. It seems likely, but it is not immediate, that Q_p^d should have larger diameter and mixing time than Q^d does. For the diameter it might be that $O(d)$ is the correct order of growth. However, whilst the mixing time of the lazy random walk on Q^d is known to be $O(d \log d)$, see, e.g., [53], it can be shown that the largest component L_1 of Q_p^d whp contains bare paths of length $\Omega(d)$. In particular, since we expect a lazy random walk starting in the middle of such a path to take at least $\Omega(d^2)$ steps before reaching either endpoint, it follows that whp the mixing time of the lazy random walk on L_1 is $\Omega(d^2)$.

In the case of $G(d+1, p)$, Benjamini, Kozma and Wormald [4] used their description of the structure of the giant component as a decorated expander to give a $\Theta((\log d)^2)$ bound on the mixing time of the lazy random walk on the giant component in the supercritical regime. Roughly, the idea here is that the lazy random walk mixes quickly, in time $O(\log d)$, inside the expanding subgraph, but it might end up making detours of length $\Omega((\log d)^2)$ inside

the decorations. If a similar description of the giant component of Q_p^d were to hold, then we might hope that the expanding subgraph has expansion ratio α which is a small inverse power of d , and these decorations have size $O(d)$, see Lemma 3.2, in which case perhaps a more reasonable hope would be that the mixing time of the lazy random on the giant component is either dominated by the mixing time on the α -expanding subgraph, which would be of order something like α^{-2} , or by the length of the detours in the decorations, of order something like d^2 . Note that, by Claim 5.2, both of these terms must be $\Omega(d^2)$.

QUESTION 5.3. *Let $\epsilon > 0$, let $p = \frac{1+\epsilon}{d}$ and let L_1 be the largest component of Q_p^d .*

- *How large is the likely diameter of L_1 ?*
- *What is the likely mixing time of the lazy random walk on L_1 ?*

In the context of Bollobás, Kohayakawa and Łuczak's [13] question about when the diameter of the largest component of Q_p^d is superpolynomial in d , it would be interesting to know if our results can be extended partially into the weakly supercritical regime, for example when $\epsilon = o(1)$ is significantly larger than d^{-1} which was the regime considered in [12].

In terms of the circumference and the Hadwiger number of Q_p^d there are more natural conjectures to make, analogous to the case of $G(d+1, p)$, which is that whp Q_p^d contains a cycle whose length is linear in n and a complete minor of size $\Omega(\sqrt{n})$. However it seems unlikely that it is possible to prove such sharp results simply by considering the expansion properties of the giant component. In particular, we note that it may be the case that it is easier to show the likely existence of a linear length path, than that of a cycle.

QUESTION 5.4. *Let $\epsilon > 0$ and $p = \frac{1+\epsilon}{d}$.*

1. *Is it the case that whp Q_p^d contains a path of length $\Omega(n)$?*
2. *Is it the case that whp Q_p^d contains a cycle of length $\Omega(n)$?*
3. *Is it the case that whp Q_p^d contains a complete minor of order $\Omega(\sqrt{n})$?*

Furthermore, in the case of a positive answer it would also be interesting to know the dependence of the leading constants on ϵ . For example, in $G(d+1, p)$ it is known that for $p = \frac{1+\epsilon}{d}$ the giant component is of order $(2\epsilon + o(\epsilon))d$, the length of the longest cycle is of order $\Theta(\epsilon^2)d$ (see, for example, [42, Theorem 5.7]) and the size of the largest complete minor is of order $\Theta\left(\epsilon^{\frac{3}{2}}\right)\sqrt{d}$ (see [28]).

There are also some interesting open questions about the model Q_p^d in the paper of Condon, Espuny Díaz, Girão, Kühn and Osthus [19]. In particular, they used as a crucial part of their proof the fact that whp $Q_{\frac{1}{2}}^d$ contains an 'almost spanning' path, that is, a path containing $(1 - o(1))n$ vertices, and they showed that this property is in fact true for Q_p^d for any constant p .

However, analogous to the case of $G(d+1, p)$, we should perhaps expect such a path to exist for much smaller values of p . In particular, if we expect the sparse random subgraph Q_p^d with $p = \frac{c}{d}$ to contain a path of linear length $f(c)n$ for some function $f(c)$ when $c > 1$, it is natural to conjecture that $f(c) \rightarrow 1$ as $c \rightarrow \infty$.

QUESTION 5.5 ([19]). *Let $p = \omega\left(\frac{1}{d}\right)$. Is it true that whp Q_p^d contains a path of length $(1 - o(1))n$?*

Finally, other notions of random subgraphs of the hypercube Q^d have also been studied. In particular, if we let $Q^d(p)$ denote a random *induced* subgraph of Q^d , obtained by retaining

each *vertex* independently with probability p , then the typical existence of a giant component in $Q^d(p)$ when $p = \frac{1+\epsilon}{d}$ for a fixed $\epsilon > 0$ was shown by Bollobás, Kohayakawa and Łuczak [14], and this was extended to a broader range of p with $\epsilon = o(1)$ by Reidys [58]. It would be interesting to know if whp the giant component in $Q^d(p)$ also has good expansion properties. We note that random induced subgraphs of pseudo-random d -regular graphs have been studied by Diskin and Krivelevich [21], who in particular prove some likely expansion properties of the giant component in a supercritical random induced subgraph.

Acknowledgments The third author wishes to thank Asaf Nachmias for stimulating discussions and helpful remarks. The authors would also like to thank the reviewer for their careful reading of the paper and useful comments.

Funding The second author was supported by Austrian Science Fund (FWF): I3747. The third author was supported in part by USA-Israel BSF grant 2018267.

REFERENCES

- [1] AJTAI, M., KOMLÓS, J. and SZEMERÉDI, E. (1981). The longest path in a random graph. *Combinatorica* **1** 1–12.
- [2] AJTAI, M., KOMLÓS, J. and SZEMERÉDI, E. (1982). Largest random component of a k -cube. *Combinatorica* **2** 1–7.
- [3] ALON, N. and SPENCER, J. H. (2016). *The probabilistic method*, Fourth ed. Hoboken, NJ: John Wiley & Sons.
- [4] BENJAMINI, I., KOZMA, G. and WORMALD, N. (2014). The mixing time of the giant component of a random graph. *Random Structures Algorithms* **45** 383–407.
- [5] BERESTYCKI, N., LUBETZKY, E., PERES, Y. and SLY, A. (2018). Random walks on the random graph. *Ann. Probab.* **46** 456–490.
- [6] BERNSTEIN, A. J. (1967). Maximally connected arrays on the n -cube. *SIAM J. Appl. Math.* **15** 1485–1489.
- [7] BEVERIDGE, A., FRIEZE, A. and MCDIARMID, C. (1998). Random minimum length spanning trees in regular graphs. *Combinatorica* **18** 311–333.
- [8] BOLLOBÁS, B. (1983). The evolution of the cube. In *Combinatorial mathematics (Marseille-Luminy, 1981)*. *North-Holland Math. Stud.* **75** 91–97. North-Holland, Amsterdam. MR841284
- [9] BOLLOBÁS, B. (1984). The evolution of random graphs. *Trans. Am. Math. Soc.* **286** 257–274.
- [10] BOLLOBÁS, B. (1990). Complete matchings in random subgraphs of the cube. *Random Structures Algorithms* **1** 95–104. MR1068493
- [11] BOLLOBÁS, B. (2001). *Random graphs*, second ed. *Cambridge Studies in Advanced Mathematics*. Cambridge University Press. MR1864966
- [12] BOLLOBÁS, B., KOHAYAKAWA, Y. and ŁUCZAK, T. (1992). The evolution of random subgraphs of the cube. *Random Structures Algorithms* **3** 55–90. MR1139488
- [13] BOLLOBÁS, B., KOHAYAKAWA, Y. and ŁUCZAK, T. (1994). On the diameter and radius of random subgraphs of the cube. *Random Structures Algorithms* **5** 627–648. MR1300592
- [14] BOLLOBÁS, B., KOHAYAKAWA, Y. and ŁUCZAK, T. (1994). On the evolution of random Boolean functions. In *Extremal problems for finite sets (Visegrád, 1991)*. *Bolyai Soc. Math. Stud.* **3** 137–156. János Bolyai Math. Soc., Budapest. MR1319160
- [15] BOLLOBÁS, B. and RIORDAN, O. (2006). *Percolation*. Cambridge: Cambridge University Press.
- [16] BORGS, C., CHAYES, J. T., HOFSTAD, R. V. D., SLADE, G. and SPENCER, J. (2006). Random subgraphs of finite graphs. III. The phase transition for the n -cube. *Combinatorica* **26** 395–410. MR2260845
- [17] BROADBENT, S. B. and HAMMERSLEY, J. M. (1957). Percolation processes. I: Crystals and mazes. *Proc. Camb. Philos. Soc.* **53** 629–641.
- [18] CHUNG, F. and LU, L. (2001). The diameter of sparse random graphs. *Adv. Appl. Math.* **26** 257–279.
- [19] CONDON, P., ESPUNY DÍAZ, A., GIRÃO, A., KÜHN, D. and OSTHUS, D. (2020). Hamiltonicity of random subgraphs of the hypercube. *arXiv preprint arXiv:2007.02891*.
- [20] DING, J., LUBETZKY, E. and PERES, Y. (2014). Anatomy of the giant component: the strictly supercritical regime. *Eur. J. Comb.* **35** 155–168.
- [21] DISKIN, S. and KRIVELEVICH, M. (2021). Site Percolation on Pseudo-Random Graphs. *arXiv preprint arXiv:2107.13326*.

- [22] EHARD, S. and JOOS, F. (2018). Paths and cycles in random subgraphs of graphs with large minimum degree. *Electron. J. Combin.* **25** Paper No. 2.31. MR3814265
- [23] ERDE, J., KANG, M. and KRIVELEVICH, M. (2021). Large complete minors in random subgraphs. *Combin. Probab. Comput.* **30** 619–630.
- [24] ERDŐS, P. and RÉNYI, A. (1959). On random graphs. I. *Publ. Math.* **6** 290–297.
- [25] ERDŐS, P. and SPENCER, J. (1979). Evolution of the n -cube. *Comput. Math. Appl.* **5** 33–39. MR534014
- [26] FERNHOLZ, D. and RAMACHANDRAN, V. (2007). The diameter of sparse random graphs. *Random Structures Algorithms* **31** 482–516.
- [27] FOUNTOULAKIS, N., KÜHN, D. and OSTHUS, D. (2008). The order of the largest complete minor in a random graph. *Random Structures Algorithms* **33** 127–141. MR2436843
- [28] FOUNTOULAKIS, N., KÜHN, D. and OSTHUS, D. (2009). Minors in random regular graphs. *Random Structures Algorithms* **35** 444–463. MR2571779
- [29] FOUNTOULAKIS, N. and REED, B. A. (2008). The evolution of the mixing rate of a simple random walk on the giant component of a random graph. *Random Structures Algorithms* **33** 68–86.
- [30] FRIEZE, A. and KAROŃSKI, M. (2016). *Introduction to random graphs*. Cambridge University Press, Cambridge. MR3675279
- [31] FRIEZE, A. and KRIVELEVICH, M. (2013). On the non-planarity of a random subgraph. *Combin. Probab. Comput.* **22** 722–732. MR3094480
- [32] GILBERT, E. N. (1959). Random graphs. *Ann. Math. Statist.* **30** 1141–1144. MR108839
- [33] GRIMMETT, G. (1999). *Percolation*. Berlin: Springer.
- [34] HARPER, L. H. (1964). Optimal assignments of numbers to vertices. *SIAM J. Appl. Math.* **12** 131–135.
- [35] HART, S. (1976). A note on the edges of the n -cube. *Discrete Math.* **14** 157–163.
- [36] HASLEGRAVE, J., HU, J., KIM, J., LIU, H., LUAN, B. and WANG, G. (2021). Crux and long cycles in graphs. *arXiv preprint arXiv:2107.02061*.
- [37] HEYDENREICH, M. and VAN DER HOFSTAD, R. (2011). Random graph asymptotics on high-dimensional tori II: Volume, diameter and mixing time. *Probab. Theory Related Fields* **149** 397–415. MR2776620
- [38] HOFSTAD, R. v. D. and NACHMIAS, A. (2014). Unlacing hypercube percolation: A survey. *Metrika* **77** 23–50. MR3152019
- [39] HOFSTAD, R. v. D. and NACHMIAS, A. (2017). Hypercube percolation. *J. Eur. Math. Soc.* **19** 725–814. MR3612867
- [40] HOORY, S., LINIAL, N. and WIGDERSON, A. (2006). Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)* **43** 439–561. MR2247919
- [41] HULSHOF, T. and NACHMIAS, A. (2020). Slightly subcritical hypercube percolation. *Random Structures Algorithms* **56** 557–593. MR4060356
- [42] JANSON, S., ŁUCZAK, T. and RUCIŃSKI, A. (2000). *Random graphs. Wiley-Interscience Series in Discrete Mathematics and Optimization*. Wiley-Interscience. MR1782847
- [43] KAWARABAYASHI, K. and REED, B. (2010). A separator theorem in minor-closed classes. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science—FOCS 2010* 153–162. IEEE Computer Soc., Los Alamitos, CA. MR3024789
- [44] KESTEN, H. (1982). *Percolation theory for mathematicians*. Birkhäuser, Boston, MA.
- [45] KOMLÓS, J., SULYOK, M. and SZEMERÉDI, E. (1980). Second largest component in a random graph. *Studia Sci. Math. Hungar.* **15** 391–395. MR688618
- [46] KRIVELEVICH, M. (2018). Finding and using expanders in locally sparse graphs. *SIAM J. Discrete Math.* **32** 611–623. MR3769697
- [47] KRIVELEVICH, M. (2019). Expanders – how to find them, and what to find in them. In *Surveys in combinatorics 2019. London Math. Soc. Lecture Note Ser.* **456** 115–142. Cambridge Univ. Press, Cambridge. MR3967294
- [48] KRIVELEVICH, M. (2019). Long cycles in locally expanding graphs, with applications. *Combinatorica* **39** 135–151. MR3936195
- [49] KRIVELEVICH, M. and NACHMIAS, A. (2006). Coloring complete bipartite graphs from random lists. *Random Structures Algorithms* **29** 436–449. MR2268230
- [50] KRIVELEVICH, M. and SAMOTIJ, W. (2014). Long paths and cycles in random subgraphs of \mathcal{H} -free graphs. *Electron. J. Combin.* **21** Paper 1.30. MR3177525
- [51] KRIVELEVICH, M. and SUDAKOV, B. (2013). The phase transition in random graphs: A simple proof. *Random Structures Algorithms* **43** 131–138. MR3085765
- [52] LAWLER, G. F. and SOKAL, A. D. (1988). Bounds on the L^2 spectrum for Markov chains and Markov processes: A generalization of Cheeger’s inequality. *Trans. Am. Math. Soc.* **309** 557–580.
- [53] LEVIN, D. A., PERES, Y. and WILMER, E. L. (2017). *Markov chains and mixing times*. Providence, RI: American Mathematical Society.

- [54] LINDSEY, J. H. (1964). Assignment of numbers to vertices. *Amer. Math. Monthly* **71** 508-516.
- [55] ŁUCZAK, T. (1990). Component behavior near the critical point of the random graph process. *Random Structures Algorithms* **1** 287–310.
- [56] MCDIARMID, C., SCOTT, A. and WITHERS, P. (2021). The component structure of dense random subgraphs of the hypercube. *Random Structures Algorithms* **59** 3-24.
- [57] PETE, G. (2008). A note on percolation on \mathbb{Z}^d : Isoperimetric profile via exponential cluster repulsion. *Electron. Commun. Probab.* **13** 377–392.
- [58] REIDYS, C. M. (2009). Large components in random induced subgraphs of n -cubes. *Discrete Math.* **309** 3113–3124. MR2526729
- [59] RIORDAN, O. and WORMALD, N. (2010). The diameter of sparse random graphs. *Comb. Probab. Comput.* **19** 835–926.
- [60] SAPOŽENKO, A. A. (1967). Metric properties of almost all functions of the algebra of logic. *Diskret. Analiz* 10 91–119. MR223280
- [61] SINCLAIR, A. and JERRUM, M. (1989). Approximate counting, uniform generation and rapidly mixing Markov chains. *Inf. Comput.* **82** 93–133.