

# Percolation through Isoperimetry

Sahar Diskin <sup>\*</sup>      Joshua Erde <sup>†</sup>      Mihyun Kang <sup>‡</sup>  
Michael Krivelevich <sup>§</sup>

July 11, 2024

## Abstract

We provide a sufficient condition on the isoperimetric properties of a regular graph  $G$  of growing degree  $d$ , under which the random subgraph  $G_p$  typically undergoes a phase transition around  $p = \frac{1}{d}$  which resembles the emergence of a giant component in the binomial random graph model  $G(n, p)$ . We further show that this condition is tight.

More precisely, let  $d = \omega(1)$ , let  $\epsilon > 0$  be a small enough constant, and let  $p \cdot d = 1 + \epsilon$ . We show that if  $C$  is sufficiently large and  $G$  is a  $d$ -regular  $n$ -vertex graph where every subset  $S \subseteq V(G)$  of order at most  $\frac{n}{2}$  has edge-boundary of size at least  $C|S|$ , then  $G_p$  typically has a unique linear sized component, whose order is asymptotically  $y(\epsilon)n$ , where  $y(\epsilon)$  is the survival probability of a Galton-Watson tree with offspring distribution  $\text{Po}(1 + \epsilon)$ . We further give examples to show that this result is tight both in terms of its dependence on  $C$ , and with respect to the order of the second-largest component.

We also consider a more general setting, where we only control the expansion of sets up to size  $k$ . In this case, we show that if  $G$  is such that every subset  $S \subseteq V(G)$  of order at most  $k$  has edge-boundary of size at least  $d|S|$  and  $p$  is such that  $p \cdot d \geq 1 + \epsilon$ , then  $G_p$  typically contains a component of order  $\Omega(k)$ .

## 1 Introduction

### 1.1 Background and Motivation

In 1957, Broadbent and Hammersley [14] initiated the study of percolation theory in order to model the flow of fluid through a medium with randomly blocked channels. In *(bond) percolation*, given a *host graph*  $G$ , the *percolated random subgraph*  $G_p$  is obtained by retaining every edge of  $G$  independently and with probability  $p$ . A fundamental and extensively studied example of such a model is when the host graph  $G$  is the complete graph  $K_n$ , which coincides with the classical *binomial random graph model*  $G(n, p)$ .

In their seminal paper from 1960, Erdős and Rényi [21] showed that  $G(n, p)$ <sup>1</sup> undergoes a dramatic phase transition, with respect to its component structure, when the expected average degree is around 1 (that is, when  $p \cdot n \approx 1$ ). More precisely, given a constant  $\epsilon > 0$ , let us define  $y := y(\epsilon)$  to be the unique solution in  $(0, 1)$  of the equation

$$y = 1 - \exp\{-(1 + \epsilon)y\}. \tag{1}$$

Erdős and Rényi showed the following phase transition result.

---

<sup>\*</sup>School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel. Email: sahardiskin@mail.tau.ac.il.

<sup>†</sup>Institute of Discrete Mathematics, Graz University of Technology, Steyrergasse 30, 8010 Graz, Austria. Email: erde@math.tugraz.at. Research supported in part by FWF P36131. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

<sup>‡</sup>Institute of Discrete Mathematics, Graz University of Technology, Steyrergasse 30, 8010 Graz, Austria. Email: kang@math.tugraz.at. Research supported in part by FWF F1002 and I6502.

<sup>§</sup>School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel. Email: krivelev@tauex.tau.ac.il. Research supported in part by USA–Israel BSF grant 2018267.

<sup>1</sup>In fact, Erdős and Rényi worked in the closely related *uniform* random graph model  $G(n, m)$ .

**Theorem 1.1** ([21]). *Let  $\epsilon > 0$  be a small enough constant. Then, with probability tending to 1 as  $n$  tends to infinity,*

- (a) *if  $p = \frac{1-\epsilon}{n}$ , then all components of  $G(n, p)$  are of order  $O_\epsilon(\log n)$ ; and,*
- (b) *if  $p = \frac{1+\epsilon}{n}$ , then there exists a unique giant component in  $G(n, p)$  of order  $(1+o(1))y(\epsilon)n$ . Furthermore, all other components of  $G(n, p)$  are of order  $O_\epsilon(\log n)$ .*

We refer the reader to [10, 22, 28] for systematic coverage of random graphs, and to the monographs [12, 24, 29] for more background on percolation theory.

Let us say a little about a proof of Theorem 1.1. A standard heuristic (see, for example, [6, Chapter 10.4]) shows that, since every vertex in the host graph has degree  $\approx n$ , a local exploration process of the clusters in  $G(n, p)$  can be approximated by a branching process where the expected number of children is  $pn$ . From this, one can deduce that each vertex is contained in a ‘large’ component, in this case linear sized, with probability  $\approx y(\epsilon)$  where  $\epsilon = np - 1$ , and what remains is to show that these large clusters ‘merge’ into a unique ‘giant’ component (throughout the rest of the paper, we will often refer to a linear sized component as a giant component).

This heuristic implies that the threshold probability depends only on the local structure of the graph, the vertex degrees, whereas the order of the largest component above this threshold grows with the order of the host graph. In the case of  $G(n, p)$ , these two parameters are coincidentally approximately equal. However, more generally this heuristic suggests that we should see a sharp change in the order of the largest component in a random subgraph of an arbitrary  $d$ -regular host graph  $G$ , for  $d = \omega(1)$ , when  $p$  is around  $\frac{1}{d}$ .

In the subcritical regime, it is known (and it is fairly easy to show) that for  $d$ -regular graphs  $G$  the behaviour of Theorem 1.1 (a) is universal, in that when  $p = \frac{1-\epsilon}{d}$ , typically all components of  $G_p$  are of order at most  $O_\epsilon(\log |V(G)|)$  (see [16, Theorem 1] and also [35]). In the supercritical regime, consider first a disjoint union of copies of  $K_{d+1}$  — the largest component (whether we percolate or not) is then of order at most  $d$ . Thus, for such a phase transition to typically occur, one needs some additional conditions on the edge distribution of the host graph. Indeed, this phase transition has been studied in certain specific families of host graphs. In particular, in the case of the  $d$ -dimensional hypercube  $Q^d$ , the pioneering work of Ajtai, Komlós, and Szemerédi [1] and of Bollobás, Kohayakawa, and Łuczak [11] shows that the component structure of a supercritical percolated hypercube is quantitatively similar to that in Theorem 1.1 (b), in that the largest component has asymptotic order  $y(\epsilon)|V(Q^d)|$  and the second-largest component has order  $O_\epsilon(\log |V(Q^d)|)$ . This result has been further generalised by the authors to any high-dimensional product graph, whose base graphs are regular and of bounded degree [16]. In a different context, this phenomenon has also been observed in pseudo-random graphs, which lack any clear geometric structure but have fairly good expansion. In this setting Frieze, Krivelevich, and Martin [23] showed that when  $\frac{\lambda}{d} = o(1)$ , supercritical percolated  $(n, d, \lambda)$ -graphs demonstrate similar behaviour in terms of their component structure.

In all these cases, the proofs rely heavily on the *expansion properties* of the host graph. However, in this regard, high-dimensional product graphs and pseudo-random graphs are quite dissimilar. The former are well-structured with respect to their geometry and display a high level of symmetry, yet they are (generally) not very good expanders, whereas the latter are much better expanders and have very uniform edge-distribution, but lack any obvious geometric structure or symmetry.

In light of this, it is perhaps natural to ask if a sufficiently strong assumption on the expansion of the host graph  $G$  alone is sufficient to guarantee a similar phase transition as in  $G(n, p)$ , without any further assumptions of geometric structure or pseudo-randomness, and if so, what are the minimal such requirements.

**Question 1.2.** *Let  $d = \omega(1)$ , and let  $G$  be a  $d$ -regular graph on  $n$  vertices. What requirements on the expansion of  $G$  do guarantee the typical existence of a component of asymptotic size  $y(\epsilon)n$  in the supercritical regime?*

To make this question more precise, we will need some way to quantify the expansion properties of the host graph. A well-known notion, capturing in part the expansion properties, is the *isoperimetric constant* of the graph, also known as the *Cheeger constant* due to its connection to the Cheeger isoperimetric constant of a Riemannian manifold [15].

Given a graph  $G = (V, E)$  and a subset  $S \subseteq V(G)$ , we write  $\partial_G(S)$  for the size of the edge-boundary of  $S$ , that is, the number of edges in  $G$  with one endpoint in  $S$  and the other endpoint in  $S^C := V \setminus S$ . The *isoperimetric constant* of  $G$  is defined as

$$i(G) := \min_{S \subseteq V, 1 \leq |S| \leq |V|/2} \left\{ \frac{\partial_G(S)}{|S|} \right\}.$$

For example, it follows from Harper's Theorem [26] that  $i(Q^d) = 1$ , whereas if  $G$  is an  $(n, d, \lambda)$ -graph with  $\frac{\lambda}{d} = o(1)$ , then standard results imply  $i(G) = (1 + o_d(1))\frac{d}{2}$  (see, for example, [33]).

Graphs with *constant* degree and  $i(G) > 0$  are known as *expander graphs* (or *expanders*) and have turned out to be very important in diverse areas of discrete mathematics and computer science (we refer the reader to [27] for a comprehensive survey on expander graphs and their applications). Question 1.2 has been studied in the context of constant degree expander graphs. For constant degree expanders with high-girth, the phase transition occurs around  $p = \frac{1}{d-1}$  (note that when  $d = \omega(1)$  and outside the critical window, the scaling of  $\frac{1}{d}$  suffices, whereas in the constant degree case, this difference matters). Indeed, Alon, Benjamini, and Stacey [4] showed that if  $G$  is an expander, then there exists at most one linear sized component in  $G_p$  for any  $p$  and under the further assumption that the graph has high-girth, they showed the existence of a unique giant component in the supercritical regime (that is, when  $p \cdot (d-1) > 1$ ), whose asymptotic order was determined by Krivelevich, Lubetzky, and Sudakov [31]. Subsequent work by Alimohammadi, Borgs, and Saberi [3] generalised this to digraph expanders with constant *average* degree. Let us further mention the work of Benjamini, Nachmias, and Peres [8], who showed that under extra assumptions on the host graph there is an  $o(1)$ -sharp threshold for the appearance of a giant component, and conjectured its universality. Subsequent work by Benjamini, Boucheron, Lugosi, and Rossignol [7] showed the existence of a sharp threshold for the existence of a giant component of order  $cn$  for any  $c \in (0, 1)$ .

Questions of a similar flavour to Question 1.2 have also been studied under the assumption that the host graph is transitive. In particular, Easo [19] showed necessary and sufficient conditions for the existence of a percolation threshold in vertex-transitive graphs, and Easo and Hutchcroft [20] showed sufficient conditions for the existence of a unique giant component above the percolation threshold in vertex-transitive graphs. It is further worth noting the work of Alimohammadi, Borgs, and Saberi [2], which studied the emergence of a giant component when the host graph is an expander with bounded average degree, in a more general percolation setting.

## 1.2 Main results

Towards answering Question 1.2, our first main result shows that a fairly weak assumption on the isoperimetric constant of the host graph is sufficient to guarantee the existence of a component whose asymptotic order is arbitrarily close to  $y(\epsilon)n$ .

**Theorem 1.** *Let  $d = \omega(1)$ , and let  $G$  be a  $d$ -regular  $n$ -vertex graph. Let  $\epsilon > 0$  be a small constant and let  $p = \frac{1+\epsilon}{d}$ . Then, there exists a constant  $C_0 := C_0(\epsilon) > 0$ , such that for every*

$C \geq C_0$  the following holds. If  $i(G) \geq C$ , then with probability at least  $1 - o_d(1) - \exp\left\{-\frac{\sqrt{Cn}}{d}\right\}$ ,

$$||L_1| - yn| \leq \frac{10n}{C^{1/20}},$$

where  $y = y(\epsilon)$  is as defined in (1) and  $|L_1|$  denotes the number of vertices in the largest component of  $G_p$ . Furthermore, the second-largest component of  $G_p$  is of order at most  $(1 + o_d(1))\frac{10n}{C^{1/20}}$ .

We note that in fact an even weaker assumption on the isoperimetry of the host graph, restricted to sets of linear size, suffices (see Remark 4.6). It is worth noting here that in a recent work of Borgs and Zhao [13], it was shown that under somewhat similar conditions on the isoperimetry of the host graph, together with *bounded* average degree assumption, there exists a local sampling algorithm which provides, among other things, a robust estimate for the order of the largest component.

Furthermore, it suffices to take  $C_0(\epsilon) = \Omega(\epsilon^{-40})$  in Theorem 1. Observe that for constant  $C$  this does not quite answer Question 1.2, but when taking  $C = C(d) \rightarrow \infty$  so that  $\frac{\sqrt{Cn}}{d} \rightarrow \infty$ , the result implies that **whp**<sup>2</sup>  $|L_1| = (1 + o_d(1))yn$  and all other components have sublinear order. In particular, when  $G$  is an  $(n, d, \lambda)$ -graph with  $\lambda = o(d)$ , we have that  $i(G) \geq (1 - o(1))\frac{d}{2}$ , and so Theorem 1 implies the typical existence of a giant component of order  $(1 + o_d(1))yn$  in the supercritical regime, which was first shown in [23].

On the other hand, in the case of the hypercube  $Q^d$ , a subcube  $S$  of size  $2^{d-1}$  has an edge-boundary of size only  $|S|$ . Nevertheless, we note that an ad-hoc adjustment of our method, utilising Harper's inequality [26], can give a proof for the typical existence of a giant component of the correct asymptotic order in supercritical  $Q_p^d$  (see Remark 4.6).

Since, in general, our methods will always lead to some inverse polynomial dependence on  $\epsilon$ , we have not attempted to optimise the power of  $C$  in the estimate of Theorem 1, nor the power  $\epsilon$  in the estimate of  $C_0$ . Nevertheless, the above example suggests it would be very interesting to see if the dependence here could be significantly improved.

We remark that our proof technique of Theorem 1 carries on to  $p = \frac{c}{d}$  for any constant  $c > 1$ , showing that **whp**,  $|L_1| = (1 + o(1))y(c-1)n$  (under an appropriate assumption on  $C_0 = C_0(c)$ ). Furthermore, our argument allows also for  $\epsilon = \epsilon(d)$  which tends slowly to zero as  $d$  tends to infinity.

Theorem 1 is tight in two senses. Firstly, there exists a  $d$ -regular graph  $G$  whose isoperimetric constant is bounded from below and some supercritical probability  $p$  such that **whp**  $G_p$  contains no linear sized component in the supercritical regime, and in fact only (poly-)logarithmically sized components.

**Theorem 2.** *Let  $C > 1$  be a constant, let  $d = \omega(1)$ , let  $n = \omega(d^2)$ , and let  $p \leq \frac{C}{d}$ . Then, there exists a  $d$ -regular  $n$ -vertex graph  $G$  with  $i(G) \geq \frac{1}{40C}$  such that **whp** all the components in  $G_p$  are of order at most  $3d \log n$ .*

Secondly, the bound on the size of the second-largest component in Theorem 1 is close to tight, in that there exists a  $d$ -regular graph  $G$  whose isoperimetric constant is arbitrarily large and supercritical probabilities  $p$  for which the second-largest component in  $G_p$  has almost linear size.

**Theorem 3.** *Let  $d = \omega(1)$ , let  $C \geq 1$  with  $7C \leq d$ , and let  $n \geq 10C \cdot d$ . Let  $\epsilon > 0$  be a sufficiently small constant, and let  $p = \frac{1+\epsilon}{d}$ . Then, there exists a  $d$ -regular  $n$ -vertex graph  $G$  with  $i(G) \geq C$ , such that the second-largest component of  $G_p$  is of order at least  $\epsilon d$  with probability at least  $1 - o_d(1) - \exp\left\{-\frac{n}{\exp\{30C\}d}\right\}$ .*

<sup>2</sup>With high probability, that is, with probability tending to 1 as  $d$  tends to infinity.

We note that in Theorem 3,  $C$  can, but does not have to, depend on  $d$  and tend slowly to infinity. Furthermore, note that for any  $d = o\left(\frac{n}{\exp\{30C\}}\right)$ , we have that **whp** the second-largest component is of order at least  $\epsilon d$ . Finally, we remark that in the constructions given for Theorems 2 and 3, there are some suitable parity assumptions on  $n$ ,  $d$ , and  $C$ , which we leave implicit here in order to simplify the statements of the theorems.

As mentioned before, the connection between the phase transition in a percolated subgraph and the isoperimetric properties of the host graph has been studied before for *constant*-degree high-girth expanders. To be more precise, given fixed constants  $d \in \mathbb{N}$  and  $i > 0$ , in their pioneering paper, Alon, Benjamini, and Stacey [4] showed that for a  $d$ -regular  $n$ -vertex graph  $G$  with  $i(G) \geq i$  and high-girth,  $G_p$  undergoes a phase transition around  $p = \frac{1}{d-1}$ , where the order of the largest component grows from sublinear to linear in  $n$ . Furthermore, they showed that the second-largest component in  $G_p$  in the supercritical regime is typically of order  $O(n^c)$ , for some  $c < 1$ . Subsequent work by Krivelevich, Lubetzky, and Sudakov [31] determined that the typical asymptotic order of the largest component in the supercritical regime, when  $p = \frac{1+\epsilon}{d-1}$ , is  $y(\epsilon)n$ , and demonstrated the existence of a high-girth constant-degree expander  $H$  such that the second-largest component of  $H_p$  in the supercritical regime is of order  $\Omega(n^c)$ , for any  $c < 1$ .

Let us briefly compare this to our results, where instead the degree  $d$  is tending to infinity. Whilst we have a similar phase transition in terms of the size of the largest component when  $p$  is around  $\frac{1}{d}$ , we require a stronger assumption on the isoperimetric constant,  $i(G) \geq C$ , for some large  $C$ . Theorem 2 shows that this is in some sense necessary, demonstrating a difference between the constant degree and growing degree setting. Moreover, whilst our assumption on the isoperimetric constant is stronger, let us note that any  $(n, d, \lambda)$ -graph  $G$  satisfies  $i(G) = \Theta(d)$ , and that for every  $S \subseteq V(Q^d)$  with  $|S| \leq |V(Q^d)|^{1-c}$ , for any  $c < 1$ , we have  $\partial_{Q^d}(S) = \Theta(d)|S|$ . It is thus perhaps surprising that Theorem 1 holds **whp** for any large enough  $C$ , with no dependence on  $d$  in how *quickly*  $C$  needs to tend to infinity.

Furthermore, outside of the largest component, the typical component structure in the supercritical regime is quantitatively different — in the constant degree setting the second-largest component is **whp** of order  $O(n^c)$  for some  $c < 1$ , while when the degree is growing, Theorem 3 shows that the second-largest component can be typically of order  $\Omega(n/t)$ , for any function  $t$  tending to infinity arbitrarily slowly, demonstrating a stark difference between the two settings.

Finally, we will also consider what we can say about Question 1.2 when we have rather less control over the expansion of the host graph. Indeed, let us suppose that we do not have good control on the isoperimetric constant of  $G$ , or even over the expansion of linear sized sets as in Remark 4.6, but rather a more ‘restricted’ control over the expansion of  $G$ , up to sets of a fixed size  $k$ . In the spirit of Question 1.2, can we still hope to determine the existence of some ‘large’ component? In this case, a graph consisting of many disjoint expanding graphs of size  $2k$  shows that we cannot hope to find a component of order larger than  $2k$ . Our second main result demonstrates that we can guarantee **whp** a component of order  $\Omega(k)$ , when the probability  $p$  is sufficiently large with respect to our local control on the expansion of  $G$ .

**Theorem 4.** *Let  $k = \omega(1)$ ,  $d \leq k$  and let  $G$  be a graph on more than  $k$  vertices, such that every  $S \subseteq V(G)$  with  $|S| \leq k$  satisfies  $\partial_G(S) \geq d|S|$ . Let  $\epsilon > 0$  be a small constant and let  $p = \frac{1+\epsilon}{d}$ . Then, with probability tending to 1 as  $k$  tends to infinity,  $G_p$  contains a component of order at least  $\frac{k}{2}$ .*

At first sight, the assumptions of Theorem 4 seem quite dissimilar to other statements about the existence of large clusters in percolated subgraphs. However, for sets  $S$  of size one, the condition is equivalent to the condition on the minimum degree of the host graph,  $\delta(G) \geq d$ , and for such graphs it is already known (see, e.g., [34]) that when  $p \geq \frac{1+\epsilon}{d}$  **whp**  $G_p$  contains a component of order  $\Omega(d)$ . In particular, the main interest here is when  $k \gg d$ . In some sense, we can think of Theorem 4 as giving an alternative heuristic for the nature of the phase transition — here our point of criticality is controlled by the expansion ratio of subsets, whereas

the quantitative aspects of the component structure above the critical point are controlled by the *scale* on which this level of expansion holds. Indeed, in the complete graph, sets of order  $k$  have edge-expansion of  $(1 - \frac{k}{n})n \cdot k$ . Thus, choosing  $k = \delta n$ , Theorem 4 shows that when  $p \cdot n > \frac{1}{1-\delta}$ , **whp**  $G(n, p)$  contains a linear sized component. Choosing  $\delta$  sufficiently small, we have that whenever  $p \cdot n > 1$ , **whp**  $G(n, p)$  contains a linear sized component.

In this way, Theorem 4 once again demonstrates the intrinsic connection between the expansion of the host graph, both global and ‘local’, and the typical emergence of large components in the percolated subgraphs. Moreover, we prove a variant of Theorem 4 which requires a slightly stronger assumption on the probability and on the degree of the graph, but allows one to verify instead only a ‘local’ expansion property in the host graph (see Theorem 3.1, stated in Section 3).

The paper is structured as follows. In Section 2 we set our notation, collect several lemmas which we will use throughout the proofs, and describe the Breadth First Search (BFS) algorithm which we will also use. In Section 3 we give short proofs of Theorems 4 and 3.1, utilising the BFS algorithm. The proof of Theorem 1, given in Section 4, is the most involved part of the paper, and therein lie several novel ideas. In Section 5 we give two constructions proving Theorems 2 and 3. Finally, in Section 6, we discuss our results and avenues for future research.

## 2 Preliminaries

We note that throughout the paper (and, in particular, in the statements of our main results), we make a fairly standard abuse of notation: when considering a  $d$ -regular  $n$ -vertex graph  $G$ , we in fact consider a sequence  $(d_k, n_k)_{k \in \mathbb{N}}$  of pairs  $(d_k, n_k) \in \mathbb{N}^2$  giving rise to a sequence  $(G_k)_{k \in \mathbb{N}}$  of  $d_k$ -regular  $n_k$ -vertex graphs. For example, a statement such as ‘let  $d = \omega(1)$ , let  $n = \omega(d^2)$  ... there exists a  $d$ -regular  $n$ -vertex graph such that...’ should be interpreted as ‘for all sequences  $(d_k, n_k)_{k \in \mathbb{N}}$  such that for every positive constant  $C$ , there exists  $k_0$  such that for every  $k \geq k_0$ , and every pair  $(d_k, n_k) \in \mathbb{N}^2$  satisfying  $d_k \geq C$  and  $n_k \geq Cd_k^2$  ... there exists a sequence  $(G_k)_{k \in \mathbb{N}}$  of  $d_k$ -regular  $n_k$ -vertex graphs such that...’.

As mentioned in the introduction, given a graph  $H$  and a subset  $S \subseteq V(H)$ , we denote by  $\partial_H(S)$  the number of edges with one endpoint in  $S$  and the other endpoint in  $S^C = V(H) \setminus S$ . We denote by  $N_H(S)$  the external neighbourhood of  $S$  in  $H$ . Further,  $H[S]$  denotes the induced subgraph of  $H$  on the vertex set  $S$ . Given two disjoint subsets  $A, B \subseteq V(H)$ , we denote by  $e_H(A, B)$  the number of edges (in  $H$ ) with one endpoint in  $A$  and the other endpoint in  $B$ . We denote by  $e_H(A)$  the number of edges in  $H[A]$ . Given  $v \in V(H)$ , we denote by  $C_H(v)$  the connected component in  $H$  to which  $v$  belongs. In each case, if the underlying graph is clear from the context, we may omit the subscript. Furthermore, throughout the paper, we omit rounding signs for the sake of clarity of presentation.

If  $y(\epsilon)$  is defined as in (1), then we note that  $y(\epsilon)$  is the survival probability of a Galton-Watson tree with offspring distribution  $\text{Po}(1 + \epsilon)$ , and it can be shown that  $y(\epsilon)$  is an increasing continuous function on  $(0, \infty)$  with  $y(\epsilon) = 2\epsilon - O(\epsilon^2)$ .

Given an  $n$ -vertex graph  $H$  and a subset  $I \subseteq [n]$ , the  $I$ -restricted isoperimetric constant of  $H$  is defined as

$$i_I(H) := \min_{S \subseteq V(H), |S| \in I} \left\{ \frac{\partial_H(S)}{|S|} \right\}.$$

When  $I = \{k\}$ , we may abbreviate  $i_{\{k\}}(H)$  into  $i_k(H)$ .<sup>3</sup> For example, in Theorems 1 and 4, our assumptions correspond to a lower bound on the restricted isoperimetric constants  $i_{\lfloor \frac{n}{2} \rfloor}(G)$  and  $i_{\lfloor k \rfloor}(G)$ , respectively.

We will make use of two standard probabilistic bounds. The first one is a typical Chernoff-type tail bound on the binomial distribution (see, for example, Appendix A in [6]).

<sup>3</sup>We note that in some papers  $i_k$  is used to refer to  $i_{\{k\}}$

**Lemma 2.1.** Let  $n \in \mathbb{N}$ , let  $p \in [0, 1]$ , and let  $X \sim \text{Bin}(n, p)$ . Then for any  $0 < t \leq \frac{np}{2}$ ,

$$\mathbb{P}[|X - np| \geq t] \leq 2 \exp\left\{-\frac{t^2}{3np}\right\}.$$

The second one is a variant of the well-known Azuma-Hoeffding inequality (see, for example, Chapter 7 in [6]),

**Lemma 2.2.** Let  $m \in \mathbb{N}$  and let  $p \in [0, 1]$ . Let  $X = (X_1, X_2, \dots, X_m)$  be a random vector with range  $\Lambda = \{0, 1\}^m$  with  $X_\ell$  distributed according to Bernoulli( $p$ ). Let  $f : \Lambda \rightarrow \mathbb{R}$  be a function such that there exists  $C \in \mathbb{R}$  such that for every  $x, x' \in \Lambda$  which differ only in one coordinate,

$$|f(x) - f(x')| \leq C.$$

Then, for every  $t \geq 0$ ,

$$\mathbb{P}[|f(X) - \mathbb{E}[f(X)]| \geq t] \leq 2 \exp\left\{-\frac{t^2}{2mpC^2}\right\}.$$

We also require the following bound on the number of  $k$ -vertex trees in a graph  $G$ , which follows immediately from [9, Lemma 2].

**Lemma 2.3.** Let  $G$  be a graph of maximum degree at most  $d$ , let  $v \in V(G)$ , and let  $k \in \mathbb{N}$ . Let  $t_k(G, v)$  denote the number of trees on  $k$  vertices in  $G$  rooted at  $v$ . Then

$$t_k(G, v) \leq \frac{k^{k-2}d^{k-1}}{(k-1)!} \leq (ed)^{k-1}.$$

Finally, we will utilise the following lemma, allowing one to find large matchings in percolated subgraphs, which follows immediately from [17, Lemma 3.8].

**Lemma 2.4.** Let  $G$  be a  $d$ -regular graph. Let  $0 < \delta < \frac{1}{2}$  be a constant, let  $q = \frac{\delta}{d}$ , and let  $s = \Omega(d)$ . Let  $F \subseteq E(G)$  be such that  $|F| \geq s$ . Then, there exists a constant  $c = c(\delta) \geq \delta^2$  such that  $F_q$ , a random subset of  $F$  obtained by retaining each edge independently with probability  $q$ , contains a matching of size at least  $\frac{cs}{d}$  with probability at least  $1 - \exp\left\{-\frac{cs}{d}\right\}$ .

## 2.1 The Breadth First Search algorithm

The Breadth First Search (BFS) algorithm is an algorithm which explores the components of a graph  $G$  by building a maximal spanning forest.

The algorithm receives as input a graph  $G = (V, E)$  and an order  $\sigma$  on  $V$ . The algorithm maintains three sets of vertices:

- $W$ , the set of vertices whose exploration is complete;
- $Q$ , the set of vertices currently being explored, kept in a queue; and
- $U$ , the set of vertices that have not been explored yet.

The algorithm starts with  $W = Q = \emptyset$  and  $U = V(G)$ , and ends when  $Q \cup U = \emptyset$ . At each step, if  $Q$  is non-empty, then the algorithm queries  $U$  for neighbours in  $G$  of the first vertex  $v$  in  $Q$ , according to  $\sigma$ . Each neighbour which is discovered is added to the back of the queue  $Q$ . Once all neighbours of  $v$  have been discovered, we move  $v$  from  $Q$  to  $W$ . If  $Q = \emptyset$ , then we move the next vertex from  $U$  (according to  $\sigma$ ) into  $Q$ . Note that the set of edges discovered during the algorithm forms a maximal spanning forest of  $G$ . In particular, we begin exploring a component  $K$  of  $G$  when the first vertex of  $K$  (according to  $\sigma$ ) enters  $Q$ , which was empty

at that moment, and we complete exploring the component the first moment  $Q$  becomes empty again.

In order to analyse the BFS algorithm on a random subgraph  $G_p$  of a graph  $G$  with  $|E(G)|$  edges, we will utilise the *principle of deferred decisions*. That is, we will take a sequence  $(X_j : 1 \leq j \leq |E(G)|)$  of i.i.d. Bernoulli( $p$ ) random variables, which we will think of as representing a positive or negative answer to a query in the algorithm. When the  $j$ -th edge of  $G$  is queried during the BFS algorithm, we will include it in  $G_p$  if and only if  $X_j = 1$ . Note that the forest obtained in this way has the same distribution as a forest obtained by running the BFS algorithm on  $G_p$ . We say the  $j$ -th edge of  $G$  is queried at *time*  $j$  in the process, and write  $W(j), Q(j)$  and  $U(j)$  for the sets  $W, Q$  and  $U$  directly after this edge has been queried.

### 3 Large components — assuming restricted isoperimetry

We begin with the proof of Theorem 4. The proof is short, drawing inspiration from [34] and showcases how one can utilise the properties of the BFS algorithm in the setting of percolation.

*Proof of Theorem 4.* We run the BFS algorithm described in Section 2.1 on  $G_p$ . Let  $t_0$  be the first time that  $|W(t_0)| = \frac{k}{2}$ , and let  $t_1$  be the first time when  $|W(t_1)| = k$ .

Suppose towards a contradiction that  $Q$  is empty at some time  $t \in [t_0, t_1]$ , and let  $k' \in [\frac{k}{2}, k]$  be such that  $|W(t)| = k'$ . By our assumption on the isoperimetric properties of  $G$  we have  $\partial(W(t)) \geq d \cdot k'$ , and since each edge between  $W(t)$  and  $W(t)^C$  has been queried at time  $t$ , we have that  $t \geq k'd$ .

On the other hand, at any time  $t$ , we have that  $\sum_{j=1}^t X_j \leq |W(t) \cup Q(t)|$ , and in particular if  $Q(t)$  is empty, then  $|W(t)| \geq \sum_{j=1}^t X_j \geq \sum_{j=1}^{k'd} X_j$ . By Lemma 2.1, we obtain

$$\mathbb{P} \left[ \sum_{j=1}^{k'd} X_j \leq \left(1 + \frac{\epsilon}{2}\right) k' \right] \leq 2 \exp \left\{ -\frac{\epsilon^2 k'^2}{4 \cdot 3(1 + \epsilon)k'} \right\} \leq \exp \left\{ -\frac{\epsilon^2 k}{25} \right\}.$$

Recalling that  $k' \in [\frac{k}{2}, k]$ , by the union bound over the at most  $kd$  possible values of  $k'd$ , we have that the probability that  $\sum_{j=1}^t X_j \leq \left(1 + \frac{\epsilon}{2}\right) k'$  is at most

$$dk \cdot \exp \left\{ -\frac{\epsilon^2 k}{25} \right\} \leq k^2 \cdot \exp \left\{ -\frac{\epsilon^2 k}{25} \right\} = o(1),$$

where the inequality follows from our assumption that  $k \geq d$ , and the equality follows from our assumption that  $k = \omega(1)$ . Thus, **whp** at any time  $t \in [t_0, t_1]$ , we have that  $|W(t)| \geq \sum_{j=1}^t X_j \geq \sum_{j=1}^{k'd} X_j > k'$  — a contradiction.

Therefore, **whp**  $Q(t)$  is not empty for any  $t \in [t_0, t_1]$ , and all the vertices in  $W(t_1) \setminus W(t_0)$  belong to the same component. We thus conclude that **whp** there exists a component of size at least  $\frac{k}{2}$  in  $G_p$ .  $\square$

We note that in the above argument, we can choose  $t_0$  to be the first moment where  $|W(t_0)| = \delta k$  for small  $\delta$ , and with a similar analysis deduce the existence of a component of order at least  $(1 - o(1))k$ . We have chosen  $\delta = \frac{1}{2}$  for clarity of presentation.

As mentioned in the introduction, under an additional assumption on the maximum degree of  $G$  and with a slightly larger probability, we can make a similar conclusion as to the size of the largest component under a much weaker isoperimetric assumption, where we only bound the expansion of sets of size *exactly*  $k$ .

**Theorem 3.1.** *Let  $k = \omega(1)$ , let  $c_1, c_2 \in (0, 1]$ , and let  $0 < d \leq c_1(1 - c_2)k$ . Let  $G$  be a graph on more than  $k$  vertices, with maximum degree  $d$ , and with  $i_k(G) \geq c_1 d$ . Let  $\epsilon > 0$  be sufficiently small, and let  $p = \frac{1 + \epsilon}{c_2 c_1 d}$ . Then, **whp**  $G_p$  contains a component of order at least  $\frac{c_1(1 - c_2)k}{4}$ .*



Before proving Theorem 3.1, we first prove the following lemma, which allows us to translate a bound on  $i_{k_1}(G)$  to one on  $i_{[k_2]}(G)$ , with  $k_2$  being not much smaller than  $k_1$ . The proof draws on ideas from [30].

**Lemma 3.2.** *Let  $k, d$  be positive integers, let  $c_1, c_2 \in [0, 1]$  be such that  $c_1 d < k$  and let us write  $c_3 := \frac{c_1(1-c_2)}{2}$ . Let  $G$  be a graph on more than  $k$  vertices, with maximum degree  $d$ , and such that  $i_k(G) \geq c_1 d$ . Then there exists a subgraph  $G' \subseteq G$  with  $|V(G')| \geq |V(G)| - k$ , such that  $i_{[c_3 k]}(G') \geq c_1 c_2 d$ .*

*Proof.* Initialise  $G_0 = G$  and  $W_0 = \emptyset$ . At each iteration  $j \in \mathbb{N}$ , if there is a subset  $B \subseteq V(G_{j-1})$  of size  $|B| \leq c_3 k$  with  $\partial_{G_{j-1}}(B) < c_2 c_1 d |B|$ , we update  $W_j := W_{j-1} \cup B$  and  $G_j := G_{j-1}[V(G_{j-1}) \setminus B]$ . We terminate this process once there are no more such subsets  $B$ , and let  $G'$  be the resulting graph.

Suppose towards contradiction that there is some  $m \in \mathbb{N}$  such that  $|W_m| \geq k$ , where we may assume without loss of generality that  $m$  is minimal with this property. Let  $B_0$  be the last subset added to  $W_m$ , so that  $W_m = W_{m-1} \cup B_0$ . Then, there is some  $B_1 \subseteq B_0$  such that  $|W_{m-1} \cup B_1| = k$ . Let us further set  $B_2 := B_0 \setminus B_1$ .

On the one hand, by our assumption on the isoperimetric inequality on  $G$ , that is,  $i_k(G) \geq c_1 d$ , we have that

$$\partial_G(W_{m-1} \cup B_1) \geq c_1 d k. \quad (2)$$

On the other hand,

$$\begin{aligned} \partial_G(W_{m-1} \cup B_1) &\leq \partial_G(W_{m-1}) + e_G(B_1, V(G) \setminus (W_{m-1} \cup B_1)) \\ &\leq \partial_G(W_{m-1}) + e_G(B_0, V(G) \setminus (W_{m-1} \cup B_0)) + e_G(B_1, B_2). \end{aligned}$$

Now, by construction, we have that  $\partial_G(W_{m-1}) < c_1 c_2 d |W_{m-1}| < c_2 c_1 d k$ . Moreover, by our choice of  $W_{m-1}$  and  $B_0$ , we have  $e_G(B_0, V(G) \setminus (W_{m-1} \cup B_0)) = \partial_{G_{m-1}}(B_0) < c_1 c_2 d |B_0|$ . Finally, since the graph  $G$  has maximum degree  $d$ , we have that  $e_G(B_1, B_2) < d |B_0|$ . Recalling that  $|B_0| \leq c_3 k = \frac{c_1(1-c_2)k}{2}$  and  $c_1 \in [0, 1]$ , we have that

$$\begin{aligned} \partial_G(W_{m-1} \cup B_1) &< c_1 c_2 d k + c_1 c_2 d \cdot \frac{c_1(1-c_2)k}{2} + d \cdot \frac{c_1(1-c_2)k}{2} \\ &\leq c_1 d k \left( c_2 + \frac{(1+c_2)(1-c_2)}{2} \right) = c_1 d k \left( c_2 + \frac{1-c_2^2}{2} \right), \end{aligned} \quad (3)$$

where in the last inequality we used our assumption that  $c_1 \leq 1$ . Observe that  $f(c_2) = c_2 + \frac{1-c_2^2}{2}$  is continuous, increasing on  $(-\infty, 1]$ , and attains the value of 1 when  $c_2 = 1$ . Recalling that  $c_2 \in [0, 1]$ , it follows from (2) and (3) that

$$c_1 d k \leq \partial_G(W_{m-1} \cup B_1) < c_1 d k \left( c_2 + \frac{1-c_2^2}{2} \right) \leq c_1 d k,$$

a contradiction. □

We can now show that Theorem 3.1 follows from Lemma 3.2 and Theorem 4.

*Proof of Theorem 3.1.* By Lemma 3.2, there exists a non-empty subgraph  $G' \subseteq G$  such that every  $S \subseteq V(G')$  with  $|S| \leq \frac{c_1(1-c_2)k}{2}$  has  $\partial_{G'}(S) \geq c_1 c_2 d |S|$ . Applying Theorem 4 to  $G'$  with  $p = \frac{1+\epsilon}{c_1 c_2 d}$ , we conclude that **whp**  $G'_p$  (and thus  $G_p$ ) contains a component of size at least  $\frac{c_1(1-c_2)k}{4}$ . □

It would be interesting to see whether the maximum degree condition of Lemma 3.2, and thus of Theorem 3.1, can be replaced with an assumption on the *average* degree of  $G$ .

**Remark 3.3.** We note that one cannot hope to obtain a path of length  $\Omega_\epsilon(k)$  under the assumptions of Theorem 4. For example, let  $G = K_{d,d^{10}}$ , and let  $k = d^2$ . We have that  $i_{[k]}(G) \geq d$ . All paths in  $G$  have length at most  $2d$ , and thus, naturally, we cannot hope to find any path of length  $\Omega_\epsilon(k)$  in a random subgraph of  $G$ .

## 4 Existence of a giant component under weak isoperimetric assumptions

Let us start this section by defining some useful notation, and giving a broad outline of the proof of Theorem 1.

Our proof will proceed using a three-round exposure. Let  $\delta := \delta(\epsilon) > 0$  be a sufficiently small constant. We let  $p_2 = p_3 = \frac{\delta}{d}$ , and let  $p_1$  be such that  $(1 - p_1)(1 - p_2)(1 - p_3) = 1 - p$ , so that  $G_p$  has the same distribution as  $G_{p_1} \cup G_{p_2} \cup G_{p_3}$ . In order to easily describe the intermediary stages of the three-round exposure, we let  $p'$  be such that  $(1 - p_1)(1 - p_2) = (1 - p')$ , and define the following three graphs.

- $G(1) = G_{p_1}$ , noting that  $p_1 \geq \frac{1+\epsilon-2\delta}{d}$ ;
- $G(2) = G(1) \cup G_{p_2}$ , noting that  $G(2) \sim G_{p'}$  and  $p' \geq \frac{1+\epsilon-\delta}{d}$ ; and
- $G(3) = G(2) \cup G_{p_3}$ , noting that  $G(3) \sim G_p$ .

Let us write  $\rho_1 = p_1, \rho_2 = p'$  and  $\rho_3 = p$ , so that  $G(j) \sim G_{\rho_j}$  for each  $j \in \{1, 2, 3\}$ . Note that  $V(G(j)) = V(G)$  for each  $j \in \{1, 2, 3\}$ .

We now define several sets which will be crucial to our analysis in the section. Firstly,  $V_L(G(j))$  is the set of vertices which lie in ‘large’ components in  $G(j)$ . To be precise,

$$V_L(G(j)) := \{v \in V(G) : |C_{G(j)}(v)| \geq \delta d\}.$$

Secondly,  $W_L(G(j))$  is the set of vertices with many neighbours in  $V_L(G(j))$ . That is,

$$W_L(G(j)) := \{w \in V(G) : |N_G(w) \cap V_L(G(j))| \geq \delta^5 d\}.$$

Finally,  $V_S(G(j))$  is the set of vertices in ‘small’ components in  $G(j)$ . That is,

$$V_S(G(j)) := \{v \in V(G) : |C_{G(j)}(v)| \leq \log^2 d\}.$$

Our strategy for proving Theorem 1 is then broadly as follows. The first step, which is relatively standard, is to show that **whp** the right asymptotic proportion of vertices lie in ‘large’ components in each  $G(j)$ . To do this, we estimate the number of vertices which lie in ‘small’ components and then show that a negligible proportion of the vertices lies in components of intermediate size.

The key part of the proof, which contains several novel arguments, is to show that **whp** almost all vertices have many neighbours which lie in ‘large’ components in  $G(2)$ . Such a statement already appears in the seminal work of Ajtai, Komlós and Szemerédi [1], and is a key part of the analysis of the phase transition in various geometric graphs [1, 16]. However, the proofs in these settings rely heavily on the *self-symmetry* of these graphs, and a key improvement here is to prove such a statement without any structural assumptions on the host graph, in fact relying only on its *regularity*. This follows from a delicate argument analysing a modified BFS process starting at the neighbours of a fixed vertex. The rough idea here is that, before we have discovered  $\Omega(d)$  neighbours which lie in large components, **whp**  $o(d)$  vertices are contained in components which are not ‘large’, which will allow us to couple the exploration process from below with a supercritical branching process, and conclude that each neighbour has a constant probability of lying in a large component.

With these two results in hand, we can argue for the typical existence of a giant component, roughly as in [1, 11, 16]. More precisely, we show that almost every vertex in  $V_L(G(2))$  will coalesce into a single component in  $G(3)$  as follows. We say that a partition  $V_L(G(2)) = A \cup B$  is component-respecting if  $K \cap V_L(G(2))$  is fully contained in either  $A$  or  $B$ , for every component  $K$  of  $G(2)$ . Given a component-respecting partition  $V_L(G(2)) = A \cup B$ , since almost every vertex is adjacent to many vertices in large components in  $G(2)$ , we can extend this partition to an almost partition  $A' \cup B'$  of  $V(G)$  such that  $A \subseteq A'$ ,  $B \subseteq B'$  and every vertex in  $A', B'$  has many neighbours in  $A, B$ , respectively. Our assumption on the isoperimetric properties of  $G$  ensures that there are many edges between  $A'$  and  $B'$ , which we can extend to a large family of paths in  $G$  between  $A$  and  $B$  of length at most 3. Using Lemma 2.4, which provides with very high probability a large matching between  $A'$  and  $B'$  in  $G_{p_3}$ , we can argue that with very high probability one of these paths is present in  $G(3)$ , and in fact that the failure probability is small enough that **whp** this holds for *every* component-respecting partition of  $V_L(G(2))$  of relevant sizes. Hence, **whp**  $G(3)$  contains a component containing almost all the vertices in  $V_L(G(2))$ , which contains asymptotically the required number of vertices.

Let us begin then by estimating the order of  $V_S(G(j))$  for each  $j \in \{1, 2, 3\}$ . Note that the number of vertices in small components is a decreasing property, and it thus suffices to upper bound the order of  $V_S(G(1))$  and to lower bound the order of  $V_S(G(3))$ . To that end, define

$$F(c) := \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} c^{k-1} \exp\{-ck\}.$$

It is known (see for example [21, p. 346]) that given  $c > 1$ , we have that

$$F(c) = 1 - y(c - 1), \tag{4}$$

where  $y(x)$  is defined as in (1). With this in hand, we are ready to estimate  $|V_S(G(j))|$ , in a manner similar to the one used already in the seminal work of Erdős and Rényi [21].

**Lemma 4.1.** *We have that whp*

$$\begin{aligned} |V_S(G(1))| &\leq (1 + o_d(1)) (1 - y(\epsilon - 2\delta)) n, \\ |V_S(G(3))| &\geq (1 - o_d(1)) (1 - y(\epsilon)) n. \end{aligned}$$

*Proof.* We begin by bounding  $\mathbb{E}[|V_S(G(1))|]$  from above. Fix  $k \leq \log^2 d$ . Let  $X_k$  be the number of vertices contained in components of order  $k$  in  $G(1) = G_{p_1}$ . Let  $\mathcal{T}_k$  denote the set of trees on  $k$  vertices in  $G$ . Then, recalling  $\rho_1 = p_1$ , we have

$$\begin{aligned} \mathbb{E}[|V_S(G(1))|] &\leq \mathbb{E}\left[\sum_{k=1}^{\log^2 d} X_k\right] \leq \sum_{k=1}^{\log^2 d} k \sum_{T \in \mathcal{T}_k} \rho_1^{k-1} (1 - \rho_1)^{e_G(T, T^C)} \\ &\leq \frac{n}{k} \cdot k \sum_{k=1}^{\log^2 d} \frac{(dk)^{k-2}}{(k-1)!} \rho_1^{k-1} (1 - \rho_1)^{ki_k(G)}, \end{aligned}$$

where the third inequality follows from the first inequality in Lemma 2.3 (and the factor of  $\frac{n}{k}$  comes from considering all  $n$  possible different choices of  $v$  in  $t_k(G, v)$ , and noting that each tree is counted  $k$  times), and from the definition of the restricted isoperimetric constant  $i_k(G)$ . Since  $k \leq \log^2 d$  and  $G$  is  $d$ -regular, it follows that  $i_k(G) \geq d - \log^2 d$ . Therefore, since  $\rho_1 \geq \frac{1+\epsilon-2\delta}{d}$

and  $x \exp\{-ax\}$  is decreasing when  $ax > 1$ , we have

$$\begin{aligned}
\mathbb{E}[|V_S(G(1))|] &\leq n \sum_{k=1}^{\log^2 d} \frac{k^{k-1}}{k!} (d\rho_1)^{k-1} \exp\{-\rho_1 k(d - \log^2 d)\} \\
&\leq n \sum_{k=1}^{\log^2 d} \frac{k^{k-1}}{k!} (1 + \epsilon - 2\delta)^{k-1} \exp\left\{-(1 + \epsilon - 2\delta)k \left(1 - \frac{\log^2 d}{d}\right)\right\} \\
&\leq \exp\left\{\frac{2 \log^4 d}{d}\right\} n \sum_{k=1}^{\log^2 d} \frac{k^{k-1}}{k!} (1 + \epsilon - 2\delta)^{k-1} \exp\{-(1 + \epsilon - 2\delta)k\} \\
&\leq (1 + o_d(1))n \cdot F(1 + \epsilon - 2\delta) \\
&= (1 + o_d(1))(1 - y(\epsilon - 2\delta))n,
\end{aligned}$$

where the penultimate inequality and last equality follow from (4).

Let us now bound  $|V_S(G(3))|$  from below. Since  $G$  is  $d$ -regular, for every  $v \in V(G)$  a standard coupling argument implies that  $|C_{G(3)}(v)|$  is stochastically dominated by the number of vertices in a Galton-Watson tree with offspring distribution  $\text{Bin}(d, \rho_3)$ . Hence, standard results (see, for example, [18, Theorem 4.3.12]) imply that for every  $v \in V(G)$ ,

$$\mathbb{P}[|C_{G(3)}(v)| \geq \log d] \leq y(\epsilon) + o_d(1).$$

Therefore,

$$\mathbb{E}[|V_S(G(3))|] \geq (1 - o_d(1))(1 - y(\epsilon))n.$$

Finally, we show that  $|V_S(G(j))|$  is tightly concentrated around its mean for each  $j \in \{1, 2, 3\}$ . Indeed, let us consider the standard edge-exposure martingale on  $G(j)$ . Changing any edge can change the value of  $|V_S(G(j))|$  by at most  $2 \log^2 d$ . Therefore, by Lemma 2.2, we have

$$\begin{aligned}
\mathbb{P}\left[||V_S(G(j))| - \mathbb{E}[|V_S(G(j))|]| \geq n^{2/3}\right] &\leq 2 \exp\left\{-\frac{n^{4/3}}{2 \cdot \frac{dn}{2} \cdot \rho_j \cdot (2 \log^2 d)^2}\right\} \\
&\leq \exp\left\{-\frac{n^{1/3}}{5 \log^4 n}\right\} = o(1),
\end{aligned}$$

where we used  $d \leq n$  in the last inequality.  $\square$

We now show that **whp** there are  $o_d(n)$  vertices in  $V(G) \setminus (V_S(G(j)) \cup V_L(G(j)))$ .

**Lemma 4.2.** *For all  $j \in \{1, 2, 3\}$ , **whp** the number of vertices in components of order between  $\log^2 d$  and  $\delta d$  in  $G(j)$  is  $o_d(n)$ .*

*Proof.* Fix  $\log^2 d \leq k \leq \delta d$ . As in the proof of Lemma 4.1, let  $X_k^{(j)}$  be the number of vertices in components of order  $k$  in  $G(j)$ . Then, by the second inequality in Lemma 2.3 (taking into account the less than  $n$  possible ways of choosing the root of each tree component on  $k$  vertices in  $G$ ), we obtain

$$\begin{aligned}
\mathbb{E}[X_k^{(j)}] &\leq k \sum_{T \in \mathcal{T}_k} \rho_j^{k-1} (1 - \rho_j)^{e_G(T, T^C)} \\
&\leq nk(ed)^{k-1} \rho_j^{k-1} (1 - \rho_j)^{ki_k(G)}.
\end{aligned}$$

Since  $k \leq \delta d$  and  $G$  is  $d$ -regular, we have that  $i_k(G) \geq (1 - \delta)d$ . Furthermore,  $\rho_j \geq \frac{1 + \epsilon - 2\delta}{d}$  for each  $j \in \{1, 2, 3\}$ , and since  $x \exp\{-ax\}$  is decreasing when  $ax > 1$ ,

$$\begin{aligned} \mathbb{E}[X_k^{(j)}] &\leq nk(ed)^{k-1} \rho_j^{k-1} (1 - \rho_j)^{(1-\delta)dk} \\ &\leq nk [ed\rho_j \exp\{-(1-\delta)d\rho_j\}]^k \\ &\leq nd [e(1 + \epsilon - 2\delta) \exp\{-(1-\delta)(1 + \epsilon - 2\delta)\}]^k \\ &\leq nd [(1 + \epsilon - 2\delta) \exp\{-\epsilon + 4\delta\}]^k. \end{aligned}$$

Furthermore, we have that  $1 + x \leq \exp\left\{x - \frac{x^2}{3}\right\}$  for small enough  $x$  and therefore

$$\begin{aligned} \mathbb{E}[X_k^{(j)}] &\leq nd \left[ \exp\left\{\epsilon - 2\delta - \frac{(\epsilon - 2\delta)^2}{3}\right\} \exp\{-\epsilon + 4\delta\} \right]^k \\ &\leq nd \exp\left\{-\left(\frac{\epsilon^2}{3} - 3\delta\right)k\right\} \\ &\leq nd \exp\{-\epsilon^3 \log^2 d\}, \end{aligned}$$

where we used that  $k \geq \log^2 d$  and that  $\delta$  is small enough with respect to  $\epsilon$ . By Markov's inequality, we have

$$\mathbb{P}\left[X_k^{(j)} \geq nd^3 \exp\{-\epsilon^3 \log^2 d\}\right] \leq \frac{1}{d^2}.$$

Thus, by the union bound over all the at most  $\delta d$  possible values of  $k$ , **whp** there are at most  $\delta d \cdot nd^3 \exp\{-\epsilon^3 \log^2 d\} = o_d(n)$  vertices in components of order between  $\log^2 d$  and  $\delta d$ .  $\square$

We can now bound the number of vertices in  $|V_L(G(j))|$ .

**Lemma 4.3.** *For all  $j \in \{1, 2, 3\}$ , **whp**  $|V_L(G(j))| \in [(1 - o_d(1))y(\epsilon - 2\delta)n, (1 + o_d(1))y(\epsilon)n]$ .*

*Proof.* By Lemma 4.1, we have that for all  $j \in \{1, 2, 3\}$ , **whp**

$$|V_S(G(j))| \in [(1 - o_d(1))(1 - y(\epsilon))n, (1 + o_d(1))(1 - y(\epsilon - 2\delta))n].$$

Furthermore, by Lemma 4.2, for all  $j \in \{1, 2, 3\}$ , **whp**  $|V(G) \setminus (V_S(G(j)) \cup V_L(G(j)))| = o_d(n)$ . Therefore, for all  $j \in \{1, 2, 3\}$ , **whp**

$$\begin{aligned} |V_L(G(j))| &\leq n + o_d(n) - (1 - o_d(1))(1 - y(\epsilon))n = (1 + o_d(1))y(\epsilon)n, \text{ and,} \\ |V_L(G(j))| &\geq n - o_d(n) - (1 - o_d(1))(1 - y(\epsilon - 2\delta))n = (1 - o_d(1))y(\epsilon - 2\delta)n, \end{aligned}$$

as required.  $\square$

We now turn to the task of estimating  $|W_L(G(2))|$ . We begin with a lemma bounding from above the probability that a vertex has many neighbours which do not lie in large components in  $G(1)$ . Throughout the rest of the section, we suppose we have enumerated the vertices of  $V(G)$  according to some arbitrary ordering.

**Lemma 4.4.** *Let  $v \in V(G)$ . The probability that there exists  $U \subseteq N_G(v)$  such that*

$$\left| \bigcup_{u \in U} C_{G(1)}(u) \right| \in [\delta d, 2\delta d]$$

*and  $|U| \leq \delta^2 d$  is at most  $\exp\{-\delta^2 d\}$ .*

*Proof.* We restrict ourselves to  $U' \subseteq U$ , such that  $C_{G(1)}(u)$  are disjoint for each  $u \in U'$ . Let  $F$  be a spanning forest of the components meeting  $U'$  in  $G(1)$  such that  $|V(F)| = k \in [\delta d, 2\delta d]$ . This forest is composed of some  $1 \leq \ell \leq \delta^2 d$  tree components,  $B_1, \dots, B_\ell$ , so we may assume that for every  $i \neq j$ ,  $B_i \cap B_j = \emptyset$  and all the edges leaving each  $B_i$  are not in  $G(1)$ . Each  $B_i$  contains a unique vertex  $u_i \in U'$  for  $1 \leq i \leq \ell$ . Note that if there is a subset  $U \subseteq N_G(v)$  satisfying the conditions of the lemma, then such an  $F$  exists.

Let us now bound from above the probability such a forest  $F$  exists. There are at most  $\sum_{m=1}^{\delta^2 d} \binom{d}{m} \leq \left(\frac{e}{\delta^2}\right)^{\delta^2 d}$  ways to choose  $U' \subseteq N_G(v)$  with  $|U'| \leq \delta^2 d$ . We can then specify the forest  $F$  by choosing  $|V(F)| = k \in [\delta d, 2\delta d]$ , the number  $1 \leq \ell \leq \delta^2 d$  of tree components, their sizes  $|B_i| = k_i$  such that  $\sum_{i=1}^{\ell} k_i = k$  and finally the tree components  $\{B_1, \dots, B_\ell\}$ , for which, by the second inequality in Lemma 2.3, we have at most  $\prod_{i=1}^{\ell} (ed)^{k_i-1} = (ed)^{k-\ell}$  choices (note that there is no factor of  $n$  in the estimate since the roots of the tree components are determined — these are the vertices in  $U'$ ). For a fixed forest  $F$ , there are  $k - \ell$  edges which must appear in  $G(1)$ , which happens with probability  $\rho_1^{k-\ell}$ . Since  $|V(F)| \leq 2\delta d$  there are at least  $kd(1 - 2\delta)$  edges in the boundary of  $F$  which must not appear in  $G(1)$ , which happens with probability at most  $(1 - \rho_1)^{kd(1-2\delta)}$ . Thus, by the union bound, the probability such  $F$  exists is at most

$$\left(\frac{e}{\delta^2}\right)^{\delta^2 d} \sum_{k=\delta d}^{2\delta d} \sum_{\ell=1}^{\delta^2 d} \sum_{\substack{k_1, \dots, k_\ell \in \mathbb{N} \\ k_1 + \dots + k_\ell = k}} (ed)^{k-\ell} \rho_1^{k-\ell} (1 - \rho_1)^{kd(1-2\delta)}.$$

Since  $\rho_1 \geq \frac{1+\epsilon-2\delta}{d}$  and as  $x \exp\{-ax\}$  is decreasing when  $ax > 1$ ,

$$\begin{aligned} (ed)^{k-\ell} \rho_1^{k-\ell} (1 - \rho_1)^{kd(1-2\delta)} &\leq [e(1 + \epsilon - 2\delta) \exp\{-(1 - 2\delta)(1 + \epsilon - 2\delta)\}]^k \\ &\leq \left[ \exp\left\{ \epsilon - 2\delta - \frac{(\epsilon - 2\delta)^2}{3} \right\} \exp\{-\epsilon + 5\delta\} \right]^k \\ &\leq \exp\{-\epsilon^3 k\}, \end{aligned}$$

where the penultimate inequality follows from  $1 + x \leq \exp\left\{x - \frac{x^2}{3}\right\}$  for small enough  $x$ , and the last inequality follows since  $\delta$  is sufficiently small with respect to  $\epsilon$ . There are  $\binom{k-1}{\ell-1}$  ways to choose  $k_1, \dots, k_\ell \in \mathbb{N}$  such that  $\sum_{i=1}^{\ell} k_i = k$ . Recalling that  $k \in [\delta d, 2\delta d]$  and  $\ell \in [\delta^2 d]$ , we have that  $\binom{k-1}{\ell-1} \leq \left(\frac{2e}{\delta}\right)^{\delta^2 d}$ . Altogether, the probability that such  $F$  exists is at most

$$\left(\frac{2e}{\delta^2}\right)^{2\delta^2 d} \exp\{-\epsilon^3 \delta d\} \leq \exp\left\{-\frac{\epsilon^3 \delta d}{2}\right\} \leq \exp\{-\delta^2 d\},$$

where we used the fact that  $\delta$  is sufficiently small with respect to  $\epsilon$ .  $\square$

We are now ready to bound  $|W_L(G(2))|$ . Recall that

$$W_L(G(2)) = \{w \in V(G) : |N_G(w) \cap V_L(G(2))| \geq \delta^5 d\}.$$

**Lemma 4.5.** *Whp*  $|W_L(G(2))| \geq (1 - \exp\{-\delta^6 d\})n$ .

*Proof.* Let  $v \in V(G)$ , and let  $N_G(v) = \{u_1, \dots, u_d\}$ . Let us bound the probability that  $v \notin W_L(G(2))$ . To that end, we consider a modified BFS process, during which we will keep track of the sets  $Q$  and  $W$  as in Section 2.1, as well as the set  $X_L$  of vertices in large components (of order at least  $\delta d$ ) in  $G(2)$ , and a set  $X_S$  of vertices in components which might not be large, where initially  $X_L = X_S = \emptyset$ .

The process runs in a number of *epochs*, at the start of which the queue  $Q$  is empty. We begin the next epoch by moving the first  $u_i \in N_G(v) \setminus (X_L \cup X_S)$  into  $Q$ . We then run the BFS algorithm in  $G(1)$  starting at  $u_i$ , where we copy every vertex discovered during the epoch to a set  $K$ , and we empty  $K$  at the end of every epoch. We call  $u_i$  the *root* of the epoch. We then modify the BFS algorithm in the following ways:

- If  $|K| = \delta d$  at some moment during the epoch, then once we finish exploring the connected component in  $G(1)$ , we move all the vertices in  $K$  to  $X_L$ .
- Otherwise, if at some moment during the epoch, the first vertex  $w$  in the queue  $Q$  has at least  $\delta d$  neighbours in  $X_L$ , we expose the edges in  $G_{p_2}$  between  $w$  and  $X_L$ . If any of the queries are successful, we end the epoch, move all the vertices in  $Q$  to the set  $W$  of vertices whose exploration is complete, and move all the vertices in  $K$  to  $X_L$ . Otherwise, we continue exploring the component in  $G(1)$ .
- Otherwise, if we have not moved  $u_i$  to  $X_L$  due to the first two items, and we are at the end of the epoch since  $Q = \emptyset$ , we query all the unqueried edges between  $K$  and  $X_L$  in  $G(1)$ . If any of these queries are successful, we move all the vertices in  $K$  to  $X_L$ . Otherwise, we move all the vertices in  $K$  to  $X_S$ .

We stop the process either after  $\delta^2 d$  epochs, or once  $N_G(v) \setminus (X_L \cup X_S) = \emptyset$ .

Note that if all the vertices in  $K$  moved to  $X_S$  at the end of an epoch rooted at  $u_i$ , then the component of  $u_i$  in  $G(1)$  has size at most  $\delta d$ , since all edges incident to this component in  $G(1)$ , and potentially some in  $G(2)$ , were exposed during this epoch. We thus claim that by Lemma 4.4 at the end of this process  $|X_S| \leq 2\delta d$  with probability at least  $1 - \exp\{-\delta^2 d\}$ . Indeed, let  $Y$  be the set of all roots  $u_i$  of epochs with  $u_i \in X_S$ . Since there are at most  $\delta^2 d$  epochs, we have that  $|Y| \leq \delta^2 d$ . Suppose that at the end of this process  $|X_S| > 2\delta d$ . Consider the first moment when  $|X_S| \geq \delta d$ , and let  $U \subseteq Y$  be the set of all roots  $u_i$  of epochs with  $u_i \in X_S$  at that moment. Note that since  $|X_S|$  changes in size by at most  $\delta d$  at each time, at that moment we have that  $|X_S| \leq 2\delta d$ , and thus  $|X_S| \in [\delta d, 2\delta d]$ . Therefore,  $U$  satisfies the conditions of Lemma 4.4.

Now, conditioned on  $|X_S| \leq 2\delta d$ , at the end of the process we have two cases. If we had less than  $\delta^2 d$  epochs, then the process stopped since  $N_G(v) \subseteq X_L \cup X_S$ . However then, since  $|X_S| \leq 2\delta d$ , we have that  $|X_L| \geq (1 - 2\delta)d > \delta^5 d$ , and thus  $v \in W_L(G(2))$ . We may thus assume we had  $\delta^2 d$  epochs.

We claim that, if  $|X_S| \leq 2\delta d$  at the start of an epoch, then, conditioned on the full history of the process so far, the probability that the root  $u_i$  of the epoch is moved into  $X_L$  at the end of the epoch is at least  $\frac{\delta^2}{2}$ . Indeed, assume that we have yet to discover  $\delta d$  vertices during the epoch rooted at  $u_i$ . If during this epoch no vertices with more than  $\delta d$  neighbours in  $X_L$  were discovered, then by assumption each vertex discovered during the epoch has at most  $4\delta d$  neighbours which have already been discovered when it is queried — at most  $\delta d$  from the current epoch, at most  $\delta d$  from  $X_L$  and at most  $2\delta d$  from  $X_S$ . Therefore, we can couple the BFS exploration in this epoch with a Galton-Watson tree  $B$  with offspring distribution  $\text{Bin}((1 - 4\delta)d, p_1)$ , such that the BFS exploration stochastically dominates  $B$  as long as  $|B| \leq \delta d$ . Hence, standard results imply that with probability at least  $y((1 - 4\delta)dp_1 - 1) \geq \delta$ ,  $|B|$  grows to infinity, and in particular the BFS exploration discovers at least  $\delta d$  vertices.

Conversely, if during this epoch a vertex  $w$  with at least  $\delta d$  neighbours in  $X_L$  was discovered, then, the probability that  $u_i \in X_L$  is at least the probability that one of the  $\delta d$  edges from  $w$  to  $X_L$  lies in  $G_{p_2}$ , which is at least

$$1 - (1 - p_2)^{\delta d} \geq 1 - \exp\{-\delta d p_2\} = 1 - \exp\{-\delta^2\} \geq \frac{\delta^2}{2},$$

where the last inequality follows since  $\exp\{-x\} \leq 1 - \frac{x}{2}$  for small enough  $x$ .

Hence, conditioned on the event that  $|X_S| \leq 2\delta d$ , which holds with probability at least  $1 - \exp\{-\delta^2 d\}$  by Lemma 4.4, and assuming that we had  $\delta^2 d$  epochs, then  $|X_L \cap N_G(v)|$  stochastically dominates  $\text{Bin}\left(\delta^2 d, \frac{\delta^2}{2}\right)$ . Hence, by Lemma 2.1, with probability at least  $1 - \exp\{-\delta^5 d\}$  we have that  $|X_L \cap N_G(v)| \geq \delta^5 d$ .

Therefore, putting all these together, the probability that  $|X_L| < \delta^5 d$  is at most

$$\exp\{-\delta^2 d\} + \exp\{-\delta^5 d\} \leq 2 \exp\{-\delta^5 d\}.$$

It follows that the probability that any particular  $v \in V(G)$  is not in  $W_L(G(2))$  is at most  $2 \exp\{-\delta^5 d\}$ , and so by Markov's inequality, **whp**  $|V(G) \setminus W_L(G(2))| \leq \exp\{-\delta^6 d\}n$ .  $\square$

We are now ready to prove Theorem 1.

*Proof of Theorem 1.* By Lemma 4.3, **whp**

$$|V_L(G_p)| \leq (1 + o_d(1))y(\epsilon)n, \quad |V_L(G(2))| \geq (1 - o_d(1))y(\epsilon - 2\delta)n \geq y(\epsilon)n - 5\delta n.$$

We claim that, after sprinkling with probability  $p_3$ , **whp** all but of most  $\delta n$  of the vertices in  $V_L(G(2))$  merge into a single component. Let  $A \sqcup B$  be a partition of the components forming  $V_L(G(2))$ , satisfying  $\delta n \leq k := |A| \leq |B|$ . Let  $A'$  be  $A$  together with all the vertices in  $V(G) \setminus B$  which have at least  $\frac{\delta^5 d}{2}$  neighbours in  $A$ , and let  $B'$  be  $B$  together with all the vertices in  $V(G) \setminus A'$  which have at least  $\frac{\delta^5 d}{2}$  neighbours in  $B$ . Note that  $A' \cap B' = \emptyset$  and  $V_L(G(2)) \cup W_L(G(2)) \subseteq A' \cup B'$ .

Fix the partition  $A \sqcup B$  of (the components forming)  $V_L(G(2))$ . Note that  $|A'|, |B'| \geq k$  and one of them is of order at most  $\frac{n}{2}$ . We may assume without loss of generality that  $k \leq |A'| \leq \frac{n}{2}$ . Thus, by our assumption on the isoperimetric constant of  $G$ , that is,  $i(G) \geq C$ , there are at least  $Ck$  edges in  $G$  between  $A'$  and  $(A')^C$ . Since  $G$  is  $d$ -regular, and by Lemma 4.5 **whp**  $|V(G) \setminus W_L(G(2))| \leq \exp\{-\delta^6 d\}n = o(n/d)$ , it follows that  $e_G(A', B') \geq Ck/2$ .

Let us now expose these edges (between  $A'$  and  $B'$ ) with probability  $p_3$ . By Lemma 2.4, with probability at least  $1 - \exp\left\{-\frac{\delta^2 Ck}{d}\right\}$  there is a matching of size at least  $\frac{\delta^2 Ck}{d}$  between  $A'$  and  $B'$  in  $G_{p_3}$ .

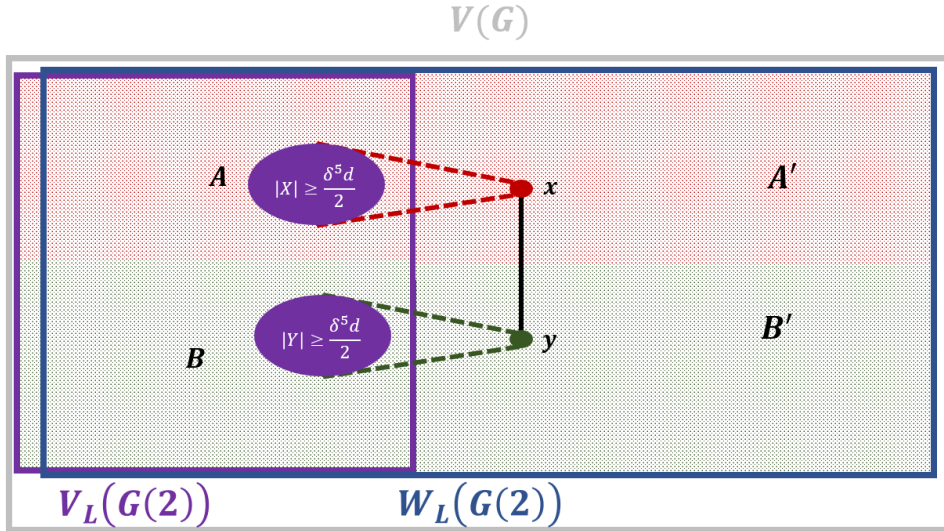


Figure 1: Illustration of the sets utilised in the proof of Theorem 1 and their properties. By Lemma 4.5, **whp**  $|W_L(G(2))| = n - o(n/d)$  (appears in dark blue).  $V_L(G(2))$  appears in purple, together with a partition of it into  $A$  (red-dotted) and  $B$  (green-dotted). This extends to  $A'$  (red-dotted) and  $B'$  (green-dotted), which cover  $W_L(G(2))$ . In the figure, the edge  $xy$  between  $A'$  and  $B'$  appears in solid black.  $X$ , the neighbourhood of  $x$  in  $A$ , appears in purple, and its size is lower-bounded by construction. Similarly,  $Y$ , the neighbourhood of  $y$  in  $B$ , appears in purple with its size also lower-bounded by construction. When sprinkling with probability  $p_3$ , with very high probability there will be many black edges between  $A'$  and  $B'$  in  $G_{p_3}$ , which we can then extend to paths of length three between  $A$  and  $B$ .



Assume such a matching exists. By the definition of  $A'$  and  $B'$ , from the endpoints of each edge in the matching there are at least  $\frac{\delta^5 d}{2}$  edges into  $A$ , and at least  $\frac{\delta^5 d}{2}$  edges into  $B$ . The probability that a given edge in the matching does not extend to a path of length three between  $A$  and  $B$  is thus at most  $2(1 - p_3)^{\delta^5 d/2} \leq 2 \exp\{-\delta^6/2\} \leq \exp\{-\delta^7\}$ . These events are independent for the edges of the matching, and thus the probability that there is no path in  $G_{p_3}$  connecting  $A$  and  $B$  is at most

$$\exp\left\{-\frac{\delta^7 \cdot \delta^2 Ck}{d}\right\} \leq \exp\left\{-\frac{C\delta^{10}n}{d}\right\}.$$

Since there are at most  $2^{\frac{n}{\delta d}}$  possible component-respecting partitions  $A \sqcup B$  of  $V_L(G(2))$ , by choosing  $C_0 = \frac{2}{\delta^{20}}$  (and by our assumption that  $C \geq C_0$ ), we conclude that with probability at least

$$1 - o_d(1) - \exp\left\{-\frac{C\delta^{10}n}{d} + \frac{n}{\delta d}\right\} \geq 1 - o_d(1) - \exp\left\{-\frac{\sqrt{C}n}{d}\right\},$$

$G_p$  contains a component of order at least  $y(\epsilon)n - 5\delta n - \delta n = y(\epsilon)n - \frac{10}{C^{1/20}}n$ , as required. Furthermore, with the same probability, the second-largest component is of order at most  $(1 + o_d(1))\frac{10n}{C^{1/20}}$ .  $\square$

**Remark 4.6.** *A careful reading of the proof of Theorem 1 shows that we do not use the full extent of our assumption  $i(G) \geq C$ . In fact, a milder requirement suffices: for any  $S \subseteq V(G)$  with  $\delta n \leq |S| \leq \frac{n}{2}$ , we require that  $\partial_G(S) \geq C\delta n$ . Furthermore, we note that it suffices to choose  $\delta = \frac{\epsilon^2}{10}$ , and so we can take  $C_0 = \Omega(\delta^{-20}) = \Omega(\epsilon^{-40})$ . Finally, we note that key techniques of the proof of Theorem 1 can easily be adapted to prove that a percolated hypercube  $Q_p^d$  in the supercritical regime, when  $p = \frac{1+\epsilon}{d}$ , typically contains a unique giant component of the correct asymptotic order, namely,  $y(\epsilon)|V(Q^d)|$ . Indeed, by Harper's inequality [26], which states that  $i_k(Q^d) \geq d - \log_2 k$ , we have that the edge-boundary of an arbitrary set  $S$  of order  $|S| = \frac{|V(Q^d)|}{2^C} = 2^{d-C}$  has size at least  $C|S|$ . Following the lines of our proof arguments of Theorem 1, we can show that **whp**  $Q_p^d$  contains constantly many components of linear order, whose sizes sum up asymptotically to  $y(\epsilon)|V(Q^d)|$ . Using a standard sprinkling argument, we can then show that **whp** these linear sized components merge into a single giant component.*

## 5 Tightness of Theorem 1

We begin by describing the construction which will be used for the proof of Theorem 2.

### 5.1 Construction for Theorem 2

**First construction** Let  $d, d_1$ , and  $n$  be even positive integers satisfying the following conditions:

$$d = \omega(1) \text{ is such that } d+2 \text{ divides } n, \quad d_1 = \frac{d}{Ce^2}, \text{ and } n = \omega(d^2). \quad (5)$$

Let  $H_0$  be an  $\left(\frac{n}{d+2}, d_1, \lambda\right)$ -graph with  $\lambda = O(\sqrt{d_1})$  (note that  $d_1 \cdot \frac{n}{d+2}$  is even, since  $d_1$  is even and  $d+2$  divides  $n$ ). Indeed, **whp** a random  $d_1$ -regular graph on  $\frac{n}{d+2}$  vertices satisfies this (see [33], [37], [36] and the references therein for even stronger results). Note that, by the expander mixing lemma (see [5]), it follows that  $i(H_0) \geq \frac{d_1}{3}$ .

For every  $v \in V(H_0)$ , let  $F_0(v)$  be a  $(d+1)$ -clique and let  $M(v) \subseteq E(F_0(v))$  be an arbitrary matching of size  $\frac{d-d_1}{2}$ . Let  $F_1(v) = F_0(v) - M(v)$  be the graph obtained by deleting the edges of this matching from the clique.

We now consider the graph  $G$  formed by taking the union of  $H_0$  and  $\bigcup_{v \in V(H_0)} F_1(v)$  and joining each  $v \in V(H_0)$  to each vertex in  $V(M(v)) \subseteq V(F_1(v))$  (see Figure 2 for an illustration of  $G$ ).

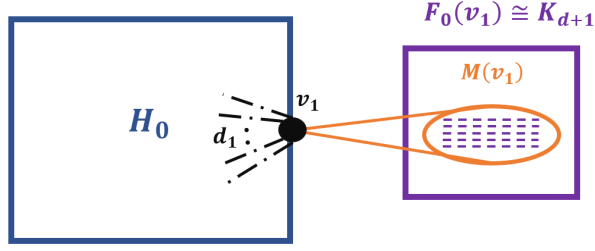


Figure 2: An illustration of the **First construction**

Let us first show that our graph  $G$  satisfies the assumptions of Theorem 2, noting that  $\frac{d_1}{5d} \geq \frac{1}{5Ce^2} \geq \frac{1}{40C}$ .

**Claim 5.1.** *Let  $G$  be the graph built according to the **First construction**. Then  $G$  is a  $d$ -regular  $n$ -vertex graph with  $i(G) \geq \frac{d_1}{5d}$ .*

*Proof.* We first note that  $G$  is  $d$ -regular. Indeed, every vertex  $v \in V(H_0)$  has  $d_1$  neighbours in  $H_0$ , and  $d - d_1$  neighbours in  $V(M(v)) \subseteq V(F_1(v))$ . Furthermore, every vertex in  $V(F_1(v)) \setminus V(M(v))$  has  $d$  neighbours in  $G$ , since  $F_0(v) \cong K_{d+1}$  and the degree of vertices not in  $V(M(v))$  has not been changed.

Furthermore,  $G$  has  $n$  vertices. Indeed,  $|V(H_0)| = \frac{n}{d+2}$  and for each  $v \in V(H_0)$ ,  $|F_1(v)| = d+1$ . Therefore,

$$|V(G)| = \frac{n}{d+2} + (d+1) \frac{n}{d+2} = n.$$

Let us now show that  $i(G) \geq \frac{d_1}{5d}$ . Given  $S \subseteq V(G)$  with  $|S| \leq \frac{n}{2}$ , let  $S_1 = S \cap V(H_0)$  and for every  $v \in V(H_0)$ , let  $S_F(v) = S \cap (\{v\} \cup V(F_1(v)))$ .

Suppose first that  $|S_1| \geq \frac{4|S|}{5(d+1)}$  and  $|S_1| \leq \frac{|V(H_0)|}{2}$ . Then, since

$$\frac{e(S, S^C)}{|S|} \geq \frac{e_{H_0}(S_1, S_1^C)}{|S|},$$

it follows from the fact  $i(H_0) \geq \frac{d_1}{3}$  that

$$\frac{e_{H_0}(S_1, S_1^C)}{|S|} \geq \frac{\frac{d_1|S_1|}{3}}{|S|} \geq \frac{d_1}{4d}.$$

If  $|S_1| \geq \frac{4|S|}{5(d+1)}$  and  $\frac{|V(H_0)|}{2} \leq |S_1| \leq \frac{2|V(H_0)|}{3}$ , then

$$\frac{e_{H_0}(S_1, S_1^C)}{|S|} \geq \frac{\frac{d_1|S_1^C|}{3}}{|S|} \geq \frac{d_1|V(H_0)|}{9|S|} \geq \frac{2d_1|V(H_0)|}{9(d+2)|V(H_0)|} \geq \frac{d_1}{5d}.$$

Now, suppose that  $|S_1| < \frac{4|S|}{5(d+1)}$ . Then, we may assume that there is some subset  $S_2 \subseteq S$  of size at least  $\frac{|S|}{5}$  which lies in some union of  $S_F(v)$  where  $v \notin S$ . Then, since  $S = \bigcup_{v \in V(H_0)} S_F(v)$  we can see that

$$\frac{e(S, S^C)}{|S|} \geq \frac{1}{5} \min_{v \in S_2} \frac{e(S_F(v), S^C)}{|S_F(v)|}. \quad (6)$$

However, for each  $v \in S_2$  we have that  $S_F(v)$  meets  $F_1(v)$  and does not contain  $v$ . Since  $F_1(v)$  is a  $(d+1)$ -clique with a matching removed, we have that

$$\begin{aligned} e(S_F(v), S^C) &\geq e(S_F(v), F_1(v) \setminus S_F(v)) \\ &\geq |S_F(v)| (d+1 - |S_F(v)|) - |S_F(v)| = |S_F(v)| (d - |S_F(v)|). \end{aligned}$$

Note that as long as  $|S_F(v)| < d$ , the above is at least  $|S_F(v)|$ . On the other hand, if  $|S_F(v)| = d$ , then since  $S_F(v)$  does not contain  $v$  and  $v$  has  $d - d_1$  neighbours in  $F_1(v) \setminus \{v\}$ , we have that

$$e(S_F(v), S^C) \geq e(v, S_F(v)) \geq d - d_1 - 1.$$

Hence, in this case by (6), we have

$$\frac{e(S, S^C)}{|S|} \geq \frac{1}{5} \cdot \frac{d - d_1 - 1}{d} \geq \frac{d_1}{5d},$$

where the last inequality holds by our assumption on  $d_1$ .

Finally, suppose that  $|S_1| \geq \frac{2|V(H_0)|}{3}$ . Then, since  $|S| \leq \frac{n}{2}$ , there is a set  $S'_1 \subseteq S_1$ ,  $|S'_1| \geq \frac{|V(H_0)|}{6}$ , such that for every  $v \in S'_1$ ,  $S_F(v) \cap F_1(v) \neq F_1(v)$ . In particular, each such copy contributes at least  $d - d_1$  edges to the boundary of  $S$ . Therefore, we get

$$\frac{e(S, S^C)}{|S|} \geq \frac{(d - d_1) |V(H_0)|}{6|S|} \geq \frac{d - d_1}{3(d+2)} \geq \frac{d_1}{5d},$$

where once again the last inequality holds by our assumption on  $d_1$ .  $\square$

Note that  $i(G) = \Theta\left(\frac{d_1}{d}\right)$ . Indeed, let  $S = \{v\} \cup F_1(v)$ . Then  $|S| = d + 2$  and  $\partial_G(S) = d_1$ , and thus  $i(G) \leq \frac{d_1}{d+2}$ .

Theorem 2 will be an immediate corollary of the following Lemma.

**Lemma 5.2.** *Let  $G$  be the graph built according to the [First construction](#). Let  $p = \frac{C}{d}$ . Then, **whp** all the components of  $G_p$  have size at most  $3d \log n$ .*

*Proof of Lemma 5.2.* Let  $k \geq 3d \log n$  and let  $\mathcal{A}_k$  be the event that  $G_p$  contains a component of size  $k$ . Note that if  $\mathcal{A}_k$  occurs, then there is a tree  $T$  of size  $k$  in  $G$ , of all whose edges lie in  $G_p$ . It follows that there is a subtree  $T' \subseteq T$  of size at least  $k_0 := \frac{k}{d+1} \geq 2 \log n$ , such that  $T' \subseteq H_0 \subseteq G$ , all of whose edges lie in  $G_p$ . Recalling that  $H_0$  is a  $d_1$ -regular graph on  $\frac{n}{d+2}$  vertices, by the second inequality of Lemma 2.3 there are at most  $\frac{n}{d+2} (ed_1)^{k_0-1}$  ways to choose  $T'$  (where we considered the at most  $\frac{n}{d+2}$  possible roots), and we retain its edges in  $G_p$  with probability  $p^{k_0-1}$ . Therefore, since  $pd_1 = \frac{1}{e^2}$ , we have

$$\mathbb{P}[\mathcal{A}_k] \leq \frac{n}{d+2} \cdot (ed_1)^{k_0-1} p^{k_0-1} = \frac{n}{d+2} \cdot \left(\frac{1}{e}\right)^{k_0-1} = o\left(\frac{1}{n}\right).$$

Hence, by the union bound, **whp**  $\mathcal{A}_k$  does not hold for all  $3d \log n \leq k \leq n$ .  $\square$

## 5.2 Construction for Theorem 3

**Second construction** Let  $C, d, n$ , and  $p$  be as in the statement of Theorem 3. Let  $C_1 := 3C$ , let  $d_1 := d - C_1$  and let  $t := \frac{n}{d_1+1}$ . We further assume that  $C_1, d_1, t$ , and  $n$  satisfy the needed parity assumptions for what follows.

Let  $H$  be a  $C_1$ -regular graph on  $n$  vertices, with  $i(H) \geq \frac{C_1}{3} = C$  (indeed, **whp** a random  $C_1$ -regular graph on  $n$  vertices satisfies this). Since  $t = \frac{n}{d_1+1} = \frac{n}{d-3C+1} \geq \frac{n}{d} \geq 10C$ , we conclude that there exists an equitable (proper) colouring of  $H$  in  $t$  colours,  $A_1, \dots, A_t$ , with each colour class containing exactly  $d_1 + 1$  vertices [25]. We form  $G$  by adding to  $H$  all the possible edges in  $H[A_j]$ , that is,  $G[A_j] \cong K_{d_1+1}$ , for every  $j \in [t]$ . Since  $d = d_1 + C_1$ , note that  $G$  is a  $d$ -regular graph on  $t(d - 3C + 1)$  vertices. Furthermore,  $i(G) \geq i(H) \geq C$ .

With this construction at hand, we are ready to prove Theorem 3.

*Proof of Theorem 3.* Note that, for every  $j \in [t]$ , the edges between  $A_j$  and  $V(G) \setminus A_j$  are those in  $H$ . Let  $X$  be the number of sets  $A \in \{A_1, \dots, A_t\}$ , such that  $e_{H_p}(A, V(H) \setminus A) = 0$ .

For each fixed  $j \in [t]$  we have that  $e_H(A_j, V(H) \setminus A_j) = C_1(d_1 + 1)$ . The probability this set is disjoint from the rest of the graph in  $H_p$  is

$$(1 - p)^{C_1(d_1+1)} \geq \exp \left\{ -2 \frac{C_1(d - C_1 + 1)}{d} \right\} \geq \exp \{-4C_1\},$$

where we used  $d \geq 7C = \frac{7C_1}{3}$ . Hence,  $\mathbb{E}[X] \geq \exp\{-4C_1\}t$ . Now, note that changing one edge can change the value of  $X$  by at most two. Hence, by Lemma 2.2, we have

$$\mathbb{P} \left[ X \leq \frac{\mathbb{E}[X]}{2} \right] \leq 2 \exp \left\{ -\frac{\exp\{-8C_1\}t^2/4}{2 \cdot C_1 n p \cdot 4} \right\} \leq \exp \left\{ -\frac{\exp\{-8C_1\}t}{12} \right\}.$$

Thus, with probability at least  $1 - \exp\{-\exp\{-10C_1\}t\} \geq 1 - \exp\left\{-\frac{n}{\exp\{30C\}d}\right\}$ , there are at least  $\frac{\exp\{-4C_1\}t}{2}$  disjoint sets  $A \in \{A_1, \dots, A_t\}$  in  $G_p$ . Since  $G[A_j] \cong K_{d_1+1}$  for every  $j \in [t]$ , we have that with probability  $1 - o_d(1)$  there exists a component of order  $(2\epsilon - O(\epsilon^2))d_1$  in  $G_p[A_j]$ . Therefore, with probability at least  $1 - \exp\left\{-\frac{n}{\exp\{30C\}d}\right\} - o_d(1)$ , there are at least two components of order  $(2\epsilon - O(\epsilon^2))d_1 \geq \epsilon d$  in  $G_p$ , where we used our assumption that  $d \geq 7C$ .  $\square$

## 6 Discussion and avenues for future research

In this paper, we aimed to capture the minimal requirements on a host graph  $G$ , guaranteeing the percolated subgraph  $G_p$  undergoes a similar phase transition, with respect to the size of the largest component, as that in  $G(n, p)$ . Theorem 1, together with the two constructions (Theorems 2 and 3) showing that it is qualitatively tight, gives a qualitative answer to this question. In Theorems 3.1 and 4, we further demonstrated the intrinsic connection between the isoperimetry of the host graph, both global and local, to the existence of large components in the percolated subgraph.

There are several natural questions which arise in this context. While Theorem 1 gives a qualitative answer to Question 1.2, one can ambitiously aim for a more precise *quantitative* answer as to the minimum requirements on  $i(G)$ , or a more restricted notion of expansion, which guarantees this behaviour.

**Question 6.1.** *Given  $\epsilon > 0$  and  $d \rightarrow \infty$ , what is the minimal  $C(\epsilon, d)$  such that every  $d$ -regular  $n$ -vertex graph  $G$  with  $i(G) \geq C(\epsilon, d)$  is such that **whp**  $G_p$  contains a unique component of linear order, specifically one of order asymptotic  $y(\epsilon)n$ , when  $p = \frac{1+\epsilon}{d}$ ?*

Theorem 1 implies that any function  $C(\epsilon, d) = \omega(1)$  is sufficient, and Theorem 2 implies that it is necessary to take  $C(\epsilon, d) \geq \frac{1}{40(1+\epsilon)}$ . It would be interesting to determine if the answer is in fact independent of  $d$  (or even  $\epsilon$ ). Furthermore, we note that the regularity assumption in Theorem 1 can be somewhat weakened. With slight technical care, we can assume instead that  $\frac{\Delta(G)}{\delta(G)} = 1 + o_d(1)$ , where  $\Delta(G)$  is the maximum degree of  $G$  and  $\delta(G)$  is the minimum degree of  $G$ . It could be interesting to determine the (irregular) degree assumptions that are necessary for Theorem 1 to hold.

It is known (see, for example, [16, Theorem 1]) that in subcritical percolation on any  $d$ -regular graph the largest component typically has logarithmic order. Theorem 1.1 demonstrates that in  $G(n, p)$  the second-largest component is of logarithmic order in the supercritical regime, which is a particularly simple facet of the *duality principle* which holds for  $G(n, p)$ , which broadly says if we remove the giant component from a supercritical random graph, what remains resembles a subcritical random graph. Similar behaviour, at least in terms of the order of the second-largest component, is known to hold in many other percolation models such as percolation on

$(n, d, \lambda)$ -graphs [23] or on hypercubes [11] (and in fact on any regular high-dimensional product graph [16]).

Whilst Theorem 1 shows that relatively weak assumptions on the expansion of  $G$  are sufficient to demonstrate a threshold for the existence of a linear sized component, Theorem 3 shows that such assumptions alone are not sufficient to effectively bound the size of the second-largest component in the supercritical regime.

It is an interesting open problem to determine natural conditions on the host graph which guarantee that the phase transition quantitatively resembles that of  $G(n, p)$ .

**Question 6.2.** *Let  $G$  be a  $d$ -regular  $n$ -vertex graph with  $d \rightarrow \infty$ . Let  $\epsilon > 0$  be a sufficiently small constant, and let  $p = \frac{1+\epsilon}{d}$ . What are the minimal requirements on  $G$  such that **whp** the largest component in  $G_p$  is of asymptotic order  $y(\epsilon)n$ , and the second-largest component is of order  $O(\log n)$ ?*

It is worth noting that whilst the assumptions of Theorem 4 impose a restriction on the *minimum* degree of the host graph  $G$ , an averaging argument shows that the assumptions of Theorem 3.1 impose a restriction only on the *average* degree of  $G$ . Motivated by their work on percolation on graphs with large minimum degree, Krivelevich and Samotij [32] ask whether there is a quantitatively similar threshold for the existence of a large component in graphs with large average degree. In particular, if  $G$  has average degree  $d$  and  $p = \frac{1+\epsilon}{d}$ , they ask if **whp**  $G_p$  contains a component (or indeed, even a cycle) with  $\Omega(d)$  vertices.

As we mentioned before, our argument of Theorem 1 allows also for  $\epsilon := \epsilon(d)$  which tends *slowly* to zero as  $d$  tends to infinity. We have not tried to optimise this dependency. It could be very interesting to analyse the correct dependency, and in particular to see what could be said about the size of the largest component, and of the component sizes in general, in the slightly supercritical regime.

## References

- [1] M. Ajtai, J. Komlós, and E. Szemerédi. Largest random component of a  $k$ -cube. *Combinatorica*, 2(1):1–7, 1982.
- [2] Y. Alimohammadi, C. Borgs, and A. Saberi. Algorithms using local graph features to predict epidemics. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3430–3451. SIAM, 2022.
- [3] Y. Alimohammadi, C. Borgs, and A. Saberi. Locality of random digraphs on expanders. *Ann. Probab.*, 51(4):1249–1297, 2023.
- [4] N. Alon, I. Benjamini, and A. Stacey. Percolation on finite graphs and isoperimetric inequalities. *Ann. Probab.*, 32(3):1727–1745, 2004.
- [5] N. Alon and F. R. K. Chung. Explicit construction of linear sized tolerant networks. *Discrete Mathematics*, 72(1-3):15–19, 1988.
- [6] N. Alon and J. H. Spencer. *The probabilistic method*. Hoboken, NJ: John Wiley & Sons, fourth edition, 2016.
- [7] I. Benjamini, S. Boucheron, G. Lugosi, and R. Rossignol. Sharp threshold for percolation on expanders. *Ann. Probab.*, 40(1):130–145, 2012.
- [8] I. Benjamini, A. Nachmias, and Y. Peres. Is the critical percolation probability local? *Probab. Theory Relat. Fields*, 149(1-2):261–269, 2011.

- [9] A. Beveridge, A. Frieze, and C. McDiarmid. Random minimum length spanning trees in regular graphs. *Combinatorica*, 18(3):311–333, 1998.
- [10] B. Bollobás. *Random graphs*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, second edition, 2001.
- [11] B. Bollobás, Y. Kohayakawa, and T. Łuczak. The evolution of random subgraphs of the cube. *Random Structures Algorithms*, 3(1):55–90, 1992.
- [12] B. Bollobás and O. Riordan. *Percolation*. Cambridge University Press, Cambridge, 2006.
- [13] C. Borgs and G. Zhao. Locality via global ties: Stability of the 2-core against misspecification. *arXiv:2304.06170*, 2023.
- [14] S. B. Broadbent and J. M. Hammersley. Percolation processes. I: Crystals and mazes. *Proc. Camb. Philos. Soc.*, 53:629–641, 1957.
- [15] J. Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. *Probl. Analysis, Sympos. in Honor of Salomon Bochner, Princeton Univ. 1969, 195-199 (1970)*., 1970.
- [16] S. Diskin, J. Erde, M. Kang, and M. Krivelevich. Percolation on high-dimensional product graphs. *arXiv:2209.03722*, 2022.
- [17] S. Diskin, J. Erde, M. Kang, and M. Krivelevich. Isoperimetric inequalities and supercritical percolation on high-dimensional graphs. *Combinatorica*, to appear, doi:10.1007/s00493-024-00089-0.
- [18] R. Durrett. *Probability: theory and examples*. Cambridge University Press, Cambridge, 2019.
- [19] P. Easo. Existence of a percolation threshold on finite transitive graphs. *Int. Math. Res. Not.*, 2023(21):18781–18802, 2023.
- [20] P. Easo and T. Hutchcroft. Supercritical percolation on finite transitive graphs I: Uniqueness of the giant component. *Duke Mathematical Journal*, to appear, arXiv:2112.12778.
- [21] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.
- [22] A. Frieze and M. Karoński. *Introduction to random graphs*. Cambridge University Press, Cambridge, 2016.
- [23] A. Frieze, M. Krivelevich, and R. Martin. The emergence of a giant component in random subgraphs of pseudo-random graphs. *Random Structures Algorithms*, 24(1):42–50, 2004.
- [24] G. Grimmett. *Percolation*. Springer, Berlin, 1999.
- [25] A. Hajnal and E. Szemerédi. Proof of a conjecture of P. Erdős. In *Combinatorial theory and its applications, I-III (Proc. Colloq., Balatonfüred, 1969)*, volume 4 of *Colloq. Math. Soc. János Bolyai*, pages 601–623. North-Holland, Amsterdam-London, 1970.
- [26] L. H. Harper. Optimal assignments of numbers to vertices. *SIAM J. Appl. Math.*, 12:131–135, 1964.
- [27] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561, 2006.
- [28] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, 2000.

- [29] H. Kesten. *Percolation theory for mathematicians*. Birkhäuser, Boston, MA, 1982.
- [30] M. Krivelevich. Expanders — how to find them, and what to find in them. In *Surveys in combinatorics 2019*, volume 456 of *London Math. Soc. Lecture Note Ser.*, pages 115–142. Cambridge Univ. Press, Cambridge, 2019.
- [31] M. Krivelevich, E. Lubetzky, and B. Sudakov. Asymptotics in percolation on high-girth expanders. *Random Structures & Algorithms*, 56(4):927–947, 2020.
- [32] M. Krivelevich and W. Samotij. Long paths and cycles in random subgraphs of  $\mathcal{H}$ -free graphs. *Electron. J. Combin.*, 21(1):Paper 1.30, 2014.
- [33] M. Krivelevich and B. Sudakov. Pseudo-random graphs. In *More sets, graphs and numbers*, volume 15 of *Bolyai Soc. Math. Stud.*, pages 199–262. Springer, Berlin, 2006.
- [34] M. Krivelevich and B. Sudakov. The phase transition in random graphs: A simple proof. *Random Structures Algorithms*, 43(2):131–138, 2013.
- [35] A. Nachmias and Y. Peres. Critical percolation on random regular graphs. *Random Structures Algorithms*, 36(2):111–148, 2010.
- [36] A. Sarid. The spectral gap of random regular graphs. *Random Structures Algorithms*, 63(2):557–587, 2023.
- [37] K. Tikhomirov and P. Youssef. The spectral gap of dense random regular graphs. *Ann. Probab.*, 47(1):362–419, 2019.