Asymptotics in percolation on high-girth expanders

Michael Krivelevich[*]   Eyal Lubetzky[†]   Benny Sudakov[‡]

Abstract

We consider supercritical bond percolation on a family of high-girth d-regular expanders. Alon, Benjamini and Stacey (2004) established that its critical probability for the appearance of a linear-sized ("giant") component is $p_c = 1/(d-1)$. Our main result recovers the sharp asymptotics of the size and degree distribution of the vertices in the giant and its 2-core at any $p > p_c$. It was further shown in [1] that the second largest component, at any $0 < p < 1$, has size at most n^ω for some $\omega < 1$. We show that, unlike the situation in the classical Erdős–Rényi random graph, the second largest component in bond percolation on a regular expander, even with an arbitrarily large girth, can have size $n^{\omega'}$ for ω' arbitrarily close to 1. Moreover, as a by-product of that construction, we answer negatively a question of Benjamini (2013) on the relation between the diameter of a component in percolation on expanders and the existence of a giant component. Finally, we establish other typical features of the giant component, e.g., the existence of a linear path.

1   Introduction

A graph $\mathcal{G}$ is called a (b, d)-expander if its maximum degree is d, and for every subset $S \subset V(\mathcal{G})$ of at most $|V(\mathcal{G})|/2$ vertices there are at least $b|S|$ edges in the cut between S and $V(\mathcal{G}) \setminus S$. For $0 < p < 1$ and a given graph $\mathcal{G}$, let $\mathcal{G}_p$ denote the corresponding bond percolation on $\mathcal{G}$, i.e., the distribution over spanning subgraphs of $\mathcal{G}$ where each edge is present, independently, with probability p. Our primary focus will be $\mathcal{G}_p$ for a $(d\text{-})$regular (b, d)-expander $\mathcal{G}$ whose girth (length of the shortest cycle) is large.

The pioneering paper of Alon, Benjamini and Stacey [1] showed that if $\mathcal{G}$ is an expander and $p > 0$ is fixed, there is at most a single linear component in $\mathcal{G}_p$ with high probability (w.h.p.). Moreover, when $\mathcal{G}$ is d-regular and its girth tends to ∞, the authors of [1] identified the critical percolation probability to be $p_c = 1/(d-1)$, showing that, for every fixed $p > p_c$, w.h.p. there exists a linear ("giant") component in $\mathcal{G}_p$ (see the formal statement below). The latter result was thereafter extended by Peres et al. [4] to any family of sparse graphs that converge as $n \to \infty$ in the Benjamini–Schramm sense.

More formally, write $\mathcal{C}_i(G)$ (for $i = 1, 2, \ldots$) for the connected components of G in decreasing order of their sizes. (By a slight abuse of notation, we also let $\mathcal{C}_v(G)$ (for $v \in V(G)$) denote the connected

---

[*]School of Mathematical Sciences, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 6997801, Israel. Email: `krivelev@post.tau.ac.il`. Research supported in part by BSF grant 2014361 and ISF grant 1261/17.

[†]Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA. Email: `eyal@courant.nyu.edu`. Research supported in part by NSF grant DMS-1513403 and BSF grant 2014361.

[‡]Department of Mathematics, ETH, Ramistrasse 101, 8092 Zurich, Switzerland. Email: `benjamin.sudakov@math.ethz.ch`. Research supported in part by SNSF grant 200021-175573.

component of v in G.) It was shown in [1] that if $\mathcal{G}$ is a regular (b,d)-expander on n vertices (more precisely, a sequence of such graphs) with girth tending to ∞ with n, then for every fixed $p > 1/(d-1)$,

$$\text{for } \textit{some} \text{ fixed } c = c(p,d) > 0: \quad \lim_{n \to \infty} \mathbb{P}\left(|\mathcal{C}_1(\mathcal{G}_p)| > cn\right) = 1,$$

whereas for every fixed $p < 1/(d-1)$,

$$\text{for } \textit{every} \text{ fixed } c > 0: \quad \lim_{n \to \infty} \mathbb{P}\left(|\mathcal{C}_1(\mathcal{G}_p)| > cn\right) = 0.$$

That the lack of a linear component w.h.p. at $p < 1/(d-1)$ extends also to $p = 1/(d-1)$ follows from the work of Nachmias and Peres [13] (who proved, more generally, that $\mathbb{P}(|\mathcal{C}_1| \leq An^{2/3}) \to 0$ as $A \to \infty$ in percolation with parameter $p = 1/(\Delta - 1)$ on any family of graphs with maximal degree Δ).

Pittel [14] refined the results of Alon et al. in the special case where $\mathcal{G}$ is a *random* (uniformly chosen) d-regular graph on n vertices for $d \geq 3$ fixed — well-known to be a expander w.h.p. — showing that the phase transition of $|\mathcal{C}_1|$ mirrors that in the Erdős–Rényi graph $\mathcal{G}(n,p)$: w.h.p., $|\mathcal{C}_1| \leq C(p) \log n$ at $p < p_c = \frac{1}{d-1}$ vs. $|\mathcal{C}_1| \sim \theta_1 n$ at $p > p_c$ for an explicit $\theta_1(p)$, whereas $|\mathcal{C}_1| = n^{2/3+o(1)}$ at the critical p_c (the precise order of $n^{2/3}$ and correct scaling of the critical window were subsequently found in [13]).

In this work we obtain the asymptotic size (lower bounded by $c(p,d)n$ above) and degree distribution of the giant component $\mathcal{C}_1(G)$ for any high-girth expander $G \sim \mathcal{G}_p$ at $p > 1/(d-1)$, as well as its 2-core.

For $d \geq 3$ and $1 < \lambda < d - 1$, let

$$p = \lambda/(d-1)$$

and let $q = q(\lambda, d)$ be the unique solution in $[0,1)$ of the equation

$$q = (1 - p + pq)^{d-1}, \tag{1.1}$$

well known (cf. [2]) to coincide with the extinction probability of a $\text{Bin}(d-1, p)$-Galton–Watson tree.

Recall that the 2-core of a connected component $\mathcal{C}$, denoted here $\mathcal{C}^{(2)}$, is its maximum induced subgraph with minimum degree at least 2, and the (tree) excess of $\mathcal{C}$ is the minimum number of edges that need to be removed from $\mathcal{C}$ to turn it into a tree (note that the excess of $\mathcal{C}$ equals that of $\mathcal{C}^{(2)}$). Our first main result establishes the asymptotic number of vertices and edges in the largest (giant) component in supercritical bond percolation on a high girth d-regular expander, as well as its 2-core, and consequently their asymptotic excess.

Theorem 1 (the giant). *Fix $d \geq 3$ and $\frac{1}{d-1} < p < 1$, and letting q be as in* (1.1), *define*

$$\theta_1 := 1 - q(1-p) - pq^2, \qquad\qquad \eta_1 := \tfrac{1}{2}pd(1 - q^2), \tag{1.2}$$
$$\theta_2 := 1 - q - (d-1)pq(1-q), \qquad \eta_2 := \tfrac{1}{2}pd(1-q)^2. \tag{1.3}$$

For every $\varepsilon > 0$ and $b > 0$ there exist some $c, C, R > 0$ such that, if $\mathcal{G}$ is a regular (b,d)-expander on n vertices with girth at least R, then with probability at least $1 - Ce^{-cn}$, the random graph $G \sim \mathcal{G}_p$ has

$$\left|\tfrac{1}{n}|V(\mathcal{C}_1)| - \theta_1\right| < \varepsilon, \qquad \left|\tfrac{1}{n}|E(\mathcal{C}_1)| - \eta_1\right| < \varepsilon, \tag{1.4}$$
$$\left|\tfrac{1}{n}|V(\mathcal{C}_1^{(2)})| - \theta_2\right| < \varepsilon, \qquad \left|\tfrac{1}{n}|E(\mathcal{C}_1^{(2)})| - \eta_2\right| < \varepsilon. \tag{1.5}$$

In particular, with probability at least $1 - C\exp(-cn)$, the excess of $\mathcal{C}_1$ is within $2\varepsilon n$ of $(\eta_1 - \theta_1)n$, and the excess of $\mathcal{C}_1^{(2)}$ is within $2\varepsilon n$ of $(\eta_2 - \theta_2)n$.
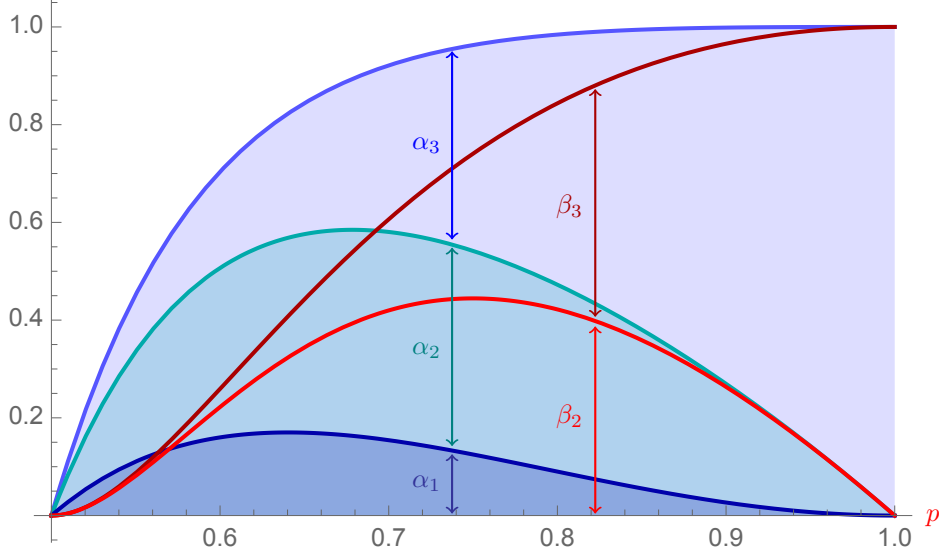
2

Figure 1: Asymptotic degree distributions of the giant component ($\sim \alpha_k n$ degree-k vertices (blue)) and its 2-core ($\sim \beta_k n$ degree-k vertices (red)) in supercritical percolation on a high girth 3-regular expander.

Remark 1.1 (degree distributions of the giant and the 2-core). In the setting of Theorem 1, we in fact asymptotically obtain the entire degree distributions of the giant component and 2-core (Theorem 2.1). For instance, for a fixed $b > 0$, consider a sequence $\mathcal{G}^{(n)}$ of regular $(b, 3)$-expanders on n vertices whose girth tends to ∞ (arbitrarily slowly) with n. For $\frac{1}{2} < p < 1$, let

$$\alpha_1 = \frac{3}{p}(1-p)^2(2p-1)\,, \quad \alpha_2 = \frac{3}{p^2}(1-p)(1-4p+6p^2-4p^3)\,, \qquad \alpha_3 = p^3\left(1 - \left(\tfrac{1-p}{p}\right)^6\right),$$

$$\beta_2 = \frac{3}{p^3}(1-2p)^2(1-p)\,, \qquad\qquad\qquad \beta_3 = \left(\frac{2p-1}{p}\right)^3.$$

Then w.h.p., the giant component $\mathcal{C}_1$ of $G \sim \mathcal{G}_p^{(n)}$ has $(\alpha_k + o(1))n$ vertices of degree k for each $k \in \{1, 2, 3\}$, and its 2-core $\mathcal{C}_1^{(2)}$ has $(\beta_k + o(1))n$ vertices of degree k for each $k \in \{2, 3\}$; see Figure 1.

Remark 1.2 (Limits for large d). For $p = (1 + \xi)/(d - 1)$ with $\xi > 0$ and $d \to \infty$ with n, one has

$$q \to 1 - 2\xi + \tfrac{8}{3}\xi^2 - \tfrac{28}{9}\xi^3 + O(\xi^4)$$

and

$$\theta_1 \to 2\xi - \tfrac{8}{3}\xi^2 + \tfrac{28}{9}\xi^3 + O(\xi^4)\,, \quad \eta_1 \to 2\xi - \tfrac{8}{3}\xi^2 + \tfrac{34}{9}\xi^3 + O(\xi^4)\,, \quad \eta_1 - \theta_1 \to \tfrac{2}{3}\xi^3 + O(\xi^4)\,,$$

$$\theta_2 \to 2\xi^2 - 4\xi^3 + O(\xi^4)\,, \qquad \eta_2 \to 2\xi^2 - \tfrac{10}{3}\xi^3 + O(\xi^4)\,, \qquad \eta_2 - \theta_2 \to \tfrac{2}{3}\xi^3 + O(\xi^4)\,.$$

Compare this to $\mathcal{G}(n, p)$ for $p = \frac{1+\xi}{n}$ with $\xi = o(1)$: there $|V(\mathcal{C}_1)| \sim 2\xi n$, $|E(\mathcal{C}_1)| \sim 2\xi n$, $|V(\mathcal{C}_1^{(2)})| \sim 2\xi^2 n$ and the excess is $(1 + o(1))\tfrac{2}{3}\xi^3 n$ (see, e.g., [6, 10, 15] and the structure theorems in [7, 8]).

3

In the context of the second largest component of $\mathcal{G}_p$, it was shown by Alon et al. [1] that, if $\mathcal{G}$ is a (b,d)-expander (not necessarily regular) on n vertices then there exists $\omega = \omega(b,d) < 1$ such that, w.h.p., for every sequence $0 < p_n < 1$,

$$\lim_{n\to\infty} \mathbb{P}(|\mathcal{C}_2(\mathcal{G}_{p_n})| < n^\omega) = 1\,.$$

Indeed, it is well-known (cf., e.g., [6]) that in the Erdős–Rényi random graph $\mathcal{G}(n,p)$ (where $\mathcal{G}$ is the complete graph) such is the case, as the supercritical regime $p > 1/n$ admits a single giant component w.h.p., and all other components are logarithmic in size (see, e.g., [6] and [10]). Such is also the case in percolation on random regular graphs (cf. [13]). In light of this — and in line with results of Pittel [14] on $|\mathcal{C}_2|$ in random regular graphs — one may believe that the same holds for percolation on expanders, whereby the n^ω above could be replaced by some $C(b,d) \log n$ with probability arbitrarily close to 1.

Perhaps surprisingly, it turns out that even on a family of regular expanders with arbitrarily large girth, the above polynomial bound of n^ω is essentially best possible, as the second largest component can have size $n^{\omega'}$ with ω' arbitrarily close to 1.

Theorem 2 (second component). *For every $d \geq 3$, $R \geq 1$, $p \in (\frac{1}{d-1}, 1)$ and $\alpha \in (0,1)$ there exist $b > 0$ and a regular (b,d)-expander $\mathcal{G}$ on n vertices with girth at least R where $G \sim \mathcal{G}_p$ has $|V(\mathcal{C}_2)| \gtrsim n^\alpha$ w.h.p.*

Remark 1.3. Using essentially the same construction, for any sequence $0 < \alpha_1 \leq \alpha_2 \leq \ldots \leq \alpha_k < 1$ with k fixed, one can construct an expander $\mathcal{G}$ such that w.h.p. $G \sim \mathcal{G}_p$ has components whose sizes have orders $n^{\alpha_1}, \ldots, n^{\alpha_k}$, respectively (in addition to the linearly sized component).

Towards a construction of an infinite graph that disproves the existence of "expanders at all scales," Benjamini asked the following question (a positive answer to which would be a step in said construction), on the relation between the diameter of the connected component of v in $\mathcal{G}_p$ to the existence of a giant.

Question (Benjamini [3, Q. 5.5]). *Let $\mathcal{G}$ be a bounded degree expander. Further assume that there is a fixed vertex $v \in \mathcal{G}$, so that $G \sim \mathcal{G}_{1/2}$ satisfies*

$$\mathbb{P}\big(\operatorname{diam}(\mathcal{C}_v(G)) > \tfrac{1}{2}\operatorname{diam}(\mathcal{G})\big) > \tfrac{1}{2}\,.$$

Is there a giant component w.h.p.?

A variant of the construction in Theorem 2 gives a negative answer to this question.

Theorem 3. *For every $\varepsilon > 0$ and $0 < p < 1$ there exist $b, d, \delta > 0$ and, for infinitely many values of n, a (b,d)-expander $\mathcal{G}$ on n vertices with a prescribed vertex v, such that the graph $G \sim \mathcal{G}_p$ satisfies*

$$\mathbb{P}\big(\operatorname{diam}(\mathcal{C}_v(G)) \geq (1-\varepsilon)\operatorname{diam}(\mathcal{G})\big) \geq 1 - \varepsilon\,,$$

and yet there are no components of size larger than $n^{1-\delta}$ in G w.h.p.

Finally, in §4, we explore additional typical properties of percolation on high-girth expanders (where we find the typical order of those random variables but not their precise asymptotics): the existence of linear simple paths (Corollary 4.2), the fact that linearly many vertices must be removed in order to disconnect the giant component (Theorem 4.1), and the existence of complete minors (Corollary 4.4).

Main techniques.

The starting point in proving Theorem 1, and furthermore its more detailed version Theorem 2.1 which gives the typical degree profile in the giant and 2-core of $\mathcal{G}_p$, is a refinement of the elegant sprinkling argument of Alon et al. [1]. While that argument was used in [1] and later also in [4] to show that w.h.p. a *fixed proportion* of the vertices whose local neighborhoods in $\mathcal{G}_{p-\varepsilon}$ are "large" will belong to the giant after adding $\mathcal{G}_\varepsilon$, in fact *all but a negligible proportion* of such vertices will be in the giant. Namely, the local property of whether the connected component of a vertex x has size at least R turns out to be a predictor, for all but $\varepsilon(R)n$ vertices, as to whether $x \in \mathcal{C}_1(\mathcal{G}_p)$ (similarly for edges) — see Proposition 2.2. One can then readily use this local predictor to read off the degree profile, and more generally, the entire local geometry in $\mathcal{C}_1(\mathcal{G}_p)$ for a high girth expander $\mathcal{G}$, as demonstrated next.

Example 1.4. [Density of local neighborhoods in $\mathcal{C}_1$] Let $G \sim \mathcal{G}_p$ for $\frac{1}{d-1} < p < 1$ and a d-regular expander $\mathcal{G}$ on n vertices whose girth tends to ∞ with n. Set q as in (1.1), let T be a rooted tree with k levels and ℓ leaves (for some fixed k, ℓ), and let $\alpha_T = \mathbb{P}(\mathcal{T}_p \simeq T)$, where $\mathcal{T}$ is the tree on k levels that is d-regular except at its leaves (and $\simeq$ denotes graph isomorphism). The number of vertices in $\mathcal{C}_1(G)$ whose k-radius neighborhood is isomorphic to T is then w.h.p. asymptotically $\left(1 - q^\ell\right) \alpha_T n$.

Example 1.5. [Asymptotic density of paths in $\mathcal{C}_1$] For G as in Example 1.4 and for any fixed $\ell \geq 1$, the number of paths with ℓ edges in $\mathcal{C}_1$ is w.h.p. asymptotically $\frac{1}{2}d(d-1)^{\ell-1}p^\ell\left(1-q^{\ell+1}(1-p+pq)^{1-\ell}\right)n$.

The analysis of cycles in G (and its 2-core) is substantially more delicate. Naively, one might expect our results to suggest that, if a certain edge xy in $G \sim \mathcal{G}_p$ is such that in the graph $H_{xy} := G \setminus \{xy\}$ the components of x and y are both "large," then both should typically belong to the giant in H (which is a subset of the giant of G), hence the edge xy should lie on a cycle in $\mathcal{C}_1(G)$. However, turning this intuition into a rigorous argument is problematic in light of the fact that we cannot actually delete xy (as we aim to carry this analysis simultaneously for all edges xy), and our mechanism of securing that the components of x and y would belong to the giant was non-constructive (arguing that there are small cuts whence sprinkling would patch most such components together); in particular, that argument gave no control over whether or not the edge xy itself is a bridge as opposed to a cycle edge in $\mathcal{C}_1^{(2)}(G)$.

To remedy this, we introduce the notion of a *k-thick* set — roughly put, a set that can be covered by disjoint connected components of size at least k each (see Definition 2.7) — and show, in what may be of independent interest, that if G is an expander and H is a slightly percolated subgraph of G (that is, $H \sim G_{1-\delta}$ for some small $\delta > 0$), then every linearly-sized k-thick set expands in H (cf. Claim 2.8). Since we seek to analyze the effect of sprinkling on components that are "large," such expansion suffices, and does in fact hold in H, supporting the above framework of the proof.

Finally, our constructions in Theorem 2 and 3 exploit the source of some of the obstacles described above: while we have full understanding of the geometry of G in the microscopic scale (locally the graph is a regular tree up to an arbitrarily large radius), and some control over it in the macroscopic scale (the expansion implies bounds on the number of edges between every two linearly-sized sets), we have limited control over it in the intermediate scales. For instance, if the k-radius ball around a vertex x grows as $(d-1)^k$ then percolation is supercritical, as opposed to growing only as b^k whence it might be subcritical (yet still satisfying the expansion property). Our constructions combine both these types of expansion in the mesoscopic scales to create components of size $\Theta(n^\alpha)$ $(0 < \alpha < 1)$ in $\mathcal{G}_p$.

5

2 Asymptotics of the giant component and its 2-core

The goal in this section is to prove the following result (from which Theorem 1 follows) describing the typical degree distribution of the giant component and its 2-core in $\mathcal{G}_p$ up to an error of $o(n)$.

Theorem 2.1. *Fix $d \geq 3$ and $1 < \lambda < d-1$, let $p = \lambda/(d-1)$, and with q as in (1.1), define*

$$\alpha_k = \binom{d}{k} p^k (1-p)^{d-k} (1-q^k) \qquad\qquad (k = 1, \ldots, d), \qquad\qquad (2.1)$$

$$\beta_k = \binom{d}{k} \frac{p^k (1-q)^k q}{(1-p+pq)^{k-1}} \qquad\qquad (k = 2, \ldots, d). \qquad\qquad (2.2)$$

For a given d-regular graph $\mathcal{G}$, let D_k ($1 \leq k \leq d$) be the number of degree-k vertices in $\mathcal{C}_1$, the largest component of the graph $G \sim \mathcal{G}_p$, and let D_k^* denote the number of degree-k vertices in its 2-core $\mathcal{C}_1^{(2)}$. Then for every $\varepsilon > 0$ and $b > 0$ there exist some $c, R > 0$ so that, if $\mathcal{G}$ is a regular (b,d)-expander on n vertices with girth at least R, then with probability at least $1 - \exp(-cn)$,

$$\left| D_k/n - \alpha_k \right| < \varepsilon \quad \text{for all } 1 \leq k \leq d \qquad \text{and} \qquad \left| D_k^*/n - \beta_k \right| < \varepsilon \quad \text{for all } 2 \leq k \leq d .$$

Indeed, Theorem 1 follows from verifying (recalling (1.1)) that the quantities θ_i and η_i ($i = 1, 2$), as defined in (1.2)–(1.3), satisfy the following w.r.t. the above defined α_k ($1 \leq k \leq d$) and β_k ($2 \leq k \leq d$):

$$\theta_1 = \sum_{k=1}^d \alpha_k , \quad \eta_1 = \tfrac{1}{2} \sum_{k=1}^d k\alpha_k , \quad \theta_2 = \sum_{k=2}^d \beta_k , \quad \eta_2 = \tfrac{1}{2} \sum_{k=2}^d k\beta_k .$$

2.1 The giant component

Fix $\varepsilon > 0$ small enough so that $p' := p - \varepsilon$ satisfies $p' > 1/(d-1)$. There exists R_0 so that, for all $R \geq R_0$, the probability of survival to depth R in a Galton–Watson (GW) tree with offspring distribution $\mathrm{Bin}(d-1, p')$ is at least $1 - q'$, and the corresponding probability in a GW-tree with offspring distribution $\mathrm{Bin}(d-1, p)$ is at most $1 - q + \varepsilon$. Since $\lambda \mapsto q(\lambda)$ is continuous, one has $q' \uparrow q$ as $\varepsilon \downarrow 0$. We will couple $G' \sim \mathcal{G}_{p'}$ and $G \sim \mathcal{G}_p$ by letting $E(G) = E(G') \cup E(F')$ for $F' \sim \mathcal{G}_{\varepsilon'}$, where

$$\varepsilon' := \frac{p - p'}{1 - p'} = \frac{\varepsilon}{1 - p'} .$$

For a graph H and an ordered pair of vertices $x, y \in V(H)$, define

$$\mathcal{A}_{x,y}^R = \mathcal{A}_{x,y}^R(H) = \{|\mathcal{C}_y(H \setminus \{xy\})| \geq R\} \qquad\qquad (2.3)$$

(where $H \setminus \{xy\}$ is obtained by deleting the edge xy if present), and for an unordered pair $x, y \in V(H)$ let

$$\mathcal{B}_{xy}^R = \mathcal{B}_{xy}^R(H) = \left(\mathcal{A}_{x,y}^R \cup \mathcal{A}_{y,x}^R \right) \cap \{xy \in E(H)\} ; \qquad\qquad (2.4)$$

i.e., $\mathcal{A}_{x,y}^R$ says that after removing xy (if present) the component of y has at least R vertices, while $\mathcal{B}_{xy}^R$ says that, in addition to this, after removing xy the component of x also has at least R vertices, and that the edge xy does belong to H. Further define, for every vertex $x \in V(H)$,

$$\mathcal{B}_x^R = \bigcup_y \left(\mathcal{A}_{x,y}^R \cap \{xy \in E(H)\} \right) . \qquad\qquad (2.5)$$

Finally, let

$$E_1(H) = \{xy \in E(H) \ : \ \mathcal{B}_{xy}^R(H) \text{ holds}\},$$
$$V_1(H) = \{x \in V(H) \ : \ \mathcal{B}_x^R(H) \text{ holds}\}. \tag{2.6}$$

The main result we wish to prove in this subsection is as follows.

Proposition 2.2. *For every $\varepsilon, b > 0$ there exist R and $c > 0$ such that, if $\mathcal{G}$ is a regular (b, d)-expander on n vertices with girth greater than $2R$, and $G \sim \mathcal{G}_p$, then*

$$\mathbb{P}\Big(\big|E_1(G) \triangle E(\mathcal{C}_1(G))\big| > \varepsilon n\Big) \leq \exp(-cn), \tag{2.7}$$

$$\mathbb{P}\Big(\big|V_1(G) \triangle V(\mathcal{C}_1(G))\big| > \varepsilon n\Big) \leq \exp(-cn). \tag{2.8}$$

Proof. Observe that, for any graph H with maximum degree d, if $|\mathcal{C}_x| > dR$ then the vertex x must be incident to some y such that $\mathcal{A}_{x,y}^R$ holds; similarly, if an edge xy belongs to a component of size at least $2R$ then at least one of the events $\mathcal{A}_{x,y}^R, \mathcal{A}_{y,x}^R$ must hold. That is,

$$\bigcup\{E(\mathcal{C}) \ : \ \mathcal{C} \text{ is a connected component of } H \text{ with } |\mathcal{C}| \geq 2R\} \subseteq E_1(H).$$
$$\bigcup\{V(\mathcal{C}) \ : \ \mathcal{C} \text{ is a connected component of } H \text{ with } |\mathcal{C}| > dR\} \subseteq V_1(H). \tag{2.9}$$

By the assumptions that $R \geq R_0$ for a large enough R_0 and that the girth is greater than $2R$,

$$1 - q \leq \mathbb{P}(\mathcal{A}_{x,y}^R(G)) \leq 1 - q + \varepsilon, \qquad 1 - q' \leq \mathbb{P}(\mathcal{A}_{x,y}^R(G')) \leq 1 - q' + \varepsilon.$$

When $H \sim \mathcal{G}_p$ or $H \sim \mathcal{G}_{p'}$, the standard edge-exposure martingale (see, e.g., [10, Sec. 2.4]) — noting that adding/deleting an edge influences at most $2(d-1)^R$ edges — implies via Hoeffding–Azuma that

$$\mathbb{P}\left(\Big|\big|E_1(H)\big| - \mathbb{E}[|E_1(H)|]\Big| \geq a\right) \leq \exp\left(-\frac{a^2}{4dn(d-1)^{2R}}\right). \tag{2.10}$$

and similarly,

$$\mathbb{P}\left(\Big|\big|V_1(H)\big| - \mathbb{E}[|V_1(H)|]\Big| \geq a\right) \leq \exp\left(-\frac{a^2}{4dn(d-1)^{2R}}\right). \tag{2.11}$$

Note that, in $G \sim \mathcal{G}_p$, since the girth of $\mathcal{G}$ is greater than $2R$ (hence $\mathcal{A}_{x,y}^R$ and $\mathcal{A}_{y,x}^R$ are independent),

$$p(1 - q^2) \leq \mathbb{P}(\mathcal{B}_{xy}^R) \leq p(1 - (q - \varepsilon)^2),$$

and, using that $q = (1 - p + pq)^{d-1}$ and the mutual independence of $\mathcal{A}_{x,y}^R$ for all neighbors y of x in $\mathcal{G}$,

$$1 - q(1 - p + pq) \leq \mathbb{P}(\mathcal{B}_x^R) \leq 1 - (1 - p + p(q - \varepsilon))^d.$$

We can therefore deduce that

$$\tfrac{1}{2}pd(1 - q^2)n \leq \mathbb{E}[|E_1(G)|] \leq \tfrac{1}{2}pd(1 - (q - \varepsilon)^2)n, \tag{2.12}$$

$$\big(1 - q(1 - p + pq)\big)n \leq \mathbb{E}[|V_1(G)|] \leq \big(1 - (1 - p + p(q - \varepsilon))^d\big)n. \tag{2.13}$$

7

Combining this with (2.9)–(2.10) implies that, in $G \sim \mathcal{G}_p$, with probability at least $1 - \exp(-cn)$,

$$\sum_{\mathcal{C}:|V(\mathcal{C})|\geq 2R} |E(\mathcal{C})| \leq |E_1(G)| \leq \left(\tfrac{1}{2}pd(1 - (q - \varepsilon)^2) + \varepsilon\right) n \,, \tag{2.14}$$

$$\sum_{\mathcal{C}:|V(\mathcal{C})|>dR} |V(\mathcal{C})| \leq |V_1(G)| \leq \left(1 - (1 - p + p(q - \varepsilon))^d + \varepsilon\right) n \,. \tag{2.15}$$

(In particular, this gives upper bounds on $|E(\mathcal{C}_1(G))|$ and on $|V(\mathcal{C}_1(G))|$.)

Next, we consider G', and note that using (2.10)–(2.11), together with the analogs of (2.12)–(2.13) for G', yields

$$\mathbb{P}\left(|E_1(G')| < (\tfrac{1}{2}p'd(1 - q'^2) - \varepsilon)n\right) \leq \exp(-cn) \,, \tag{2.16}$$

$$\mathbb{P}\left(|V_1(G')| < (1 - q'(1 - p' + p'q') - \varepsilon)n\right) \leq \exp(-cn) \,. \tag{2.17}$$

Claim 2.3. *For every $\varepsilon, b, d > 0$ there exist $c, R > 0$ such that the following holds for large enough n. If $\mathcal{G}$ is a regular (b, d)-expander with n vertices, and $\mathcal{S}$ is a collection of disjoint vertex subsets of $\mathcal{G}$, each of size at least R, then the probability that there exist two subsets $\mathcal{S}_1, \mathcal{S}_2 \subset \mathcal{S}$, with a total of at least εn vertices in each, and no path between them in $H \sim \mathcal{G}_\varepsilon$, is at most $\exp(-cn)$.*

Proof. By Menger's Theorem and our hypothesis on the edge expansion of $\mathcal{G}$, for every two disjoint subsets $A, B \subset V(\mathcal{G})$ of size at least εn each, there are at least $b\varepsilon n$ edge-disjoint paths between them in $\mathcal{G}$. Since the total number of edges in $\mathcal{G}$ is $dn/2$, it follows that for every two such subsets A, B,

there are at least $\lceil \tfrac{b\varepsilon}{2}n \rceil$ edge-disjoint paths of length at most $\lfloor \tfrac{d}{b\varepsilon} \rfloor$ between A, B in $\mathcal{G}$. (2.18)

In particular, this holds for every two subsets $\mathcal{S}_1$ and $\mathcal{S}_2$ of $\mathcal{S}$ each containing at least εn vertices. The probability that none of these short paths between $\mathcal{S}_1$ and $\mathcal{S}_2$ survive in $\mathcal{G}_\varepsilon$ is at most

$$\left(1 - \varepsilon^{d/(b\varepsilon)}\right)^{\frac{1}{2}b\varepsilon n} \leq \exp\left[-\tfrac{1}{2}b\varepsilon^{1+d/b\varepsilon}n\right] \,.$$

Altogether, a union bound over at most $2^{2n/R}$ possible pairs of subsets of $\mathcal{S}$ shows that the probability that there exist $\mathcal{S}_1, \mathcal{S}_2$ violating the statement of the claim is at most

$$\exp\left[\left(R^{-1}2\log 2 - \tfrac{1}{2}b\varepsilon^{1+d/b\varepsilon}\right)n\right] \,.$$

Taking R large enough completes the proof. $\blacksquare$

Corollary 2.4. *For every $\varepsilon, b, d > 0$ there exist $c, R > 0$ so that the following holds for large enough n. If $\mathcal{G}$ is a regular (b, d)-expander on n vertices with girth greater than $2R$, then with probability $1 - O(e^{-cn})$ there exists a connected component $\mathcal{C}$ of $G' \cup F'$ containing all but at most $2\varepsilon n$ vertices of the set*

$$\overline{V}_1(G') := \left\{y \in V(G) \,:\, \mathcal{A}^R_{x,y}(G') \text{ holds for some } x \in V(G)\right\} \,.$$

In particular, $\mathcal{C}$ contains all but at most $2\varepsilon n$ of the vertices $V_1(G')$.

8

Proof. Let $\mathcal{S}(G') = \{\mathcal{C}_y(G') : y \in \overline{V}_1(G')\}$, noting that every component in $\mathcal{S}(G')$ is of size at least R by definition. By Claim 2.3, with probability at least $1 - \exp(-cn)$, one cannot partition $\mathcal{S}(G')$ into two subsets $\mathcal{S}_1, \mathcal{S}_2$, each of size at least εn, such that there will be no path connecting them in $G = G' \cup F'$. With this in mind, let U be a set of vertices obtained by collecting (in an arbitrary way) connected components of $\mathcal{S}(G')$ in G, until $|U \cap \overline{V}_1(G')| \geq \varepsilon n$. Note that $|U \cap \overline{V}_1(G')| < \varepsilon n + |\mathcal{C} \cap \overline{V}_1(G')|$, where $\mathcal{C}$ is a connected component in G (the last one that joined U). If $|\mathcal{C} \cap \overline{V}_1(G')| \leq |\overline{V}_1(G')| - 2\varepsilon n$, then the sets $\mathcal{S}_1 = U \cap \overline{V}_1(G')$ and $\mathcal{S}_2 = \overline{V}_1(G') \setminus U$ (each of size at least εn) violate the above property. ∎

Note that, on the event stated in the above corollary, the component $\mathcal{C}$ contains all but at most $2\varepsilon dn$ edges of $E_1(G')$ (losing at most d edges per vertex in $V_1(G') \setminus \mathcal{C}$). Therefore, using (2.16), with probability at least $1 - \exp(-cn)$ we have that $\mathcal{C}$ is of linear size, and moreover,

$$|E(\mathcal{C})| > \tfrac{1}{2} p'd(1 - q'^2)n - (2d+1)\varepsilon n \, .$$

In particular, the component $\mathcal{C}$ belongs to the set of components in the left-hand side of (2.9), so

$$E(\mathcal{C}) \subseteq E_1(G) \, , \qquad V(\mathcal{C}) \subseteq V_1(G) \, .$$

By (2.14), the total number of edges in all components $\mathcal{C}' \neq \mathcal{C}$ with $|\mathcal{C}'| \geq 2R$ is at most

$$|E_1(G) \setminus E(\mathcal{C})| \leq \left(\tfrac{1}{2} pd(1 - (q - \varepsilon)^2) - \tfrac{1}{2} p'd(1 - q'^2) \right) n + (2d+2)\varepsilon n \, .$$

For small enough $\varepsilon > 0$, the right-hand is at most $\tilde{\varepsilon} n$, where $\tilde{\varepsilon} \downarrow 0$ as $\varepsilon \downarrow 0$. The proof of (2.7) is therefore concluded by the fact that, for a small enough $\varepsilon > 0$, one has $\tilde{\varepsilon} < \eta_1/2$, whence the total number of edges in components $\mathcal{C}' \neq \mathcal{C}$ with $|\mathcal{C}'| \geq 2R$ is strictly less than that in $\mathcal{C}$, thus in fact $\mathcal{C} = \mathcal{C}_1$.

Finally, to establish (2.8), recall that, by (2.17), with probability at least $1 - \exp(-cn)$ we have

$$|V(\mathcal{C})| \geq |V_1(G')| - 2\varepsilon n \geq (1 - q'(1 - p' + p'q') - 3\varepsilon)n \, .$$

Comparing this with the upper bound on $|V_1(G)|$ in (2.15), and recalling that $p' \uparrow p$ and $q' \uparrow q$ as $\varepsilon \downarrow 0$, thereby concludes the proof of Proposition 2.2. ∎

From the proposition and (2.10),(2.12), we deduce the required estimate on $|E(\mathcal{C}_1)|$, while (2.11),(2.13) analogously imply the required estimate on $|V(\mathcal{C}_1)|$.

Furthermore, from (2.7) we see that with probability $1 - \exp(-cn)$, at most $2\varepsilon n$ vertices in $G \sim \mathcal{G}_p$ have a discrepancy between their degree in $\mathcal{C}_1$ and that in $E_1(G)$. The statement of Theorem 2.1 that $|D_k/n - \alpha_k| < \varepsilon$ thus follows from the fact that, for every x and $1 \leq k \leq d$, the probability that $\#\{y : xy \in E_1(G)\} = k$ corresponds to α_k (up to replacing $1 - q$ by $1 - q + \varepsilon$ in that expression), as this event occurs iff x has exactly k neighbors in G, out of which there exists at least one vertex y which satisfies $\{|\mathcal{C}_y(\mathcal{C}_1 \setminus \{xy\})| \geq R - 1\}$ (up to replacing $R - 1$ by R in the case $k = 1$).

2.2 The 2-core of the giant component

Recall the definition of $\mathcal{A}_{x,y}$ in (2.3), and define

$$E_2(H) = \{xy \in E(H) \; : \; \mathcal{A}_{x,y}^R(H) \cap \mathcal{A}_{y,x}^R(H) \cap \{xy \in E(H)\}\},$$
$$V_2(H) = \{x \in V(H) \; : \; xy \in E_2(H) \text{ for some } y \in V(H)\}. \tag{2.19}$$

The main result we wish to prove in this subsection is the following characterization of the 2-core of a typical random graph $G \sim \mathcal{G}_p$. It consists of a local rule for inclusion of vertices and edges in $\mathcal{C}_1^{(2)}(G)$, the 2-core of its largest component, which determines it up to at most εn vertices. We also show that all other components contribute a combined total of at most εn vertices to the 2-core of G. (It is easy to see that, in both cases, a linear error of some $\varepsilon' n$ vertices must be allowed, e.g., when every vertex is part of a cycle of length $O(R)$.)

Proposition 2.5. *For every $\varepsilon, b > 0$ and $d \geq 3$ there exist some R and $c > 0$ such that, if $\mathcal{G}$ is a regular (b, d)-expander on n vertices with girth greater than $2R$, and $G \sim \mathcal{G}_p$, then*

$$\mathbb{P}\Big(\big|E_2(G) \triangle E(\mathcal{C}_1^{(2)}(G))\big| > \varepsilon n\Big) \leq \exp(-cn), \tag{2.20}$$

$$\mathbb{P}\Big(\big|V_2(G) \triangle V(\mathcal{C}_1^{(2)}(G))\big| > \varepsilon n\Big) \leq \exp(-cn), \tag{2.21}$$

whereas

$$\mathbb{P}\Big(\sum_{i \geq 2}|\mathcal{C}_i^{(2)}(G)| > \varepsilon n\Big) \leq \exp(-cn). \tag{2.22}$$

Proof. First observe that, for any graph H with girth greater than $2R$,

$$\bigcup\Big\{E(\mathcal{C}^{(2)}) \; : \; \mathcal{C} = \mathcal{C}_x(H) \text{ for some } x \in V(H)\Big\} \subset E_2(H);$$

indeed, for any such H and edge $xy \in E(H)$, if $\mathcal{A}_{x,y}^R$ does not hold then the component $\mathcal{C}_y(H \setminus \{xy\})$ has size less than R, and hence it is a tree by the girth assumption, so xy cannot belong to the 2-core. In particular,

$$\bigcup_{i \geq 1} E(\mathcal{C}_i^{(2)}(G)) \subset E_2(G) \quad \text{and} \quad \bigcup_{i \geq 1} V(\mathcal{C}_i^{(2)}(G)) \subset V_2(G). \tag{2.23}$$

Establishing the following bound (for some R, c depending on b, d, ε) will allow us to conclude the proof:

$$\mathbb{P}\left(|E(\mathcal{C}_1^{(2)}(G))| > (1 - \varepsilon)\tfrac{1}{2}dp(1 - q)^2 n\right) = 1 - O(\exp(-cn)). \tag{2.24}$$

In order to see that this indeed implies the statement of the proposition, note that, as in the argument preceding (2.10), applying Hoeffding's inequality to the appropriate edge-exposure martingale (where the Lipschitz constant is as before) implies that, if $G \sim \mathcal{G}_p$ then

$$\mathbb{P}\left(|E_2(G)| - \mathbb{E}[|E_2(G)|]| \geq a\right) \leq \exp\left(-\frac{a^2}{4dn(d - 1)^{2R}}\right), \tag{2.25}$$

10

where

$$\tfrac{1}{2}dp(1-q)^2 n \le \mathbb{E}[|E_2(G)|] \le \tfrac{1}{2}dp(1-q+\varepsilon)^2 n\,; \tag{2.26}$$

thus, the combination of (2.23)–(2.24) will indeed imply (2.20), and in particular will also give (2.21), as well as the upper bound (2.22) on the cumulative size of the 2-cores of $\mathcal{C}_i^{(2)}(G)$ for $i > 1$.

It remains to prove (2.24). As before, let $p' = p - \varepsilon > 1/(d-1)$ where $\varepsilon' = \varepsilon/(1-p')$ such that, if we take $G' \sim \mathcal{G}_{p'}$ and $F' \sim \mathcal{G}_{\varepsilon'}$ then we can produce $G \sim \mathcal{G}_p$ via $E(F') \cup E(G')$. The key to the proof would be to randomly partition $E(\mathcal{G})$ into $E_B \cup E_R$ — denoting E_B as BLUE edges and E_R as RED edges — and then connect the appropriate proportion of edges $(x,y) \in E(\mathcal{G})$ in the giant component $\mathcal{C}_1(G)$ via one BLUE edge e and one RED path to establish that $e \in E(\mathcal{C}_1^{(2)}(G))$. In what follows, if H is a subgraph of $\mathcal{G}$, we let H_B denote the (BLUE) subgraph of H whose edges are $E(H) \cap E_B$ and, similarly, we let H_R denote the (RED) subgraph of H whose edges are $E(H) \cap E_R$.

Remark 2.6. In principle, to show that $e \in E_2(G')$ is also in $E(\mathcal{C}_1^{(2)}(G))$, one would only need to show that $e \in \mathcal{C}_1$ and provide a path that connects the two clusters — each of size at least R — that are at its endpoints *without using said edge*. However, we must resort to witnesses in the form of a RED path between the endpoints x, y of the edge e which itself is BLUE due to our mechanism to guarantee that $e \in \mathcal{C}_1$: the latter uses sprinkling (new edges in F') to connect large clusters (ones of size at least R) to one another, and potentially might use, for instance, the edge e to connect the clusters of x and y. To remedy this, we independently color each edge of $\mathcal{G}$ in BLUE with some probability $\hat{\varepsilon}$, for a lower bound on the number of BLUE edges in $\mathcal{C}_1^{(2)}(G)$, which will turn out to be asymptotically tight, despite insisting on sprinkling only RED edges to connect the large clusters (see Remark 2.12).

The random partition $E(\mathcal{G}) = E_B \cup E_R$ is defined as follows: we let each $e \in E(\mathcal{G})$, independently, belong to E_B with probability $\hat{\varepsilon}$, given by

$$\hat{\varepsilon} := 3d\sqrt{\varepsilon}\,, \tag{2.27}$$

and further suppose that

$$\hat{\varepsilon} < \tfrac{1}{4}(24d)^{-2/b}\,, \tag{2.28}$$

which we may assume w.l.o.g. since the event we wish to estimate in (2.24) is monotone in ε.

Our goal will be to show that the edges in the set

$$\widehat{\mathcal{U}} := \big\{ e \in E(G'_B) \,:\, \mathcal{A}_{x,y}^R(G'_R) \cap \mathcal{A}_{y,x}^R(G'_R) \text{ holds} \big\}$$

represent, up to an arbitrarily small error, an $\hat{\varepsilon}$-proportion of the 2-core $\mathcal{C}_1^{(2)}(G)$. This would entail adapting our strategy of connecting small components via sprinkling to be restricted to RED edges (so we could guarantee $e \in \widehat{\mathcal{U}}$ would be part of a cycle), towards which we introduce the following notion.

Definition 2.7 (*k*-thick subsets). *We say a subset S of vertices of a graph H is* k*-thick if there exists a collection $\{S_i\}$ of disjoint connected subsets of H, each of size at least k, such that $S = \bigcup S_i$.*

The idea behind this definition is that, although $\mathcal{G}_R$ is not an expander — for instance, it contains a linear proportion of isolated vertices — w.h.p., sets that are k-thick maintain edge expansion in $\mathcal{G}_R$:

11

Claim 2.8. *Let $\mathcal{G}$ be a regular (b,d)-expander on n vertices, let $\varepsilon > 0$ and set $E_{\mathrm{B}}, E_{\mathrm{R}}$ as above for $\hat{\varepsilon}$ as in (2.27). There exists $k(\varepsilon, b, d)$ such that, with probability $1 - O(2^{-\varepsilon n})$,*

$$\#\{(x,y) \in E_{\mathrm{R}} \,:\, x \in S\,,\, y \in S^c\} \geq \tfrac{1}{2}b|S| \quad \textit{for every k-thick } S \subset V(\mathcal{G}) \textit{ with } \varepsilon n \leq |S| \leq n/2. \quad (2.29)$$

Proof. By the edge expansion assumption, the probability that a given k-thick subset S as above — denoting its size by s — violates the statement of the claim, is at most

$$\binom{\lceil bs \rceil}{\lceil bs/2 \rceil} \hat{\varepsilon}^{bs/2} \leq [(4 + o(1))\hat{\varepsilon}]^{bs/2} \,.$$

On the other hand, if $m = \lceil s/k \rceil$, then there are at most $2^{2s}\binom{n}{m}(ed)^s$ such subsets, as we may enumerate over all possible ways to write $s = s_1 + \ldots + s_m$ for the subset sizes (some possibly empty; there are at most m components since nonempty ones have size at least k), and then for each subset S_i we first root it in some $v_i \in V(\mathcal{G})$ — the total number of choices of these roots would be $\binom{n}{m}$ — and then specify its spanning tree out of at most $(ed)^{s_i}$ options (as all degrees are at most d; see, e.g., [5, Lemma 2]).

Hence, the probability that there exists some set S violating the statement of the claim is at most

$$\sum_{\varepsilon n \leq s \leq n/2} \left[(1 + o(1))\left(\frac{en}{s/k}\right)^{1/k + o(1)} 4ed(4\hat{\varepsilon})^{b/2}\right]^s \leq \sum_{s \geq \varepsilon n}\left[12d(4\hat{\varepsilon})^{b/2}\right]^s \leq \sum_{s \geq \varepsilon n} 2^{-s} = O(2^{-\varepsilon n})\,,$$

where we used that k is large enough such that $4e(e\varepsilon^{-1}k)^{1/k} < 12$, as well as (2.28). ∎

Using the above claim, we can produce a version of Claim 2.3 that will only consider E_{R} for sprinkling.

Claim 2.9. *For every $b, d, \varepsilon > 0$ there are $k, c > 0$ such that the following holds. If $\mathcal{G}$ is a regular (b,d)-expander on n vertices, $E(\mathcal{G}) = E_{\mathrm{B}} \cup E_{\mathrm{R}}$ satisfying (2.29), $\mathcal{S}$ is a family of disjoint connected subsets of $\mathcal{G}$, each of size at least k, and $H \sim \mathcal{G}_\varepsilon$ independently, then with probability $1 - \exp(-cn)$, every two subsets $\mathcal{S}_1, \mathcal{S}_2$ of $\mathcal{S}$, with at least εn vertices in each, are connected by a path in H_{R}.*

Proof. Each of the subsets $\mathcal{S}_1$ and $\mathcal{S}_2$ addressed by the claim is by definition k-thick in $\mathcal{G}$ (by the hypothesis on $\mathcal{S}$) and contains at least εn vertices. Thus, for any such $\mathcal{S}_1$ and $\mathcal{S}_2$, if (U, U^c) is a minimal cut separating these subsets of vertices in $\mathcal{G}_{\mathrm{R}}$ (i.e., $\mathcal{S}_1 \subset U$ and $\mathcal{S}_2 \subset U^c$ for U minimizing the number of edges between U, U^c in E_{R}) then w.l.o.g. U is also k-thick and $|U| \leq n/2$ (suppose U is the smaller part of the cut; we proceed by moving vertices that are not connected to $\mathcal{S}_1$ from U to U^c, noting that this would not increase the cut size). Hence, (2.29) implies that there are, with probability $1 - O(2^{-\varepsilon n})$, at least $\frac{1}{2}\varepsilon bn$ edges in this cut. By Menger's Theorem, as in the proof of Claim 2.3, we thus conclude that (2.18) holds for $A = \mathcal{S}_1$ and $B = \mathcal{S}_2$ with $b/2$ replacing b, and the rest of the argument in the proof of Claim 2.3 holds as before (with b replaced by $b/2$). ∎

Following the same short proof of Corollary 2.4 with the single modification of using the above claim instead of Claim 2.3 now yields its following analog w.r.t. sprinkling only RED edges.

12

Corollary 2.10. *For every $b, d, \varepsilon > 0$ there are $c, R > 0$ so that the following holds. If $\mathcal{G}$ is a regular (b, d)-expander on n vertices with girth greater than $2R$, and $E(\mathcal{G}) = E_{\mathrm{B}} \cup E_{\mathrm{R}}$ satisfies (2.29), then with probability $1 - O(e^{-cn})$, there is a component $\mathcal{C}$ of $G'_{\mathrm{R}} \cup F'_{\mathrm{R}}$ containing all but at most $2\varepsilon n$ vertices of*

$$\overline{V}_1(G'_{\mathrm{R}}) := \left\{ y \in V(G) : \mathcal{A}^R_{x,y}(G'_{\mathrm{R}}) \text{ holds for some } x \in V(G) \right\} .$$

Condition on the partition of $E(\mathcal{G})$ into $E_{\mathrm{B}} \cup E_{\mathrm{R}}$, and suppose that (2.29) holds (which occurs with probability $1 - O(\exp(-cn))$ as per Claim 2.8). Observe that if $e = (x, y) \in \widehat{\mathcal{U}}$ then $x, y \in \overline{V}_1(G'_{\mathrm{R}})$ thanks to the events $\mathcal{A}^R_{y,x}(G'_{\mathrm{R}}), \mathcal{A}^R_{x,y}(G'_{\mathrm{R}})$, respectively. By the last corollary, with probability $1 - O(\exp(-cn))$, all the edges of $\widehat{\mathcal{U}}$ up to at most $2d\varepsilon n$ will thus belong to some component $\mathcal{C}$ of $G' \cup F'_{\mathrm{R}}$. Moreover, each such edge is BLUE, and its endpoints will be connected in $\mathcal{C}$ by a RED path (in $G'_{\mathrm{R}} \cup F'_{\mathrm{R}}$). In particular, with probability $1 - O(\exp(-cn))$, all but at most $2d\varepsilon n$ edges of $\widehat{\mathcal{U}}$ are in the 2-core of $\mathcal{C}$.

Note that, as argued above (2.12), the girth assumption on $\mathcal{G}$ implies that the events $\mathcal{A}^R_{x,y}$ and $\mathcal{A}^R_{y,x}$ are independent, and by their definition these are also independent of the event $\{(x, y) \in E_{\mathrm{B}}\}$ (which occurs with probability $\hat{\varepsilon}$). Therefore, our bounds on $\mathbb{P}(\mathcal{A}^R_{x,y})$ yield

$$\tfrac{1}{2} dp'(1 - q')^2 \hat{\varepsilon} n \leq \mathbb{E}|\widehat{\mathcal{U}}| \leq \tfrac{1}{2} dp'(1 - q' + \varepsilon)^2 \hat{\varepsilon} n \,,$$

and, as argued before, the standard concentration estimate via the edge-exposure martingale yields

$$\mathbb{P}\left(\left| |\widehat{\mathcal{U}}| - \mathbb{E}|\widehat{\mathcal{U}}| \right| \geq a \right) \leq \exp\left(-\frac{a^2}{4dn(d-1)^{2R}} \right) ,$$

so that, in particular, for some $c(\varepsilon, d, R) > 0$ and every sufficiently large n,

$$\mathbb{P}\left(|\widehat{\mathcal{U}}| \geq \tfrac{1}{2} dp'(1 - q')^2 \hat{\varepsilon} n - \varepsilon n \right) = 1 - O(\exp(-cn)) \,.$$

In conclusion, the aforementioned single (linear) component $\mathcal{C}$ with most of the edges of $\widehat{\mathcal{U}}$ must w.h.p. be the largest component $\mathcal{C}_1$ in light of our result in the previous section.

Denoting the number of edges and BLUE edges in the 2-core of $\mathcal{C}_1(G)$, respectively, by

$$M := \left| E(\mathcal{C}_1^{(2)}(G)) \right| \quad \text{and} \quad \widehat{M} := \left| E(\mathcal{C}_1^{(2)}(G)) \cap E_{\mathrm{B}} \right| ,$$

we conclude from the above analysis of $\widehat{\mathcal{U}}$ that, for some $c(\varepsilon, d, R) > 0$,

$$\mathbb{P}\left(\widehat{M} \geq \tfrac{1}{2} dp'(1 - q')^2 \hat{\varepsilon} n - (2d + 1)\varepsilon n \right) = 1 - O(\exp(-cn)) \,.$$

At the same time, since the partition $E(\mathcal{G}) = E_{\mathrm{B}} \cup E_{\mathrm{R}}$ was performed independently of $G \sim \mathcal{G}_p$, we have that, conditional on G, the random variable $\widehat{M}$ is distributed as $\mathrm{Bin}(M, \hat{\varepsilon})$. In particular,

$$\mathbb{P}\left(\widehat{M} \leq M\hat{\varepsilon} + \varepsilon n \mid G \right) \geq 1 - \exp\left[-(\varepsilon n)^2/(2M) \right] \geq 1 - \exp\left[-(\varepsilon^2/d)n \right]$$

by Hoeffding's inequality and the fact that $M \leq dn/2$ deterministically. Combining the last two inequalities, it follows that, for some $c(\varepsilon, d, R) > 0$,

$$\mathbb{P}\left(M \geq \tfrac{1}{2} dp'(1 - q')^2 n - (2d + 2)(\varepsilon/\hat{\varepsilon})n \right) = 1 - O(\exp(-cn)) \,,$$

13

and the fact that $(2d + 2)/(\varepsilon/\hat{\varepsilon}) < \sqrt{\varepsilon}$ by our definition of $\hat{\varepsilon}$ in (2.27) (so that, with this probability, $M \geq \frac{1}{2}dp'(1-q')^2n - \sqrt{\varepsilon}\varepsilon n$, where as before, $\frac{1}{2}dp'(1-q')^2 \to \frac{1}{2}dp(1-q)^2$ as $\varepsilon \to 0$) now establishes (2.24) and thereby concludes the proof of Proposition (2.5). ∎

Remark 2.11. Note that our lower bound on $|\mathcal{C}_1^{(2)}(G)|$ — which was tight up to at most εn vertices — consisted of edges in the 2-core that lie on a cycle; thus, w.h.p. there are at most εn bridges in $\mathcal{C}_1^{(2)}(G)$.

Remark 2.12. A subtler aspect of the lower bound on $|\mathcal{C}_1^{(2)}(G)|$ is that counted edges that are BLUE and lie on cycles that, apart from this edge, are entirely RED; up to an arbitrarily small error, this matched the correct number of edges in $\mathcal{C}_1^{(2)}(G)$, where edges may lie on cycles that have mixed colors, thus the effect of such mixed cycles is negligible! It is important to note that this is *not* the case for the number of such cycles (e.g., consider the case when the girth tends to infinity), but rather for the sake of determining whether a given edge is contained in such a cycle.

From Proposition 2.5, Eq. (2.20), and (2.25)–(2.26), we deduce the required estimate on $|E(\mathcal{C}_1^{(2)})|$. Moreover, with probability $1 - \exp(-cn)$, at most $2\varepsilon n$ vertices in $G \sim \mathcal{G}_p$ have a discrepancy between their degrees in $\mathcal{C}_1^{(2)}(G)$ and in $E_2(G)$. As before, the statement of Theorem 2.1 that $|D_k^*/n - \beta_k| < \varepsilon$ follows from the fact that, for every x and $k \geq 2$, the probability that $\#\{y : xy \in E_2(G)\} = k$ corresponds to β_k, as this occurs iff x has exactly k neighbors y in G such that $\{|\mathcal{C}_y(\mathcal{C}_1^{(2)} \setminus \{xy\})| \geq R\}$. (By (1.1), the expression for β_k in (2.2) equals $\binom{d}{k}p^k(1-q)^k (1 - p + pq)^{d-k}$, i.e., $\mathbb{P}(\mathrm{Bin}(d, p(1-q)) = k)$.)

3 Second largest component

3.1 Proof of Theorem 2

Fix $d \geq 3$ and $R \geq 1$, and let $\frac{1}{d-1} < p < 1$ and $0 < \alpha < 1$. We need the following result, which (although it may be proved directly) follows immediately from our results in §2.

Claim 3.1. *For every* $b > 0$, $d \geq 3$ *and* $\frac{1}{d-1} < p < 1$ *there exist some* $c, \delta, \varepsilon, R > 0$ *such that, if* $\mathcal{G} = (V, E)$ *is a regular* (b, d)-*expander on* n *vertices with girth at least* R, *then there exists a subset* M *of at least* δn *vertex-disjoint edges of* $\mathcal{G}$ *such that the graph* $\mathcal{G}_0 = (V, E \setminus M)$ *satisfies that* $G_0 \sim (\mathcal{G}_0)_p$ *contains a connected component of size at least* εn *with probability* $1 - O(e^{-cn})$.

Proof. Using the notation in §2.2, let each edge in $\mathcal{G}$ be BLUE independently with probability $\hat{\varepsilon}$, as defined in (2.27) and RED otherwise; denote the RED and BLUE edges by E_R and E_B, respectively. Corollary 2.10 guarantees that, with probability $1 - O(e^{-cn})$, the red graph $\mathcal{G}_R = (V, E_R)$ satisfies that $G_R \sim (\mathcal{G}_R)_p$ contains a component of size εn for some $\varepsilon > 0$. Finally, E_B is of size at least $(\hat{\varepsilon}/2)dn/2$ with probability $1 - O(e^{-cn})$, whence, in particular, it contains a matching M of size at least $\hat{\varepsilon}n/4$. ∎

Our construction of a regular (b, d)-expander $\mathcal{G}$ on n vertices is as follows.

- Fix some arbitrary $b_1 > 0$. Let R_1 be the maximum of R and the girth requirement from Theorem 1 w.r.t. $\varepsilon = \theta_1/2$, and construct an expander H_1 on $n_1 \asymp n^\alpha$ vertices with girth at least

14

R_1, where $\hat{n}_1 \asymp n^\alpha$ vertices have degree 2 and the rest have degree d, in the following way: As per Claim 3.1, let H_1 be obtained from a regular (b_1, d)-expander on $n_1 = \lfloor n^\alpha \rfloor$ vertices with girth at least R_1 by subdividing each of the edges of M (given by that claim) into a path of length 2 via a new degree-2 vertex (for a total of $\hat{n}_1 \asymp n^\alpha$ new degree-2 vertices). Let $\hat{V}_1$ denote the $\hat{n}_1$ vertices of degree 2 in H_1.

- On each vertex $v \in \hat{V}_1$ in H_1: connect it via new edges to $d-2$ new $(d-1)$-ary trees T_v, each of depth

$$h = \left\lceil \frac{1}{2}(1-\alpha) \log_{d-1} n \right\rceil .$$

- Let F be a d-regular graph on m vertices, for some fixed m, with girth at least R — e.g., for concreteness, the Erdős–Sachs [9] graph — and an arbitrary edge (x, y) removed from it. Take

$$L = \left\lceil \frac{2 + \alpha}{(1-\alpha) \log_{d-1}(1/p)} \right\rceil ,$$

and replace every edge (a, b) of each $(d-1)$-tree T_v by a path of L copies of F (with $\{(x_i, y_i)\}_{i=1}^{L}$ denoting their deleted edges), where adjacent copies have y_i connected to x_{i+1} by a new edge, the vertex a is connected by a new edge to x_1, and b is connected by a new edge to y_L.

Note that this did not modify any of the degrees in the original vertices of the $(d-1)$-ary tree T_v (which are thus d everywhere except for the degree-1 leaves).

- Identify each of the $\hat{n}_2 \asymp n^{\frac{1}{2}(1+\alpha)}$ total leaves in all trees T_v as above, in an arbitrary way, to the vertices of one final regular b_2-expander H_2 with $\hat{n}_2$ vertices of degree $d-1$ and n_2 vertices of degree d, whose girth is at least R_1, where $n_2 \asymp n$ is such that altogether there are n vertices.

It is easy to see that the $\mathcal{G}$ is a regular (b, d)-expander for some fixed $b > 0$, depending only on b_1, b_2, d, as every set of s vertices must have at least $s/3$ vertices belong either to H_1, or to the interior vertices of the trees $\{T_v : v \in H_1\}$, or to H_2, thus the required expansion can be inferred from the one within that corresponding expander (out of the three) by itself. It remains to examine $\mathcal{G}_p$.

By Theorem 1, the size of the largest component $\mathcal{C}_1$ in $(H_1)_p$ is of order n^α (thanks to Claim 3.1) with probability at least $1 - \exp(-cn^\alpha)$. Each vertex v of this component connects to H_2 in $\mathcal{G}_p$ (i.e., there is a path in $(T_v)_p$ from the root to one of the leaves) with probability at most $(d-1)^h p^{hL}$; thus, the probability that H_1 is connected to H_2 in $\mathcal{G}_p$ is at most

$$\hat{n}_2 \, p^{hL} \asymp n^{(1+\alpha)/2} p^{hL} \leq n^{-1/2} = o(1) .$$

Furthermore, the size of this augmented component in $\mathcal{G}_p$ is stochastically dominated by the combined size of $O(n^\alpha)$ i.i.d. GW-trees with offspring distribution $\mathrm{Bin}(d-1, p^L)$ (multiplied by an extra factor of $L|V(F)| = O(1)$ due to the L copies of the graph F). Since $p^L < d^{-2}$, in particular these GW-trees are subcritical, hence the size of this component is of order n^α w.h.p.

On the other hand, the entire graph has girth at least R and is a d-regular expander, hence contains a linear component with probability at least $1 - \exp(-cn)$ by Theorem 1. $\blacksquare$

15

Remark 3.2. In the above construction, unlike the case of $\mathcal{G}(n,p)$, some edges can have a polynomially small probability of belonging to the giant, even though the girth can be made arbitrarily large and in particular the local neighborhoods (to an arbitrary distance) look all alike.

3.2 Proof of Theorem 3

The basic building block in the construction, which was already used for the construction in the proof of Theorem 2, is the tree $T(k, h, h^*, L^*)$, for integers $k \geq 2$ and $h, h^*, L^* \geq 1$, obtained by

(a) taking a k-ary tree with h levels, and

(b) subdividing all edges between parent nodes in the last h^* levels (i.e., in every level $j \geq h - h^*$) and their children to a path of length L^*.

Let $0 < p < 1$, fix some sufficiently small $\varepsilon > 0$, and set

$$d := \lceil (1/p)^{2/\varepsilon} \rceil \,, \tag{3.1}$$

and define

$$h_n := \lfloor \log_d n \rfloor \,, \quad \alpha := \left\lceil \varepsilon^{-1} \log_{1/p} d \right\rceil \,, \quad \beta := \left\lceil 2 \log_{1/p} d \right\rceil \,.$$

With these notations, our construction is the following graph $\mathcal{G}$ on $|V(\mathcal{G})| \asymp n$ vertices.

(i) Let $T_1 = T(d, h_n, \lfloor \varepsilon h_n \rfloor, \alpha)$. Denote its root by v and its leaves by $\{u'_1, \ldots, u'_N\}$ (so $n/d < N \leq n$).

(ii) Let $T_2 = T(d, h_n, \lfloor \varepsilon h_n \rfloor, \beta)$ and denote its leaves by $\{u''_1, \ldots, u''_N\}$.

(iii) Let F be the graph formed by taking some (arbitrary) d-regular expander F_0 on the vertex set $\{w_1, \ldots, w_N\}$, and then subdividing each edge of F_0 into a path of length β.

(iv) Add the edges (u'_i, w_i) and (u''_i, w_i) for all $i = 1, \ldots, N$.

Claim 3.3. *For large enough n, the graph $\mathcal{G}$ above satisfies* $\mathrm{diam}(\mathcal{G}) \leq (1 + 6\varepsilon) \log_{1/p} n$.

Proof. First note that the height of T_2 is

$$\ell_2 := h_n + (\beta - 1)\lfloor \varepsilon h_n \rfloor \leq \log_d n + 2\varepsilon \log_{1/p} n \leq \tfrac{5}{2}\varepsilon \log_{1/p} n \,,$$

where we used the fact that $\log_d n < (\varepsilon/2) \log_{1/p} n$ by (3.1). Therefore

$$\mathrm{dist}_{\mathcal{G}}(x, y) \leq R := 2\ell_2 + \beta + 2 = 5\varepsilon \log_{1/p} n + O(1) \quad \text{for every } x, y \in V(F) \cup V(T_2) \,,$$

where the additive $\beta + 2$ accounts for possibly traversing from the vertex x to the closest vertex w_i to it and then to v_i (at total distance at most $\beta/2 + 1$ from x), and similarly for y.

Next, consider $x \in V(T_1)$ vs. $y \in V(F) \cup V(T_2)$. If u'_i is some descendent of x in the tree T_1, then as established above, $\mathrm{dist}(y, u'_i) \leq R$, while $\mathrm{dist}(x, u'_i)$ is at most the height of T_1, given by

$$\ell_1 := h_n + (\alpha - 1)\lfloor \varepsilon h_n \rfloor \leq \log_d n + \log_{1/p} n \leq (1 + \varepsilon/2) \log_{1/p} n \,,$$

16

and overall
$$\mathrm{dist}(x,y) \le \ell_1 + R \quad \text{for every } x \in V(T_1) \text{ and } y \in V(F) \cup V(T_2) \,.$$

Finally, consider $x, y \in V(T_1)$, let u_i' and u_j' be some descendants of x, y, respectively, and write $d_x = \mathrm{dist}(x, u_i')$ and $d_y = \mathrm{dist}(y, u_j')$. If $d_x + d_y \le \ell_1$, then $\mathrm{dist}(x, y) \le d_x + d_y + \mathrm{dist}(u_i', u_j') \le \ell_1 + R$, and otherwise $\mathrm{dist}(x, y) \le \mathrm{dist}(v, x) + \mathrm{dist}(v, y) \le (\ell_1 - d_x) + (\ell_2 - d_y) < \ell_1$, thus overall,
$$\mathrm{dist}(x,y) \le \ell_1 + R \quad \text{for every } x \in V(T_1) \text{ and } y \in V(\mathcal{G}) \,.$$

Plugging in the values of ℓ_1 and R we see that $\mathrm{diam}(\mathcal{G}) \le \ell_1 + R \le (1 + 6\varepsilon) \log_{1/p} n$ for all sufficiently large n (absorbing the additive $O(1)$-term in R via the increased pre-factor of $\varepsilon \log_{1/p} n$), as required. ∎

Claim 3.4. *W.h.p., the random graph $G \sim \mathcal{G}_p$ satisfies $|\mathcal{C}_1(G)| \le n^{1-\varepsilon/3}$.*

Proof. Let U_1 denote the vertices $x \in V(\mathcal{G})$ which are at distance at most $(1 - \varepsilon/2 + \alpha\varepsilon/2)h_n$ from v, the root of T_1 (these are the vertices of the subtree of T_1 in the first $(1 - \varepsilon/2)h_n$ levels of the tree before the subdivision, as well as the new vertices in the subdivided edges between them). Similarly, let U_2 denote the vertices $x \in V(\mathcal{G})$ whose distance from the root of T_2 is at most $(1 - \varepsilon/2 + \beta\varepsilon/2)h_n$, and let $U_3 = V(\mathcal{G}) \setminus (U_1 \cup U_2)$. By construction, $|U_i| = O(d^{(1-\varepsilon/2)h_n}) = O(n^{1-\varepsilon/2})$ for $i = 1, 2$.

At the same time, if $x \in U_3$ then exploring the neighborhood of x in $\mathcal{G}$ via Breadth-First-Search reveals edges that were subdivided (in either F, T_1 or T_2) into paths of length at least β, to within depth (prior to the subdivision) of at least $\ell := \lceil (\varepsilon/2)h_n \rceil$. In particular, this exploration process in $G \sim \mathcal{G}_p$ is stochastically dominated by a Galton–Watson tree with offspring variable $\mathrm{Bin}(d, p^\beta)$. Thus, if ζ_i are i.i.d. such random variables, and Λ is the total progeny of this Galton–Watson tree, then

$$\mathbb{P}(\Lambda > k) \le \mathbb{P}\left(\sum_{i=1}^{k}(\zeta_i - 1) \ge 0\right) \le \mathbb{P}\left(\mathrm{Bin}(kd, p^\beta) \ge k\right) \le \binom{kd}{k}p^{\beta k} \le \left(edp^\beta\right)^k,$$

which, for $k = \ell$ and using $p^\beta < d^{-2}$, is at most $(d/e)^{-\ell} = O\big(n^{-\frac{2}{5}\varepsilon}\big)$ provided that ε is sufficiently small. So, if $U_3' = \{x \in U_3 : |\mathcal{C}_x(G)| > \ell\}$, then $\mathbb{E}|U_3'| = O(n^{1-\frac{2}{5}\varepsilon})$. As $|\mathcal{C}_1(G)| \le \max\{\ell, |U_1| + |U_2| + |U_3'|\}$, we find that $\mathbb{E}|\mathcal{C}_1(G)| = O(n^{1-\frac{2}{5}\varepsilon})$, which implies the statement of the claim. ∎

Claim 3.5. *For large enough n, the random graph $G \sim \mathcal{G}_p$ satisfies*

$$\mathbb{P}\left(\mathrm{diam}(\mathcal{C}_v(G)) > (1 - \varepsilon) \log_{1/p} n\right) \ge 1 - \varepsilon \,.$$

Proof. Let A denote the vertices of T_1 at distance $h_n^* = h_n - \lfloor \varepsilon h_n \rfloor$ from its root v, and let B be the set of vertices whose distance from v, prior to the subdivision of the edges of the tree T_1, was $h_n^* + \ell$, where $\ell = \lfloor (\varepsilon - 2\varepsilon^2)h_n \rfloor$. The probability that percolation on the subtree of T_1 rooted at some $x \in A$ survives to intersect B is at least

$$p^{\alpha\ell} \le d^{-\varepsilon^{-1}(\varepsilon - 2\varepsilon^2)h_n} = n^{-1+2\varepsilon} \,.$$

Recall that, if Z_t is the size of generation t in a Galton–Watson tree with offspring variable $\mathrm{Bin}(d,p)$, such that $m = dp > 1$ and its extinction probability is $0 < q < 1$, then $Z_t m^{-t}$ converges a.s. as $t \to \infty$ to a random variable W which, except for a mass of $1 - q$ at $W = 0$, has an absolutely continuous distribution on $(0, \infty)$ (see, e.g., [2, §I.12]). In particular, there exists $\delta > 0$ such that $\mathbb{P}(Z_t \geq \delta m^t) > 1 - q - \varepsilon/3$ for every sufficiently large t. Further, since q is monotone decreasing in d, our choice of d in (3.1) readily implies (recall (1.1)) that $q < \varepsilon/3$ provided ε is small enough, thus overall $\mathbb{P}(Z_t \geq \delta m^t) > 1 - \frac{2}{3}\varepsilon$ for every large enough t. Specialized to our setting, we take $t = h_n^*$, and noting that

$$m^{h_n^*} \geq d^{-1} n^{1-\varepsilon} p^{(1-\varepsilon)\log_d n} \geq d^{-1} n^{1 - \frac{3}{2}\varepsilon + \frac{1}{2}\varepsilon^2}$$

(using that $\log(1/p) < \frac{1}{2}\varepsilon \log d$ by (3.1)), we infer that the following bound on the size of A', the set of all vertices $x \in A$ such that there exists a path from v to x in G: for every sufficiently large n,

$$\mathbb{P}\left(|A'| \geq n^{1 - \frac{3}{2}\varepsilon}\right) \geq 1 - \tfrac{2}{3}\varepsilon \,.$$

On the event $|A'| \geq n^{1 - \frac{3}{2}\varepsilon}$, the size of the set B' of vertices $y \in B$ that are connected to v in G stochastically dominates a random variable $Z \sim \mathrm{Bin}(n^{1 - \frac{3}{2}\varepsilon}, n^{-1+2\varepsilon})$ by our bound on the event that the subtree of $x \in A$ survives the percolation to intersect B. Since $Z > 0$ w.h.p. (being concentrated around $n^{\varepsilon/2}$), it then follows that $\mathbb{P}(B' \neq \emptyset) \geq 1 - \frac{2}{3}\varepsilon - o(1)$, whereas every vertex $y \in B' \cap \mathcal{C}_v(G)$ would imply that

$$\mathrm{diam}(\mathcal{C}_v(G)) \geq \mathrm{dist}(v, y) \geq h_n^* + \alpha\ell > (1 - 2\varepsilon)\log_{1/p} n \,. \qquad \blacksquare$$

As argued in the proof of Theorem 2, the graph obtained by connecting the expander F to T_i is clearly also an expander, hence (iterating this) so is the entire graph $\mathcal{G}$. The combination of Claims 3.3, 3.4 and 3.5 thus concludes the proof of Theorem 3. $\qquad \blacksquare$

4   Separators, long paths and complete minors

In this section we research further typical properties of a random subgraph of a high girth constant degree expander. The properties are: non-existence of a small separator, existence of a linearly long path, and existence of a large complete minor. The results obtained are "soft", so to say, and lack the precision of the conclusions of Theorem 1; still, we believe they are of interest and complement nicely the more accurate results. We start with arguing that a random subgraph of a high girth constant degree expander typically has all its separators linear in n. Given a graph $G = (V, E)$ on n vertices, a vertex set $S \subset V$ is called a *separator* if there is a partition $V = A \cup B \cup S$ of the vertex set of G such that G has no edges between A and B, and $|A|, |B| \leq 2n/3$. Separators serve to measure quantitatively the connectivity of large vertex sets in graphs; the fact that all separators in G are large indicates that it is costly to break G into large pieces not connected by an edge.

Theorem 4.1. *Fix $d \geq 3$ and $\frac{1}{d-1} < p < 1$. For every $b > 0$ there exist some $\delta, c_1, R > 0$ such that, if $\mathcal{G}$ is a regular (b, d)-expander on n vertices with girth exceeding R, then with probability $1 - O(e^{-c_1 n})$, the largest connected component $\mathcal{C}_1$ of the random graph $G \sim \mathcal{G}_p$ has no separator of size at most δn.*

Proof. We first describe the idea of the proof. We will argue that if $G \sim \mathcal{G}_p$ has a small separator S, then by deleting all edges touching S we get a graph $G' \sim \mathcal{G}_{p'}$ without a connected component of size as large as dictated by Theorem 1, for some $\frac{1}{d-1} < p' < p$. Since deleting a relatively small number of edges touching S incurs a relatively small penalty when going from G to G', we will thus conclude that the probability of G to have a small separator must be exponentially small to begin with.

Now we provide a full proof, implementing rigorously the above described outline. For $\frac{1}{d-1} < p' < p$, to be chosen momentarily, denote

$$\theta_1 := \theta_1(p), \quad \theta_1' := \theta_1(p'),$$

where we apply (1.2) to define θ_1, θ_1'. Choose p' so $\theta_1' = \frac{5}{6}\theta_1$ (this is possible by the continuity of $\theta_1(p)$), and let

$$\rho = \frac{p'}{p}.$$

Notice that a random graph $G' \sim \mathcal{G}_{p'}$ can be obtained first by drawing a random graph $G \sim \mathcal{G}_p$, and then by retaining every edge of G with probability ρ independently.

Let $\mathcal{A}$ be the following event addressing $G \sim \mathcal{G}_p$ for some (small enough) $\delta > 0$ to be set later:

$$\mathcal{A} = \left\{ \exists S \subset [n] : \ |S| \leq \delta n \text{ and } |\mathcal{C}_1(G \setminus S)| \leq \tfrac{11}{15}\theta_1 n \right\},$$

and let $\mathcal{B}$ be the following event addressing $G' \sim \mathcal{G}_{p'}$:

$$\mathcal{B} = \left\{ |V(\mathcal{C}_1(G'))| \leq \tfrac{11}{15}\theta_1 n \right\}.$$

Suppose G satisfies $\mathcal{A}$, and choose S as in the definition of $\mathcal{A}$. There are at most $d|S| \leq \delta dn$ edges touching S in G. The probability to erase all these edges when going from G to G' is at most $(1-\rho)^{\delta dn}$. However, if none of these edges belongs in G', then $|\mathcal{C}_1(G')| \leq \tfrac{11}{15}\theta_1 n = \tfrac{22}{25}\theta_1' n$. It thus follows that

$$\mathbb{P}(\mathcal{B}) \geq \mathbb{P}(\mathcal{A}) \, (1-\rho)^{\delta dn}.$$

On the other hand, by Theorem 1, the order of $\mathcal{C}_1(G')$ is very close to $\theta_1' n$ with probability exponentially close to 1, making the event $\mathcal{B}$ exponentially unlikely. Specifically, applying Theorem 1 in $\mathcal{G}_{p'}$ with $\varepsilon = \tfrac{3}{25}\theta_1'$, we have $\mathbb{P}(\mathcal{B}) \leq Ce^{-cn}$ for $C, c > 0$. Hence,

$$\mathbb{P}(\mathcal{A}) \leq Ce^{-cn} \, (1-\rho)^{-\delta dn} \leq e^{-c'n},$$

for $\delta > 0$ small enough (as a function of other parameters).

Finally, observe that if $G \sim \mathcal{G}_p$ does not satisfy $\mathcal{A}$, then for every subset S of at most δn vertices, the graph $G - S$ has some connected component of size more than $\tfrac{11}{15}\theta_1 n$. Invoking Theorem 1 once again, this time in $\mathcal{G}_p$ with $\varepsilon = \tfrac{1}{10}\theta_1$, we derive that with probability exponentially close to 1, the random graph $G \sim \mathcal{G}_p$ does not satisfy $\mathcal{A}$, and its largest connected component satisfies $|\mathcal{C}_1(G)| \leq \tfrac{11}{10}\theta_1 n$. Therefore, in this case, recalling the definition of a separator, we derive that no subset S of at most δn vertices can be a separator in $\mathcal{C}_1(G)$. ∎

As we have indicated, non-existence of small separators is a key fact in deriving other typical properties of (the giant component) of a percolated high girth expander in the super-critical regime. They are given in the following two corollaries.

Corollary 4.2. *Fix $d \geq 3$ and $\frac{1}{d-1} < p < 1$. For every $b > 0$ there exist some $\delta, c, R > 0$ such that, if $\mathcal{G}$ is a regular (b, d)-expander on n vertices with girth exceeding R, then with probability at least $1 - \exp(-cn)$, the random graph $G \sim \mathcal{G}_p$ contains a path of length at least δn.*

Proof. We need the following result (see, e.g. [12, Proposition 2.1], including a simple proof):

Lemma 4.3. *Let k, l be positive integers. Assume $G = (V, E)$ is a graph on more than k vertices, where every $A \subset V$ of size $|A| = k$ has at least l neighbors in $V \setminus A$. Then G contains a path of length l.*

According to Theorems 1 and 4.1, with probability exponentially close to 1, the giant component $\mathcal{C}_1$ of $G \sim \mathcal{G}_p$ is of size close to $\theta_1 n$ and has no separator of size at most δn; we can assume $\delta \ll \theta_1$. Let A be a subset of $\mathcal{C}_1$ of cardinality $|A| = |V(\mathcal{C}_1)|/3$. Then by the definition of a separator, A has at least δn neighbors outside. It thus follows by Lemma 4.3 that $\mathcal{C}_1$ contains a path of length at least δn. ■

As it frequently happens in random graphs, getting a linearly long cycle from a linearly long path is pretty easy, here is a brief sketch of the argument. Choose $\frac{1}{d-1} < p' < p < 1$ and argue that a random graph $G \sim \mathcal{G}_{p'}$ has w.h.p. a path P of length $\delta' n$, for some constant $\delta' > 0$. Let P_1, P_2 be the first and the last thirds of P, respectively. Then by applying Menger's Theorem to the base graph $\mathcal{G}$, we derive that it contains linearly many constant length paths between $V(P_1)$ and $V(P_2)$. With high probability at least one of these short paths survives sprinkling (taking us from $\mathcal{G}_{p'}$ to $\mathcal{G}_p$); its union with P contains a linearly long cycle.

We now discuss embedding complete minors in percolated expanders. Let $G = (V, E)$, and let $t > 0$ be an integer. We say that G contains a *minor of the complete graph K_t* if there is a collection $(V_1, \ldots, V_t)$ of pairwise disjoint vertex subsets in V such that each V_i spans a connected subgraph in G, and in addition G has an edge between every pair of subsets V_i, V_j. Observe trivially that if G contains a minor of K_t, then $|E(G)| \geq \binom{t}{2}$; this trivial bound provides an obvious but meaningful benchmark for minor embedding statements.

Kawarabayshi and Reed proved in [11] that a graph G on n vertices has a minor of the complete graph K_h or a separator of order $O(h\sqrt{n})$. Since by Theorem 4.1 the random graph $G \sim \mathcal{G}_p$ has all separators of size at least δn in its giant component $\mathcal{C}_1$ (whose size is much larger than δn by Theorem 1) with probability exponentially close to 1, we conclude:

Corollary 4.4. *Fix $d \geq 3$ and $\frac{1}{d-1} < p < 1$. For every $b > 0$ there exist some $\delta, c, R > 0$ such that, if $\mathcal{G}$ is a regular (b, d)-expander on n vertices with girth exceeding R, then with probability $1 - O(e^{-cn})$, the random graph $G \sim \mathcal{G}_p$ contains a minor of $K_{\delta\sqrt{n}}$.*

The order of magnitude for the maximal h such that G contains a minor of K_h is obviously optimal, as the base graph $\mathcal{G}$ has only linearly many edges to begin with.

References

[1] N. Alon, I. Benjamini, and A. Stacey. Percolation on finite graphs and isoperimetric inequalities. *Ann. Probab.*, 32(3A):1727–1745, 2004.

[2] K. B. Athreya and P. E. Ney. *Branching processes.* Dover Publications, Inc., Mineola, NY, 2004. Reprint of the 1972 original [Springer, New York].

[3] I. Benjamini. Euclidean vs. graph metric. In *Erdős Centennial*, volume 25 of *Bolyai Soc. Math. Stud.*, pages 35–57. János Bolyai Math. Soc., Budapest, 2013.

[4] I. Benjamini, A. Nachmias, and Y. Peres. Is the critical percolation probability local? *Probab. Theory Related Fields*, 149(1-2):261–269, 2011.

[5] A. Beveridge, A. Frieze, and C. McDiarmid. Random minimum length spanning trees in regular graphs. *Combinatorica*, 18(3):311–333, 1998.

[6] B. Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[7] J. Ding, J. H. Kim, E. Lubetzky, and Y. Peres. Anatomy of a young giant component in the random graph. *Random Structures Algorithms*, 39(2):139–178, 2011.

[8] J. Ding, E. Lubetzky, and Y. Peres. Anatomy of the giant component: the strictly supercritical regime. *European J. Combin.*, 35:155–168, 2014.

[9] P. Erdős and H. Sachs. Reguläre Graphen gegebener Taillenweite mit minimaler Knotenzahl. *Wiss. Z. Martin-Luther-Univ. Halle-Wittenberg Math.-Natur. Reihe*, 12:251–257, 1963.

[10] S. Janson, T. Łuczak, and A. Ruciǹski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[11] K.-i. Kawarabayashi and B. Reed. A separator theorem in minor-closed classes. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science—FOCS 2010*, pages 153–162. IEEE Computer Soc., Los Alamitos, CA, 2010.

[12] M. Krivelevich. Long paths and hamiltonicity in random graphs. In N. Fountoulakis and D. Hefetz, editors, *Random Graphs, Geometry and Asymptotic Structure*, London Mathematical Society Student Texts, pages 4–27. Cambridge University Press, 2016.

[13] A. Nachmias and Y. Peres. Critical percolation on random regular graphs. *Random Structures Algorithms*, 36(2):111–148, 2010.

[14] B. Pittel. Edge percolation on a random regular graph of low degree. *Ann. Probab.*, 36(4):1359–1389, 2008.

[15] R. van der Hofstad. *Random graphs and complex networks. Vol. 1.* Cambridge Series in Statistical and Probabilistic Mathematics, [43]. Cambridge University Press, Cambridge, 2017.