Combinatorics, Probability and Computing (2014) 23, 477–504. © Cambridge University Press 2014
doi:10.1017/S0963548314000248

On the Spread of Random Graphs

LOUIGI ADDARIO-BERRY[1], SVANTE JANSON[2]
and COLIN McDIARMID[3]

[1]Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street West,
Montréal, Québec, H3A 2K6, Canada
(e-mail: louigi@gmail.com
http://www.math.mcgill.ca/~louigi/)

[2]Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden
(e-mail: svante.janson@math.uu.se
http://www.math.uu.se/~svante/)

[3]Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK
(e-mail: cmcd@stats.ox.ac.uk
http://www.stats.ox.ac.uk/~cmcd/)

Received 10 August 2012; revised 13 March 2014

The *spread* of a connected graph G was introduced by Alon, Boppana and Spencer [1], and measures how tightly connected the graph is. It is defined as the maximum over all Lipschitz functions f on $V(G)$ of the variance of $f(X)$ when X is uniformly distributed on $V(G)$. We investigate the spread for certain models of sparse random graph, in particular for random regular graphs $G(n,d)$, for Erdős–Rényi random graphs $G_{n,p}$ in the supercritical range $p > 1/n$, and for a 'small world' model. For supercritical $G_{n,p}$, we show that if $p = c/n$ with $c > 1$ fixed, then with high probability the spread of the giant component is bounded, and we prove corresponding statements for other models of random graphs, including a model with random edge lengths. We also give lower bounds on the spread for the barely supercritical case when $p = (1 + o(1))/n$. Further, we show that for d large, with high probability the spread of $G(n,d)$ becomes arbitrarily close to that of the complete graph K_n.

2010 *Mathematics subject classification*: 60C05

1. Introduction

If G is a graph, a *Lipschitz function* f on G is a real-valued function defined on the vertex set $V(G)$ such that $|f(v) - f(w)| \leqslant 1$ for every pair of adjacent vertices v and w. We may regard a function $f : V(G) \to \mathbb{R}$ on a graph G as a random variable by evaluating f at a random, uniformly distributed, vertex. We may thus talk about the mean, median and variance of f. For example, if G has n vertices, the mean $\mathbb{E} f$ is $\sum_v f(v)/n$, and the variance of f is

$$\frac{1}{n}\sum_{v}(f(v)-\mathbb{E} f)^2 = \frac{1}{2n^2}\sum_{v,w}(f(v)-f(w))^2. \tag{1.1}$$

For a fixed connected graph G, we define the *spread* of G to be the supremum of the variance of f over all Lipschitz functions f on G, and we denote this quantity by spread(G). (Note that the supremum would be infinite if we considered a disconnected graph.) The spread of a graph was introduced by Alon, Boppana and Spencer in [1], and considered further in [3]. In particular it is shown in [1] that the spread yields the optimal coefficient in the exponent in a natural asymptotic isoperimetric inequality: we discuss this briefly below. The spread is a natural measure of the overall connectivity of a graph, and the purpose of this paper is to investigate the spread for certain models of random graph.

1.1. The spread of a graph

Before we introduce our results concerning random graphs, let us give some background on the spread of a graph. Observe first that spread is an edge-monotone function in the sense that if we add an edge to a graph then the set of Lipschitz functions becomes smaller, and thus the spread becomes smaller or remains the same.

For every connected graph G, spread(G) is attained, so we can replace supremum by maximum. In fact, it is shown in Theorem 2.1 of [1] that there is always a Lipschitz function f achieving spread(G) which is integer-valued and of the following simple form: if S denotes the set of vertices v with $f(v)=0$, then each component H of $G\setminus S$ has a sign $g(H)=\pm 1$, and for each vertex v in such an H, $f(v)$ is $g(H)$ times the graph distance between v and S. An integer-valued Lipschitz function on G may be regarded as a homomorphism from G to a suitably long path with a loop at each vertex, and spread(G) measures how widely distributed along the path we can make the images of the vertices of G.

It is easy to see that the complete graph K_n has spread $1/4$ if n is even and $1/4-1/(4n^2)$ if n is odd. This of course gives the minimum possible values of the spread for graphs of order n. The maximum is $(n^2-1)/12$, attained by the path P_n.

Denote the graph distance between vertices u and v by $d_G(u,v)$, and let diameter(G) be the maximum value of $d_G(u,v)$. It is easily seen from (1.1) that

$$\mathrm{spread}(G) \leqslant \frac{1}{4}\,\mathrm{diameter}(G)^2,$$

and similarly that

$$\mathrm{spread}(G) \leqslant \frac{1}{2n^2}\sum_{v,w} d_G(v,w)^2,$$

so the spread is at most half the mean squared distance between vertices. Our results will imply that the spread is typically much smaller for random graphs.

Given a list of graphs $G_1,\dots,G_d$, the *Cartesian product* $\prod_i G_i$ is the graph with vertex set $\prod_i V(G_i)$, in which two vertices $(u_1,\dots,u_d)$ and $(v_1,\dots,v_d)$ are adjacent if and only if they differ in exactly one coordinate i and u_i and v_i are adjacent in G_i. It is implicit in [1] and explicit in [3] that, assuming the G_i are connected,

$$\operatorname{spread}\left(\prod_i G_i\right) = \sum_i \operatorname{spread}(G_i).$$

For example, the hypercube Q^d is K_2^d (the product of d copies of K_2); since $\operatorname{spread}(\mathsf{K}_2) = 1/4$ we see that $\operatorname{spread}(\mathsf{Q}^d) = d/4$.

Alon, Boppana and Spencer in [1] considered the case of a fixed connected graph G, and were interested in tight isoperimetric inequalities concerning G^d for large d. Given a graph H, a set S of vertices of H and $t > 0$, let $B(S,t)$ denote the set of vertices at distance at most t from S (the *t-ball around S*), and let

$$g(H,t) = \max_{|S|\geqslant|V(H)|/2} \frac{|V(H)\setminus B(S,t)|}{|V(H)|},$$

where the maximum is over subsets S of at least half the vertices of H. Thus $g(H,t)$ is the maximum proportion of vertices at distance $> t$ from a set of at least half the vertices. From Theorem 1.1 in [1] (which gives a more general result), we have the following.

Theorem 1.1. *Let G be a connected graph and let $\gamma = \operatorname{spread}(G)$. Then, for $d^{1/2} \ll t \ll d$,*

$$g(G^d,t) = e^{-\frac{t^2}{2d\gamma}(1+o(1))} \quad \text{as } d\to\infty.$$

1.2. Our results on the spread of random graphs

We use w.h.p. (*with high probability*) for events with probability $1-o(1)$ as $n\to\infty$. Our focus is on whether or not the spread is bounded w.h.p. in various models of sparse random graph. In these models typical degrees are small and w.h.p. the mean path length is $\Theta(\log n)$ (and so the mean squared path length is $\Omega(\log^2 n)$): see for example Durrett [7] for results on diameter and mean path length, and the discussion at the end of Section 4. (We use log to denote natural logarithm, though often the base is irrelevant, as in $O(\log n)$.)

We start with random regular graphs $G(n,d)$ with fixed degree $d\geqslant 3$, as that is the easiest case. It is well known that w.h.p. $G(n,d)$ is connected [4], and so we may talk of $\operatorname{spread}(G(n,d))$. In Section 2 we prove that such graphs with high probability have uniformly bounded spread.

Theorem 1.2. *There exists a constant C_1 such that, for every fixed $d\geqslant 3$, w.h.p.*

$$\operatorname{spread}(G(n,d)) \leqslant C_1.$$

In fact we prove a stronger result, Theorem 2.1, giving an exponential tail inequality for Lipschitz functions; we derive this from a corresponding deterministic result for expander graphs, Theorem 2.5.

In Section 3 we study the random graph $G_{n,c/n}$ with fixed $c > 1$, the supercritical case. This random graph is w.h.p. disconnected, so we consider the spread of the largest component of $G_{n,c/n}$, which we denote by $H_{n,c/n}$. (Recall that for $c > 1$, there is w.h.p. a unique giant component $H_{n,c/n}$ of order $\sim \gamma(c)n$ for some $\gamma(c) > 0$.) It was shown in [19] that there is a $f(c) > 0$ such that the diameter of $H_{n,p}$ is $f(c)\log n + O_p(1)$. However, the spread stays bounded.

Theorem 1.3. *For each fixed $c > 1$ there exists a constant $C_2 = C_2(c) > 0$ such that w.h.p.* $\mathrm{spread}(H_{n,c/n}) \leqslant C_2$.

As with Theorem 1.2 above, we actually prove a stronger result, Theorem 3.1, giving a tail inequality for Lipschitz functions which is exponential in $\sqrt{n}$.

In Section 4 we study the random graph $G_{n,c/n}$ in the barely supercritical case when $c = 1 + \varepsilon$, and show that the spread tends to infinity (in probability) at least at the rate ε^{-2}.

Theorem 1.4. *Let $p = (1 + \varepsilon)/n$ with $\varepsilon = \varepsilon(n) \to 0$ and $\varepsilon^3 n \to \infty$ as $n \to \infty$. Then w.h.p. the giant component $H_{n,p}$ of $G_{n,p}$ satisfies*

$$\mathrm{spread}(H_{n,p}) = \Omega(1/\varepsilon^2).$$

We do not have matching upper bounds here, but there are precise results on the diameter which yield upper bounds that complement (though do not quite match) the lower bound. For ε as here, w.h.p. the diameter of $H_{n,(1+\varepsilon)/n}$ is $(3 + o(1))\varepsilon^{-1}\log(\varepsilon^3 n)$ (Ding, Kim, Lubetsky and Peres [5] and Riordan and Wormald [19]), and it follows immediately that w.h.p. $\mathrm{spread}(H_{n,p}) = O(\varepsilon^{-2}\log^2(\varepsilon^3 n))$, which is within a $\log^2(\varepsilon^3 n)$ factor of the lower bound in Theorem 1.4.

In Section 5 we study the random regular graph $G(n,d)$ in the large-d case. As noted above, the spread of a connected n-vertex graph is always at least $\mathrm{spread}(\mathsf{K}_n) \geqslant 1/4 - 1/(4n^2)$. Indeed, for graphs with bounded average degree the spread is bounded above $1/4$: see Proposition 5.1. However, the spread of $G(n,d)$ approaches $1/4$ as d becomes large, in the following sense.

Theorem 1.5. *For each $\varepsilon > 0$ there exists a constant d_0 such that for each $d \geqslant d_0$ we have w.h.p.*

$$\mathrm{spread}(G(n,d)) < 1/4 + \varepsilon.$$

In Section 6 we study a basic 'small world' model $R_{n,c/n}$ of a random graph, following Watts and Strogatz [21] and Newman and Watts [16]; see also, for example, Durrett [7]. We start with a cycle on the vertices $1,\ldots,n$ (with each i and $i+1$ adjacent, where $n+1$ means 1). Then the other possible 'short-cut' edges are added independently with probability c/n.

Theorem 1.6. *For each fixed $c > 0$ there exists a constant $C_3 = C_3(c)$ such that w.h.p.*

$$\mathrm{spread}(R_{n,c/n}) \leqslant C_3.$$

Again, the proof gives a stronger result with a tail inequality for Lipschitz functions which is exponential in $\sqrt{n}$. To set Theorem 1.6 in context, we shall show also that for any C there exists a constant $c_0 = c_0(C) > 0$ such that if $0 < c \leqslant c_0$ then w.h.p. $\mathrm{spread}(R_{n,c/n}) > C$.

In Section 7 we introduce edge lengths. Given a connected graph G with edge lengths $\ell(uv) \geqslant 0$, we call a real-valued function f on $V(G)$ *Lipschitz* if we always have $|f(u) - f(v)| \leqslant \ell(uv)$. The *spread* is defined to be the maximum variance of $f(X)$ for such an f, where X is uniform over V.

Theorem 1.7. *There is a constant C_4 such that for K_n with edge lengths i.i.d. uniform over $1, \ldots, n$, the spread is w.h.p. at most C_4.*

As with other theorems, the proof in fact yields a stretched exponential tail inequality for Lipschitz functions. The proof also shows that the same result holds for i.i.d. edge lengths that are exponential with mean n, or such exponentials $+1$.

Theorem 1.7 is best possible in the following sense. Given any fixed $C > 0$, if the edge lengths are uniform over $1, \ldots, \lceil Cn \rceil$ then w.h.p. the number of vertices with minimum incident edge length at least C is at least $2\lceil n/6 \rceil$, and then the spread is at least $C^2/12$. (Set $f(v) = C/2$ for $\lceil n/6 \rceil$ such vertices v, $f(v) = -C/2$ for another $\lceil n/6 \rceil$ such vertices v, and $f(v) = 0$ otherwise.)

Finally Section 8 contains some open problems arising from our work.

Given a graph G, we let $v(G)$ denote the number of vertices and $e(G)$ the number of edges. We use c_1, C_1, *etc.*, to denote various positive constants. (We use c_i for small constants and C_i for large.) In Sections 3 and 6, where we consider $G_{n,c/n}$ and $R_{n,c/n}$, these are allowed to depend on c, but they never depend on n.

2. Random regular graphs

Recall from Section 1 that $G(n,d)$ denotes the random regular graph with degree d. (If d is odd, n is required to be even.) The following result will yield Theorem 1.2 as an immediate corollary.

Theorem 2.1. *Fix $d \geqslant 3$. There exists a constant $c_1 > 0$ for which w.h.p. $G(n,d)$ is such that every Lipschitz function $f : G(n,d) \to \mathbb{R}$ satisfies*

$$|\{v : |f(v) - m| \geqslant x\}| < 2e^{-c_1 x} n \quad \textit{for all } x \geqslant 0,$$

where m is any median of f.

In principle, numerical values could be given for c_1, but we have not tried to find an explicit value, nor to optimize the arguments. However, c_1 can be taken independent of

$d \geqslant 3$; in fact, it follows by monotonicity [10, Theorem 9.36] that any constant that works for $d = 3$ will work for all larger d as well. We will thus consider $d = 3$ only in the proof. (Alternatively, and perhaps more elementarily, we are convinced that the proof below could easily be modified to an arbitrary d, but we have not checked the details.)

For $\alpha > 0$ we say that a graph G is an *α-expander* if every set $W \subset V(G)$ with $|W| \leqslant v(G)/2$ contains at least $\alpha|W|$ vertices with neighbours in $V(G) \setminus W$. (This is slightly at odds with the standard definition of expansion but is more convenient for our purposes.) Observe that an α-expander must be connected. For disjoint sets A and B of vertices in G let $E(A,B)$ be the set of edges with one end in A and one in B, and let $e(A,B) = |E(A,B)|$. The *Cheeger constant* of a graph G with vertex set V and $e(G) > 0$ is

$$\Phi(G) = \min_{\{S \subset V : 0 < \sum_{v \in S} d(v) \leqslant e(G)\}} \frac{e(S, V \setminus S)}{\sum_{v \in S} d(v)}. \tag{2.1}$$

$\Phi(\cdot)$ measures the edge expansion, rather than the vertex expansion, of graphs. We shall use the following expander property of $G(n,3)$, proved (in a more general version) in [2] (see also [14] and [8, (proof of) Lemma 5.1]).

Lemma 2.2 (Lemma 5.3 of [2]). *There is a constant $c_2 > 0$ such that w.h.p. $\Phi(G(n,3)) \geqslant c_2$.* □

Since $G(n,3)$ has constant degree, Lemma 2.2 immediately implies vertex expansion for $G(n,3)$, with the same constant. We state this as a simple lemma.

Lemma 2.3. *If G is regular, and $0 < \alpha \leqslant \Phi(G)$, then G is an α-expander.*

Proof. Let $n := v(G)$, and let d be the degree of the vertices. Note that G has precisely $dn/2$ edges. Fix a set W of vertices in G with $|W| \leqslant n/2$. Then

$$\sum_{v \in W} d(v) = d|W| \leqslant dn/2,$$

so by (2.1) there are at least $\Phi(G)d|W|$ edges from W to its complement. These edges have at least $\Phi(G)d|W|/d \geqslant \alpha|W|$ endpoints in W. □

Lemma 2.4. *$G(n,3)$ is w.h.p. a c_2-expander.*

Proof. An immediate consequence of Lemmas 2.2 and 2.3. □

The following deterministic result on expanders now yields Theorem 2.1.

Theorem 2.5. *For each $0 < \alpha \leqslant 1/2$ and each α-expander G_n on $[n]$, every Lipschitz function f for G_n satisfies*

$$|\{v : |f(v) - m| \geqslant x\}| < 2e^{-(\alpha/2)x} n \quad \textit{for all } x \geqslant 0,$$

where m is a median of f.

https://doi.org/10.1017/S0963548314000248 Published online by Cambridge University Press

Proof. Let f be a Lipschitz function on G_n, with median m. We may assume that $m = 0$; otherwise we replace f by $f - m$. Let $V_t := \{v \in [n] : f(v) \geqslant t\}$. Then $|V_t| \leqslant n/2$ for $t > 0$.

If $t > 0$ and V_t is non-empty then there is a subset of V_t of size at least $\alpha|V_t|$ of vertices x that are adjacent to at least one vertex $y \notin V_t$. Thus $f(y) < t$, and since f is Lipschitz, we have $f(x) < t + 1$ for every such x. Consequently, $|V_{t+1}| \leqslant (1-\alpha)|V_t|$ when $t > 0$. Since $|V_1| \leqslant n/2 \leqslant (1-\alpha)n$, we obtain by induction, for simplicity considering integers k only,

$$|V_k| \leqslant (1-\alpha)^k n \leqslant e^{-\alpha k} n, \quad k = 1, 2, \ldots$$

By symmetry, we have the same estimate for $\{v : f(v) \leqslant -k\}$, and thus, for every $x \geqslant 1$,

$$|\{v : |f(v)| \geqslant x\}| \leqslant 2e^{-\alpha \lfloor x \rfloor} n < 2e^{-(\alpha/2)x} n.$$

Finally, since $2e^{-\alpha/2} > 1$ the last bound also holds for each $0 \leqslant x < 1$, which completes the proof. □

3. $G_{n,c/n}$ with $c > 1$ fixed

In this section we consider supercritical random graphs, and prove the following theorem, which immediately implies Theorem 1.3.

Theorem 3.1. *Given fixed $c > 1$, there is a constant $c_3 = c_3(c)$ such that w.h.p. the giant component $H = H_{n,c/n}$ of $G_{n,c/n}$ is such that every Lipschitz function f for H satisfies*

$$|\{v : |f(v) - m| > x\}| < 2e^{-c_3\sqrt{x}} v(H) \quad \text{for all } x \geqslant 0, \tag{3.1}$$

where m is a median of f.

For this case, in place of Lemma 2.2 we can use another result of [2]. For a graph G and a set of vertices $U \subset V(G)$, we write $G \setminus U$ for the subgraph of G induced by $V(G) \setminus U$. For $0 < \alpha < 1$ we say that a connected graph H is an *α-decorated expander* if H has a subgraph F such that:

(DE1) $\Phi(F) \geqslant \alpha$;

(DE2) listing the connected components of $H \setminus V(F)$ as $D_1, \ldots, D_\nu$ for some ν,

$$|\{i : e(D_i) + e(D_i, F) \geqslant x\}| \leqslant e^{-\alpha x} e(H);$$

(DE3) no vertex $v \in V(F)$ is adjacent to ('decorated by') more than $1/\alpha$ of the components D_i.

Note that (DE1) implies that F is connected. Note further that (DE2) implies:

(DE2′) for all $x \geqslant 0$, $|\{i : v(D_i) \geqslant x\}| \leqslant e^{-\alpha x} e(H)$.

We shall use (DE2′) rather than (DE2) in what follows. Let us say that H is a *weak α-decorated expander* if (DE1), (DE2′) and (DE3) hold, and one further condition holds:

(DE4) $v(F) \geqslant \alpha v(H)$.

From Benjamini, Kozma and Wormald [2] (their Theorem 4.2 and Lemma 4.7, combined) we have the following fact.

Lemma 3.2. *Fix $c > 1$. Then there is a constant $\alpha = \alpha(c) > 0$ such that w.h.p. the giant component $H_{n,c/n}$ of $G_{n,c/n}$ is a weak α-decorated expander.* □

Since the expansion guaranteed by Lemma 3.2 is an edge expansion, we will need to do a little work to derive the vertex expansion required to prove Theorem 3.1. The following lemma will give some further, more elementary, properties of $G_{n,c/n}$ that suffice for our purposes. Given a graph G, let $V_i(G)$ be the set of vertices of degree i, and let $v_i(G) = |V_i(G)|$.

The constants $C_5, C_6, \dots$ below may depend on c and α.

Lemma 3.3. *For fixed $c > 1$, $G_{n,c/n}$ is w.h.p. such that $H = H_{n,c/n}$ satisfies the following properties, for suitable constants:*

(P1) $n' := v(H) > \gamma n/2$ *for some* $\gamma = \gamma(c) > 0$,
(P2) $e(H) \leqslant C_5 v(H)$,
(P3) $v_i(H) \leqslant e^{-i} v(H)$ *for all* $i \geqslant C_6$.

Proof. It is well known that $n'/n \xrightarrow{\text{p}} \gamma(c) > 0$. It is also well known and easy to see that $e(G_{n,c/n})/n \xrightarrow{\text{p}} c/2$. These two results yield (P1) and (P2).

For (P3), let $d_j = d_j(G_{n,c/n})$ be the degree of vertex j, and let X be the random variable $\sum_{j=1}^n e^{2d_j}$. Since each d_j has a binomial $\operatorname{Bin}(n-1, c/n)$ distribution,

$$\mathbb{E}\, X = n\,\mathbb{E}\, e^{2d_1} = n\left(1 + \frac{c}{n}(e^2 - 1)\right)^{n-1} \leqslant n e^{c(e^2-1)}. \tag{3.2}$$

A similar calculation shows that

$$\operatorname{Var} X = n \operatorname{Var}(e^{2d_1}) + n(n-1)\operatorname{Cov}(e^{2d_1}, e^{2d_2}) = O(n).$$

Consequently, by Chebyshev's inequality, w.h.p.

$$\sum_{i=0}^{\infty} e^{2i} v_i(G_{n,c/n}) = X \leqslant e^{ce^2} n.$$

The result follows, using (P1). □

Remark 3.4. The proof of (P3) shows that it could be strengthened to $v_i(H) \leqslant e^{-Ci} v(H)$ for all $i \geqslant C_6$, for any fixed C; conversely, it would for our purposes be enough that $v_i(H) \leqslant 2e^{-\alpha i} v(H)$ for all i. For simplicity, we use the version above.

Let us call a connected graph H a *well-behaved weak α-decorated expander* if it is a weak α-decorated expander and it has properties (P2) and (P3) in the above lemma for some constants C_5, C_6, where, for definiteness, we assume $C_5 = C_6 = \alpha^{-1}$. By Lemma 3.3, Lemma 3.2 can be improved (possibly reducing α) as follows.

Lemma 3.5. *Fix $c > 1$. Then there is a constant $\alpha = \alpha(c) > 0$ such that w.h.p. the giant component $H_{n,c/n}$ of $G_{n,c/n}$ is a well-behaved weak α-decorated expander.* □

Theorem 3.1 now follows immediately from the following deterministic lemma.

Lemma 3.6. *Let the connected graph H be a well-behaved weak α-decorated expander. Then* (3.1) *holds for every Lipschitz function f on H, for some c_3 depending on α.*

Proof. Fix a subgraph F of H which verifies that H is a weak α-decorated expander. Let D be the graph $H \setminus V(F)$, and let $D_1, \dots, D_\nu$ be the components of D. Fix a Lipschitz function f on H. Let $n' = v(H)$ as in Lemma 3.3.

We write $H_{\geqslant t}$ for the set of vertices $v \in V(H)$ with $f(v) \geqslant t$ and define $H_{>t}$, $H_{\leqslant t}$, $H_{<t}$ similarly. Further, we write $F_{\geqslant t}$ (and $F_{>t}$ *etc.*) for $V(F) \cap H_{\geqslant t}$, and $D_{\geqslant t}$ (*etc.*) for $H_{\geqslant t} \cap V(D) = H_{\geqslant t} \setminus V(F)$. We also assume as in the proof of Theorem 2.1 that f has median $m = 0$; hence $|H_{\leqslant 0}|, |H_{\geqslant 0}| \geqslant n'/2$.

Our plan of attack is as follows. First, we find a large subset of $V(F)$ consisting exclusively of vertices v with $f(v)$ bounded above by a constant. Such a set is not quite guaranteed by the fact that $|H_{\leqslant 0}| \geqslant n'/2$, because $H_{\leqslant 0}$ may be largely contained within $V(H) \setminus V(F)$. However, we shall use properties (DE2′) and (DE3) to find such a set. Second, we use the expansion of F to show that the sets $F_{\geqslant t}$ decay rapidly in size as t grows. Finally, we use the fact that the decorations D_i are typically small and do not attach to very many vertices of $F_{\geqslant t}$, to show that the sets $D_{\geqslant t}$ also decay rapidly in size as t grows. We now turn to the details. For simplicity we prove the theorem for x integer, which easily implies the more general statement.

For $\lambda > 0$, let F^λ be the union of F and all components D_i with $v(D_i) < \lambda$. By property (P2), $e(H) \leqslant C_5 n'$. By property (DE2′), for any $\lambda > 0$ we have

$$\begin{aligned} \sum_{\{i:\ v(D_i) \geqslant \lambda\}} v(D_i) &= \sum_{j=0}^{\infty} \sum_{\{i:\ 2^j\lambda \leqslant v(D_i) < 2^{j+1}\lambda\}} v(D_i) \\ &\leqslant \sum_{j=0}^{\infty} 2^{j+1} \lambda e^{-\alpha\lambda 2^j} \cdot C_5 n'. \end{aligned} \tag{3.3}$$

Choose $\lambda = \lambda_1$ large enough that the upper bound in (3.3) is less than $n'/4$; then F^{λ_1} contains at least $3n'/4$ vertices. Since at most $n'/2$ vertices v in H have $f(v) > 0$, it follows that at least $n'/4$ of the vertices in F^{λ_1} have $f(v) \leqslant 0$. Since each component of $F^{\lambda_1} \setminus F$ has less than λ_1 vertices, either $|F_{\leqslant 0}| \geqslant n'/8$ or at least $n'/(8\lambda_1)$ components of $F^{\lambda_1} \setminus F$ contain a vertex of $H_{\leqslant 0}$. Since all vertices in $F^{\lambda_1} \setminus F$ have distance at most λ_1 from F, property (DE3) and the Lipschitz property of f then guarantee that in either case (assuming $\lambda_1 > \alpha$ as we may)

$$|F_{\leqslant \lambda_1}| \geqslant \frac{\alpha n'}{8\lambda_1} =: c_4 n'.$$

Since every vertex of F has at least one neighbour in F, it follows that

$$\sum_{v \in F_{\leqslant \lambda_1}} d_F(v) \geqslant c_4 n'.$$

https://doi.org/10.1017/S0963548314000248 Published online by Cambridge University Press

Assuming that $\sum_{v\in F_{\leqslant\lambda_1}} d_F(v) \leqslant e(F)$, by the expansion property (DE1) we thus have that $e(F_{\leqslant\lambda_1}, F_{>\lambda_1}) \geqslant \alpha c_4 n'$. The Lipschitz property of f implies that each edge in $E(F_{\leqslant\lambda_1}, F_{>\lambda_1})$ has one endpoint in $F_{\leqslant\lambda_1+1} \setminus F_{\leqslant\lambda_1}$, and thus

$$\sum_{v\in F_{\leqslant\lambda_1+1}\setminus F_{\leqslant\lambda_1}} d_F(v) \geqslant e(F_{\leqslant\lambda_1}, F_{>\lambda_1}) \geqslant \alpha c_4 n'.$$

Repeatedly applying property (DE1) in this manner, and using property (P2), we see that w.h.p. $\sum_{v\in F_{\leqslant\lambda_2}} d_F(v) > e(F)$, where we may take $\lambda_2 = \lambda_1 + C_5/(\alpha c_4) + 1$.

We next apply the expansion of F and properties (P2)–(P3) to bound the sizes of sets $F_{>\lambda_2+i}$ for positive integers i. As i becomes large and the sets $F_{>\lambda_2+i}$ become small, the proportion of the sum $\sum_{v\in F_{>\lambda_2+i}} d_F(v)$ due to vertices of large degree may increase; this is the reason we are only able to show that the sizes of the sets $F_{>\lambda_2+i}$ decay exponentially quickly in $\sqrt{i}$.

For given $x > 0$, let a_x be the smallest integer $\geqslant C_6$ such that $\sum_{i>a_x}^{\infty} ie^{-i} \leqslant \alpha x/2$. Since

$$\sum_{i>a}^{\infty} ie^{-i} \leqslant \sum_{i>a}^{\infty} e^{-i/2} \leqslant 3e^{-a/2},$$

there exists C_7 large enough that $a_x \leqslant C_7 \log(1/x)$ for all $x \leqslant 1/2$.

For $\lambda \geqslant \lambda_2$, if $t' = \sum_{v\in F_{>\lambda}} d_F(v)$ then $t' \leqslant e(F)$ by our choice of λ_2. For $0 \leqslant t \leqslant t'$, we thus have $e(F_{>\lambda}, F_{\leqslant\lambda}) \geqslant \alpha t' \geqslant \alpha t$ by (DE1). Let

$$\partial F_{>\lambda} = \{v \in F_{>\lambda} : v \text{ has a neighbour in } F_{\leqslant\lambda}\}.$$

Then for any t as above, $\sum_{v\in\partial F_{>\lambda}} d_F(v) \geqslant e(F_{>\lambda}, F_{\leqslant\lambda}) \geqslant \alpha t$. Also, applying property (P3) and using the definition of $a_{t/n'}$,

$$\sum_{v\in F} d_F(v)\mathbf{1}[d_F(v) > a_{t/n'}] \leqslant \sum_{i>a_{t/n'}}^{\infty} ie^{-i}\cdot n' \leqslant \alpha t/2,$$

and so

$$\sum_{v\in\partial F_{>\lambda}} d_F(v)\mathbf{1}[d_F(v) \leqslant a_{t/n'}] \geqslant \alpha t - \sum_{v\in F} d_F(v)\mathbf{1}[d_F(v) > a_{t/n'}] \geqslant \alpha t/2.$$

Hence, assuming also that $t \leqslant n'/2$,

$$|\partial F_{>\lambda}| \geqslant \frac{\alpha t/2}{a_{t/n'}} \geqslant \frac{\alpha t}{2C_7\log(n'/t)} := c_5\frac{t}{\log(n'/t)}. \tag{3.4}$$

Now fix $\lambda \geqslant \lambda_2$. Taking $t = |F_{>\lambda}| \leqslant \sum_{v\in F_{>\lambda}} d_F(v) = t'$, we also have $t \leqslant |H_{>0}| \leqslant n'/2$, so (3.4) applies with this choice of t. Furthermore, the Lipschitz property of f implies that $\partial F_{>\lambda} \subseteq F_{\leqslant\lambda+1}$, and so

$$|F_{>\lambda+1}| \leqslant |F_{>\lambda}| - |\partial F_{>\lambda}| \leqslant t(1 - c_5/\log(n'/t)).$$

Next, for integers $i \geqslant 1$, let $k_i = \lceil i/c_5 \rceil$. Then for all $t \geqslant n'/2^i$, we have $(1 - c_5/\log(n'/t))^{k_i} < 1/2$. It follows immediately that for all integers $i \geqslant 1$ we have

$$|F_{>\lambda_2+\sum_{j=2}^{i} k_j}| \leqslant \frac{n'}{2^i},$$

https://doi.org/10.1017/S0963548314000248 Published online by Cambridge University Press

so there is a $C_8 > 0$ such that for all real $x \geqslant 1$, and trivially for $0 \leqslant x \leqslant 1$,

$$|F_{>C_8x^2}| \leqslant \frac{n'}{2^x}. \tag{3.5}$$

We now deal with the elements of the 'decorations' graph D, and assume that its components $D_1, \ldots, D_v$ are listed so that $v(D_1) \geqslant \cdots \geqslant v(D_v)$. We first remark that by (DE2′) and (P2), if m_k is the number of components D_i of D with $v(D_i) \geqslant k$, then $m_k \leqslant C_5 n' e^{-\alpha k}$ for all integers $k \geqslant 1$. Hence, for any real t with $0 < t \leqslant n'$, we have, with $x = \log(C_5 n'/t)/\alpha$,

$$\begin{aligned}\sum_{j=1}^{\lfloor t \rfloor} v(D_j) &= \sum_{k=1}^{\infty} \min\big(\lfloor t \rfloor, m_k\big) \leqslant \sum_{k=1}^{\infty} \min\big(t, C_5 n' e^{-\alpha k}\big) \\ &= \sum_{k \leqslant x} t + \sum_{k > x} C_5 n' e^{-\alpha k} \leqslant C_9 t(\log n' + 1 - \log t).\end{aligned} \tag{3.6}$$

Next, for $w \in V(D)$, let D^w be the component of D containing w and fix an arbitrary vertex u^w of F that is decorated with D^w. By (DE3), for any set $S \subseteq V(F)$ with $|S| \leqslant s$, the total number of components that decorate S is at most s/α. It then follows from (3.6) that

$$|\{w \in V(D) : u^w \in S\}| \leqslant \sum_{j=1}^{\lfloor s/\alpha \rfloor} v(D_j) \leqslant C_{10} s(\log n' + 1 - \log s) \tag{3.7}$$

if $s \leqslant \alpha n'$, and by taking $C_{10} \geqslant 1/\alpha$ we see that the inequality in fact holds for all $s \leqslant n'$. For $i \geqslant 0$, if $w \in D_{>i}$ then one of the following two events must occur:

(a) $v(D^w) \geqslant 3i/4$,

(b) $d(w, u^w) < 3i/4$ and then $u^w \in F_{>i/4}$.

By (DE2′) and (P2),

$$|\{w \in D : v(D^w) \geqslant 3i/4\}| \leqslant \sum_{j \geqslant 3i/4} j \cdot C_5 n' e^{-\alpha j} \leqslant C_{11} n'(i+1) e^{-3\alpha i/4}. \tag{3.8}$$

Furthermore, by (3.5),

$$|F_{>i/4}| \leqslant n'/2^{c_6\sqrt{i}}$$

and thus by (3.7) we have

$$|\{w \in D : u^w \in F_{>i/4}\}| \leqslant C_{10} \frac{n'}{2^{c_6\sqrt{i}}} \big(1 + c_6\sqrt{i} \log 2\big),$$

so for all i we have

$$|\{w \in D : u^w \in F_{>i/4}\}| \leqslant C_{12} n' e^{-c_7\sqrt{i}} \tag{3.9}$$

for suitable constants C_{12} and $c_7 > 0$. Thus, by (3.8) and (3.9),

$$|D_{>i}| \leqslant |\{w \in D : v(D^w) \geqslant 3i/4\}| + |\{w \in D : u^w \in F_{>i/4}\}| \leqslant C_{13} n' e^{-c_7\sqrt{i}}. \tag{3.10}$$

Hence, using this together with (3.5) to bound $|F_{>i}|$, we have

$$|H_{>i}| = |F_{>i}| + |D_{>i}| \leqslant C_{14} e^{-c_8\sqrt{i}} n'$$

https://doi.org/10.1017/S0963548314000248 Published online by Cambridge University Press

for fixed C_{14} sufficiently large. Now note that $-f$ is also a Lipschitz function on H with a median 0, and so for all $i \geqslant 0$

$$|\{v : |f(v)| > i\}| \leqslant 2C_{14}e^{-c_8\sqrt{i}}n'.$$

To complete the proof, let $i_0 > 0$ satisfy $2C_{14}e^{-c_8\sqrt{i_0}} \leqslant 1$, and then choose c_9 with $0 < c_9 \leqslant c_8$ satisfying $2e^{-c_9\sqrt{i_0}} > 1$. Now $2e^{-c_9\sqrt{i}} > \min\{1, 2C_{14}e^{-c_8\sqrt{i}}\}$ for each $i \geqslant 0$, and so $|\{v : |f(v)| > i\}| < 2e^{-c_9\sqrt{i}}n'$ for all $i \geqslant 0$, and the theorem follows. □

4. $G_{n,(1+\varepsilon)/n}$ with $\varepsilon \to 0$, $\varepsilon \gg n^{-1/3}$

In this section we consider the barely supercritical case, and prove Theorem 1.4. Fix a function $\varepsilon = \varepsilon(n)$ as above and let $p = (1+\varepsilon)/n$. As above, denote by $H_{n,p}$ the largest component of $G_{n,p}$. Further, write $C_{n,p}$ (resp. $K_{n,p}$) for the core (resp. kernel) of $H_{n,p}$. For such ε, it is known (see [15] and also [10], Chapter 5) that w.h.p.

$$\begin{aligned} v(H_{n,p}) &= (1+o(1))2\varepsilon n, \\ v(C_{n,p}) &= (1+o(1))2\varepsilon^2 n, \quad \text{and} \\ v(K_{n,p}) &= (1+o(1))\frac{4}{3}\varepsilon^3 n. \end{aligned} \tag{4.1}$$

For a connected graph G, we write $\kappa(G) = e(G) - v(G)$, and call κ the *excess* of G. A moment's reflection reveals that $\kappa(H_{n,p}) = \kappa(C_{n,p}) = \kappa(K_{n,p})$, and it is known [9, 10, 13] that for ε as above, w.h.p.

$$\kappa(H_{n,p}) = (1+o(1))\frac{2}{3}\varepsilon^3 n. \tag{4.2}$$

We fix $\delta < 1/10$ and say that $H_{n,p}$ *behaves* if

$$(2-\delta)\varepsilon n \leqslant v(H_{n,p}) \leqslant (2+\delta)\varepsilon n,$$

and if similar inequalities hold for $v(C_{n,p})$, $v(K_{n,p})$, and $\kappa(H_{n,p})$. By the above comments, w.h.p. $H_{n,p}$ behaves. Using this fact and one further lemma, we may prove Theorem 1.4.

The complement $H_{n,p} \setminus V(C_{n,p})$ of the core in the largest component $H_{n,p}$ is a forest consisting of trees that are attached to the core by (exactly) one edge each. We call these trees *pendant*, and denote them (in some order) by $T_1, \ldots, T_N$. We begin with an estimate of the maximum size of the pendant trees.

Lemma 4.1. *There exists a constant C_{15} such that w.h.p.*

$$\max_i v(T_i) \leqslant C_{15}\varepsilon^{-2}\log(n\varepsilon^3). \tag{4.3}$$

Proof. We create another forest by removing all edges in the core $C_{n,p}$ from $H_{n,p}$; the result is a forest where each component consists of a single vertex in $V(C_{n,p})$ together with all pendant trees attached to it (if any). We regard these trees as rooted, with the vertices in $V(C_{n,p})$ as the roots, and denote them by T_w^*, $w \in V(C_{n,p})$.

Conditioned on $V(H_{n,p})$ and $C_{n,p}$, this forest $\{T_w^*\}_w$ is a uniformly distributed forest of rooted trees, with given sets of $M := v(C_{n,p})$ roots and $m := v(H_{n,p}) - M$ non-roots.

https://doi.org/10.1017/S0963548314000248 Published online by Cambridge University Press

The maximum size of a tree in a random forest of rooted trees has been studied by Pavlov [17] (see also [12, Section 3.6] and [18]). In our case, if $H_{n,p}$ behaves and n is large enough, we have $(2-\delta)n\varepsilon^2 \leqslant M \leqslant (2+\delta)n\varepsilon^2$ and $(2-2\delta)n\varepsilon \leqslant m \leqslant (2+\delta)n\varepsilon$. In particular, $m/M \to \infty$ and $m/M^2 \leqslant (n\varepsilon^3)^{-1} \to 0$. This is the range of [17, Theorem 3 (and the remark following it)], which implies that w.h.p., conditioned on M and m,

$$\max_w v(T_w^*) = (1+o(1))\,\frac{2m^2}{M^2}\log\left(\frac{M^2}{m}\right) \leqslant C_{15}\varepsilon^{-2}\log(n\varepsilon^3).$$

The same estimate thus holds unconditionally w.h.p., and the result follows since every pendant tree is contained in some T_w^*. □

We say that $e \in E(K_{n,p})$ has *length* $\ell(e)$ if the path in $C_{n,p}$ corresponding to e contains $\ell(e)$ edges (so $\ell(e)-1$ internal vertices). To prove Theorem 1.4, we construct a function f on $V(H_{n,p})$ which has large spread; the construction is roughly as follows. Vertices v of the kernel have value $f(v)=0$. For $e \in E(K_{n,p})$, if $\ell(e)$ is large then we may choose f giving many distinct values to the vertices of the path in $C_{n,p}$ corresponding to e, while maintaining that f is Lipschitz. (If $\ell(e)$ is too small then we set $f(v)=0$ for all edges of the path corresponding to e.) We extend the domain from $V(C_{n,p})$ to $V(H_{n,p})$ by assigning to each vertex in a pendant tree the value of the unique core vertex to which the tree is attached. We now turn to details.

Proof of Theorem 1.4. Since $H_{n,p}$ behaves w.h.p., it suffices to prove that *given* that $H_{n,p}$ behaves, w.h.p. $\mathrm{spread}(H_{n,p}) = \Omega(1/\varepsilon^2)$. We shall define a Lipschitz function f on the vertices of $H_{n,p}$ for which, given that $H_{n,p}$ behaves, w.h.p. $\mathrm{Var}(f) \geqslant \gamma/\varepsilon^2$ for some fixed $\gamma > 0$. We define f in a few steps, starting from the core. We say that $e \in E(K_{n,p})$ has *length* $\ell(e)$ if the path in $C_{n,p}$ corresponding to e contains $\ell(e)$ edges (so $\ell(e)-1$ internal vertices). Since $H_{n,p}$ behaves,

$$\begin{aligned} e(K_{n,p}) &= v(K_{n,p}) + \kappa(K_{n,p}) \\ &\leqslant \left(\frac{4}{3}+\delta\right)\varepsilon^3 n + \left(\frac{2}{3}+\delta\right)\varepsilon^3 n \\ &= (2+2\delta)\varepsilon^3 n, \end{aligned} \tag{4.4}$$

and

$$\begin{aligned} |V(C_{n,p}) \setminus V(K_{n,p})| &\geqslant (2-\delta)\varepsilon^2 n - \left(\frac{4}{3}+\delta\right)\varepsilon^3 n \\ &\geqslant (2-2\delta)\varepsilon^2 n, \end{aligned} \tag{4.5}$$

for n sufficiently large.

We say that an edge $e \in E(K_{n,p})$ is *short* if

$$\ell(e) \leqslant \left\lfloor \frac{1-\delta}{2\varepsilon(1+\delta)} \right\rfloor$$

(and long otherwise), and that $v \in V(C_{n,p}) \setminus V(K_{n,p})$ is *useless* if it is contained in a path corresponding to a short edge (and useful otherwise). By (4.4) and (4.5), the number of useful vertices is at least

$$|V(C_{n,p}) \setminus V(K_{n,p})| - e(K_{n,p}) \cdot \frac{1-\delta}{2\varepsilon(1+\delta)} \geqslant (1-\delta)\varepsilon^2 n. \tag{4.6}$$

Next, let $r = r(n)$ be the largest integer divisible by 3 and with $2r \leqslant (1-\delta)/(2\varepsilon(1+\delta))$. For each long edge $e \in E(K_{n,p})$, let P_e be the path in $C_{n,p}$ corresponding to e (so the endpoints of P_e are in $K_{n,p}$), and let P'_e be a sub-path of P_e, not containing the endpoints of P_e, which is as long as possible subject to the condition that $2r$ divides $v(P'_e)$ (picked according to some rule); such a sub-path certainly exists since

$$v(P_e) = e(P_e) + 1 \geqslant \left\lfloor \frac{1-\delta}{2\varepsilon(1+\delta)} \right\rfloor + 2 \geqslant 2r + 2,$$

so P_e has at least $2r$ internal vertices. Since P_e has $v(P_e) - 2$ internal vertices, we also have that $v(P'_e) \geqslant (v(P_e) - 2)/2$, so by (4.4) and (4.6),

$$\begin{aligned} |\{v : v \in P'_e \text{ for some } e \in E(K_{n,p})\}| &\geqslant \sum_{\{e:e \text{ is long}\}} \frac{v(P_e) - 2}{2} \\ &\geqslant \frac{(1-\delta)\varepsilon^2 n}{2} - 2(1+\delta)\varepsilon^3 n \\ &\geqslant \frac{(1-\delta)\varepsilon^2 n}{3}, \end{aligned} \tag{4.7}$$

for n large enough. We now define the restriction of f to $V(C_{n,p})$ as follows.

- If $v \in V(K_{n,p})$, v is useless, or v is not in P'_e for any long edge e, then set $f(v) = 0$.
- For each long edge e, repeat the sequence of values $12\ldots(r-1)rr(r-1)\ldots1$ along P'_e (so if v is the ith or $(2r+1-i)$th vertex mod $2r$ along some path P'_e then $f(v) = i$).

To extend f from $C_{n,p}$ to the remainder of $H_{n,p}$, for each vertex $v \in V(H_{n,p})$, we define the *point of attachment* $a(v)$ to be the vertex $x \in C_{n,p}$ whose distance from v in $H_{n,p}$ is minimum, and we set $f(v) = f(a(v))$. In other words, for each pendant tree T in $H_{n,p}$ that hooks up to the core at $v \in V(C_{n,p})$, we set $f(w) = f(v)$ for all $w \in V(T)$.

To analyse the variance of f, for $i = 1, 2, 3$, let

$$B_i = \left\{ v \in V(C_{n,p}) : \frac{i-1}{3} r < f(v) \leqslant \frac{i}{3} r \right\},$$

and let B_0 be all remaining vertices of $C_{n,p}$, *i.e.*, those with $f(v) = 0$. By the definition of f and since 3 divides r, the sizes of B_1, B_2, and B_3 are identical, and are each at least $(1-\delta)\varepsilon^2 n/9$. Also, for $i = 1, 2, 3$, let B_i^+ be the set of vertices $v \in V(H_{n,p})$ with $a(v) \in B_i$. We will prove the following assertion:

($\star$) given that $H_{n,p}$ behaves, w.h.p. $|B_i^+| \geqslant \varepsilon n/44$ for each $i = 1, 2, 3$.

Assuming for the moment that ($\star$) holds, we can quickly complete the proof of the theorem. For each graph $H_{n,p}$ which behaves, the corresponding (fixed) function f satisfies

$$\begin{aligned} \mathrm{Var}(f) &= \frac{1}{2(n')^2} \sum_{x,y \in V(H_{n,p})} (f(x) - f(y))^2 \\ &\geqslant (n')^{-2} \sum_{x \in B_1^+} \sum_{y \in B_3^+} (f(x) - f(y))^2 \end{aligned}$$

https://doi.org/10.1017/S0963548314000248 Published online by Cambridge University Press

$$\begin{aligned}
&\geqslant (n')^{-2}|B_1^+||B_3^+|\, r^2/9 \\
&\geqslant \frac{(\varepsilon n/44)^2}{((2+\delta)\varepsilon n)^2}\frac{r^2}{9} \\
&= \frac{r^2}{69696(1+\delta/2)^2}.
\end{aligned}$$

But $r = \Omega(1/\varepsilon)$, and so it follows that, conditional on the event that $H_{n,p}$ behaves, w.h.p. $\mathrm{Var}(f) = \Omega(\varepsilon^{-2})$, as needed.

It thus remains to prove ($\star$), and we now turn to this. Let $X = |B_1^+|$, the number of vertices $v \in V(H_{n,p})$ with $a(v) \in B_1$. Our aim is to show that $\mathbb{P}\{X \geqslant \varepsilon n/44\} = 1 - o(1)$.

We note that given $C_{n,p}$, we can specify $H_{n,p}$ by listing the pendant subtrees of $H_{n,p}$, and their points of attachment in $C_{n,p}$, as $T_1, \ldots, T_N$ and $U_1, \ldots, U_N$. By routine calculation it is easily seen that given $C_{n,p}$ and the pendant subtrees $T_1, \ldots, T_N$, the points of attachment $U_1, \ldots, U_N$ are independent and uniformly random elements of $V(C_{n,p})$. We further note that given $C_{n,p}$ and the pendant subtrees $T_1, \ldots, T_N$, we can determine whether or not $H_{n,p}$ behaves. Then, recalling Lemma 4.1,

$$\mathbb{P}\{X \geqslant \varepsilon n/44\} \geqslant \inf_{\mathcal{S}} \mathbb{P}\{X \geqslant \varepsilon n/44 \mid C_{n,p}, T_1, \ldots, T_N\} - o(1), \tag{4.8}$$

where $\mathcal{S}$ represents all possible choices of $C_{n,p}$ and N and $T_1, \ldots, T_N$ for which $H_{n,p}$ behaves and (4.3) holds. Fix any such choice and let $t_i = v(T_i)$ for $i = 1, \ldots, N$. To shorten forthcoming formulae, let

$$\mathbb{P}_{\mathrm{c}}\{\cdot\} = \mathbb{P}\{\cdot \mid C_{n,p}, T_1, \ldots, T_N\},$$

and define $\mathbb{E}_{\mathrm{c}}$ and $\mathrm{Var}_{\mathrm{c}}$ similarly. Given $C_{n,p}$ and $T_1, \ldots, T_N$, we may write X as

$$X = |B_1| + \sum_{i=1}^{N} t_i \mathbf{1}[U_i \in B_1].$$

Since $H_{n,p}$ behaves, by the estimates above,

$$\frac{|B_1|}{v(C_{n,p})} \geqslant \frac{(1-\delta)\varepsilon^2 n/9}{(2+\delta)\varepsilon^2 n} \geqslant \frac{1-\delta}{18(1+\delta)} \geqslant \frac{1}{22}.$$

Since the points of attachment $U_1, \ldots, U_N$ of $T_1, \ldots, T_N$ in $C_{n,p}$ are uniform and

$$\sum_{i=1}^{N} t_i = |V(H_{n,p}) \setminus V(C_{n,p})|,$$

it thus follows that

$$\mathbb{E}_{\mathrm{c}}(X) = |B_1| + \frac{|B_1|}{v(C_{n,p})} \cdot |V(H_{n,p}) \setminus V(C_{n,p})| > \frac{\varepsilon n}{22}, \tag{4.9}$$

the preceding inequality holding for n sufficiently large since $H_{n,p}$ behaves. Next, given $C_{n,p}$ and $T_1, \ldots, T_N$, $|B_1|$ is determined and $X - |B_1|$ is a sum of independent random variables $t_i \mathbf{1}[U_i \in B_1]$, $i = 1, \ldots, N$. Hence,

$$\mathrm{Var}_{\mathrm{c}}(X) = \sum_{i=1}^{N} t_i^2\, \mathrm{Var}_{\mathrm{c}}(\mathbf{1}[U_i \in B_1]) \leqslant \sum_{i=1}^{N} t_i^2.$$

https://doi.org/10.1017/S0963548314000248 Published online by Cambridge University Press

By Chebyshev's inequality, when n is large enough that (4.9) holds, we thus have

$$\mathbb{P}_{\mathrm{c}}\left\{X < \frac{\varepsilon n}{44}\right\} \leqslant \frac{\sum_{i=1}^{N} t_i^2}{(\varepsilon n/44)^2}.$$

Since we have assumed that (4.3) holds, and that $H_{n,p}$ behaves,

$$\sum_{i=1}^{N} t_i^2 \leqslant \max_{1\leqslant i\leqslant N} t_i \cdot \sum_{i=1}^{N} t_i \leqslant C_{16}\varepsilon^{-2}\log(n\varepsilon^3)\cdot n\varepsilon,$$

and thus, for n large enough,

$$\mathbb{P}_{\mathrm{c}}\left\{X < \frac{\varepsilon n}{44}\right\} \leqslant C_{17}\frac{n\varepsilon^{-1}\log(n\varepsilon^3)}{(\varepsilon n)^2} = C_{17}\frac{\log(n\varepsilon^3)}{n\varepsilon^3} \to 0$$

as $n \to \infty$. An identical argument yields the same lower bound with X equal to $|B_2^+|$ or $|B_3^+|$. (We do not actually care about $|B_2^+|$.) This establishes ($\star$) and completes the proof. □

5. Regular graphs with large degrees

We saw that w.h.p. the random regular graph $G(n,d)$ has bounded spread for any fixed $d \geqslant 3$, and similarly the random graph $H_{n,c/n}$ has bounded spread for any fixed $c > 1$. As noted in the Introduction, the minimum possible values of the spread (achieved for the complete graph K_n) are $1/4$ if n is even and $1/4 - 1/(4n^2)$ if n is odd. This suggests another natural question for random graphs. How large must degrees be for the spread to be close to $1/4$? We shall see that for random regular graphs, what is needed is simply for the degree d to be big enough.

First we note the deterministic result that the average degree must be large in order for the spread to be close to $1/4$, and then we give a matching result that for random regular graphs graphs high degree is sufficient.

Proposition 5.1. *For any fixed $d \geqslant 2$ there exists $\delta > 0$ such that if the connected graph G has average degree at most d and $v(G) \geqslant 3d$ then* spread$(G) \geqslant 1/4 + \delta$. *(We can take $\delta = 1/(6d)$.)* □

Proof. We shall show that if $V(G) = [n]$ then

$$\operatorname{spread}(G) \geqslant \frac{1}{4} + \left(\frac{1}{d} - \frac{2}{n}\right)\left(1 - \frac{1}{d}\right). \tag{5.1}$$

Note that this gives spread$(G) \geqslant 1/4 + 1/6d$ if $d \geqslant 2$ and $n \geqslant 3d$.

Let $t = \lfloor n/2d \rfloor$, let T consist of t vertices of least degree, and let U be the set of vertices adjacent to a vertex in T. Note that $|U| \leqslant n/2$. Let $A \subseteq [n] \setminus T$ be such that $A \supseteq U \setminus T$ and $|A| = a := \lfloor n/2 \rfloor$. Let $B = [n] \setminus (T \cup A)$.

Let $f(v) = 0$ on B, 1 on A and 2 on T. For X uniformly distributed over the vertices, and writing f for $f(X)$, we have $\mathbb{E}f = (1/n)(a+2t)$ and $\mathbb{E}f^2 = (1/n)(a+4t)$, and hence

$$\begin{aligned}
\mathrm{Var}(f) &= (1/n)(a+4t) - (1/n^2)(a^2+4at+4t^2)\\
&= \frac{a}{n}(1-\frac{a}{n}) + \frac{4t}{n} - \frac{2t}{n} + \frac{2t}{n^2}\mathbf{1}[n \text{ odd}] - \frac{4t^2}{n^2}\\
&= 1/4 - \frac{1}{4n^2}\mathbf{1}[n \text{ odd}] + \frac{2t}{n} + \frac{2t}{n^2}\mathbf{1}[n \text{ odd}] - \frac{4t^2}{n^2}\\
&\geqslant 1/4 + \frac{2t}{n}\left(1-\frac{2t}{n}\right)\\
&\geqslant 1/4 + \left(\frac{1}{d}-\frac{2}{n}\right)\left(1-\frac{1}{d}\right).
\end{aligned}$$

□

To prove Theorem 1.5 we need an expansion result for random regular graphs with high degree. Given $\beta > 1$ and $0 < \eta < 1$, let us say that a graph $G = (V, E)$ has (β, η)-*expansion* if, for each $T \subset V$ with $|T| \leqslant (1-\eta)|V|/\beta$, we have $|T \cup N(T)| \geqslant \beta|T|$.

Lemma 5.2. *For each $\beta > 1$ and $0 < \eta < 1/2$, there exists d_0 such that for all $d \geqslant d_0$ w.h.p. $G(n, d)$ has (β, η)-expansion.*

Proof. We consider the configuration model (see [10], Section 9.1) for $G(n, d)$. Let $\alpha > 0$ (α large). For a positive integer t let $f_{n,d}(t)$ be the expected number of pairs T and U of sets of disjoint cells where $|T| = t$ and $|U| = u := \lfloor \alpha t \rfloor$, and each neighbour of a stub in a cell in T is in $T \cup U$. Let $t_0 = \lfloor (1-\eta)|V|/\beta \rfloor$. We aim to bound this quantity from above by $f_{n,d}(t)$, in order to show that $\sum_{t=1}^{t_0} f_{n,d}(t) = o(1)$. The lemma will then follow, with $\beta = 1 + \alpha$.

Note first that, since

$$\frac{d(t+u)-j}{dn-j} \leqslant \frac{t+u}{n}$$

for each $0 < j < dn$, the probability that each neighbour of a stub in a cell in T is in $T \cup U$ is at most

$$\left(\frac{t+u}{n}\right)^{dt/2}.$$

(If we choose the neighbours of the dt stubs in cells in T first, we have to make at least $dt/2$ such choices.) Hence

$$\begin{aligned}
f_{n,d}(t) &\leqslant \binom{n}{t}\binom{n}{u}\left(\frac{t+u}{n}\right)^{dt/2}\\
&\leqslant \left(\frac{ne}{t}\right)^{t}\left(\frac{ne}{u}\right)^{u}\left(\frac{t+u}{n}\right)^{dt/2}\\
&\leqslant \left(\frac{ne}{t}\right)^{t}\left(\frac{ne}{\alpha t}\right)^{\alpha t}\left(\frac{(1+\alpha)t}{n}\right)^{dt/2}
\end{aligned}$$

https://doi.org/10.1017/S0963548314000248 Published online by Cambridge University Press

$$= \left(e^{1+\alpha}\alpha^{-\alpha}(1+\alpha)^{d/2}t^{d/2-1-\alpha}n^{1+\alpha-d/2}\right)^t$$
$$= \left(e^{1+\alpha}\alpha^{-\alpha}(1+\alpha)^{1+\alpha}\left(\frac{(1+\alpha)t}{n}\right)^{d/2-1-\alpha}\right)^t.$$

Now $\alpha^{-\alpha}(1+\alpha)^{1+\alpha} = (1+\alpha)(1+1/\alpha)^\alpha \leqslant (1+\alpha)e$. So

$$f_{n,d}(t) \leqslant \left((1+\alpha)e^{2+\alpha}\left(\frac{(1+\alpha)t}{n}\right)^{d/2-1-\alpha}\right)^t.$$

Let $\alpha > 0$ be sufficiently large that $\log(1+\alpha)+2+\alpha \leqslant 2\alpha$. Let $d_0 \geqslant 6(1+\alpha)$, so that $d/2-1-\alpha \geqslant d/3$ when $d \geqslant d_0$. For such d,

$$f_{n,d}(t) \leqslant \left(e^{2\alpha}\left(\frac{(1+\alpha)t}{n}\right)^{d/3}\right)^t.$$

If $1 \leqslant t \leqslant \log^2 n$, say, then

$$f_{n,d}(t) \leqslant \left(e^{2\alpha}\left(\frac{(1+\alpha)\log^2 n}{n}\right)^{d/3}\right)^t = O(1/n),$$

since $d \geqslant 6$. Also, since

$$\frac{(1+\alpha)t}{n} \leqslant 1-\eta \leqslant e^{-\eta},$$

for $1 \leqslant t \leqslant t_0$ we have

$$f_{n,d}(t) \leqslant (e^{2\alpha}e^{-\eta d/3})^t.$$

From these bounds it is easy to complete the proof, with $\beta = 1+\alpha$. □

Lemma 5.3. *Let $\beta \geqslant 3$, $\eta = \beta^{-1}$ and $n \geqslant 6\beta + \beta^2/2$, and let $G=(V,E)$ have (β,η)-expansion. Let f be an integer-valued function on V with median 0. Let $V_{\geqslant i}$ denote $\{v \in V : f(v) \geqslant i\}$ and so on. Assume that $|V_{\geqslant 1}| \geqslant |V_{\leqslant -1}|$. Then*

$$|V_{\geqslant i}| \leqslant \beta^{-(i-1)}n/2 \quad \text{and} \quad |V_{\leqslant -i}| \leqslant 2\beta^{-i}n \quad \text{for each } i \geqslant 1. \tag{5.2}$$

Proof. Note that $|V_{\leqslant 0}| \geqslant n/2$ and $|V_{\geqslant 0}| \geqslant n/2$. Observe also that $N(V_{\geqslant i}) \subseteq V_{\geqslant i-1}$. If $|V_{\geqslant 2}| > (1-\eta)n/\beta$ then, choosing a set $T \subset V_{\geqslant 2}$ with $|T| = \lfloor (1-\eta)n/\beta \rfloor$,

$$|V_{\geqslant 1}| \geqslant |T \cup N(T)| \geqslant \beta|T| > (1-\eta)n - \beta \geqslant n/2 \geqslant |V_{\geqslant 1}|, \tag{5.3}$$

a contradiction: thus $|V_{\geqslant 2}| \leqslant (1-\eta)n/\beta$. Hence

$$|V_{\geqslant 2}| \leqslant \frac{1}{\beta}|V_{\geqslant 1}| \leqslant \frac{n}{2\beta},$$

and further, for all $i \geqslant 1$ we have $|V_{\geqslant i}| \leqslant \beta^{-(i-1)}|V_{\geqslant 1}|$. Similarly, for all $i \geqslant 1$ we have $|V_{\leqslant -i}| \leqslant \beta^{-(i-1)}|V_{\leqslant -1}|$. Hence it suffices to show (5.2) for $i=1$, *i.e.*, that $|V_{\geqslant 1}| \leqslant n/2$, which is trivial, and $|V_{\leqslant -1}| \leqslant 2n/\beta$.

https://doi.org/10.1017/S0963548314000248 Published online by Cambridge University Press

Recall that $|V_{\geqslant 1}| \geqslant |V_{\leqslant -1}|$. We consider two cases, depending on the size of $V_{\geqslant 1}$. If $|V_{\geqslant 1}| \leqslant (1-\eta)n/\beta$ then $|V_{\geqslant -1}| \leqslant |V_{\geqslant 1}| < 2n/\beta$. If $|V_{\geqslant 1}| > (1-\eta)n/\beta$ then $|V_{\geqslant 0}| \geqslant (1-\eta)n - \beta$ as in (5.3), so $|V_{\leqslant -1}| \leqslant \eta n + \beta \leqslant 2n/\beta$. □

The last lemma easily yields that sufficiently strong expansion yields spread close to 1/4.

Lemma 5.4. *For any $\varepsilon > 0$ there exists $\beta > 1$ such that each graph G with (β, β^{-1})-expansion and n large enough satisfies* $\mathrm{spread}(G) < 1/4 + \varepsilon$.

Proof. Let f be an integer-valued Lipschitz function on G. We may assume that the median of f is 0, and (by symmetry) that $|V_{\geqslant 1}| \geqslant |V_{\leqslant -1}|$. Then, if $\beta \geqslant 3$ and n is large, Lemma 5.3 yields

$$\mathrm{Var}(f) \leqslant \mathbb{E}\left|f - \frac{1}{2}\right|^2 \leqslant \frac{1}{4} + \sum_{i \neq 0,1} \frac{|V_i|}{n}(i - 1/2)^2 \leqslant \frac{1}{4} + O(\beta^{-1}). \qquad \square$$

Lemmas 5.2 and 5.4 complete the proof of Theorem 1.5.

6. Small worlds

In this section we consider the small world model $R_{n,c/n}$ for $c > 0$, and prove Theorem 1.6. We need some preliminary work so that we can appeal to Lemma 3.6. The first step is to show that we may assume that $c \geqslant 2$, by contracting sections of the ring. Now, if we delete the edges of the ring randomly, keeping each with probability c/n, we obtain a random graph $G_{n,c/n}$, whose giant component H is a well-behaved weak α-decorated expander by Lemma 3.5. We show that using the ring to join the other vertices to H yields further decorations, but w.h.p. we still have a well-behaved weak α'-decorated expander.

Step 1: Reduction to the case $c = 2$. We start with a deterministic lemma, which will show that the spread does not shrink too much when we contract sections of the ring.

Lemma 6.1. *Let G be a connected graph on V where $|V| = n$, let k be an integer with $1 \leqslant k < n$, and let $\bar{n} = \lfloor n/k \rfloor$. Let $V_1, \dots, V_{\bar{n}}$ be a partition of V such that each induced subgraph $G[V_i]$ is a connected graph with k or $k+1$ vertices. Form the graph $\bar{G}$ on $[\bar{n}]$ by contracting each V_i to a single new vertex i. Then*

$$\mathrm{spread}(G) \leqslant \frac{(k+1)^3}{k}\,\mathrm{spread}(\bar{G}) + \frac{k^2}{4}.$$

Proof. Let f be a Lipschitz function for G, with mean μ. Let μ_i be the mean of $f|_{V_i}$, that is,

$$\mu_i = (1/|V_i|) \sum_{w \in V_i} f(w).$$

https://doi.org/10.1017/S0963548314000248 Published online by Cambridge University Press

Then, by a standard decomposition of variance,

$$\begin{aligned}\mathrm{Var}(f) &= \frac{1}{n}\sum_i \sum_{w\in V_i}(f(w)-\mu_i+\mu_i-\mu)^2 \\ &= \frac{1}{n}\sum_i\Bigg(\sum_{w\in V_i}((f(w)-\mu_i)^2+(\mu_i-\mu)^2)\Bigg) \\ &= \sum_i \frac{|V_i|}{n}\,\mathrm{Var}(f|_{V_i}) + \sum_i \frac{|V_i|}{n}(\mu_i-\mu)^2.\end{aligned} \tag{6.1}$$

We consider the two terms here separately. Since the induced subgraph $G[V_i]$ has diameter at most k,

$$\mathrm{Var}(f|_{V_i}) \leqslant \mathrm{spread}(G[V_i]) \leqslant k^2/4.$$

Thus

$$\sum_i \frac{|V_i|}{n}\,\mathrm{Var}(f|_{V_i}) \leqslant \frac{k^2}{4}.$$

Now consider the second term above. Observe that for each i and each $w \in V_i$, $|f(w) - \mu_i| \leqslant k/2$. Thus if i and j are adjacent in $\bar{G}$ then $|\mu_i - \mu_j| \leqslant k+1$. Let $\bar{f}(i) = \mu_i$ for each $i \in [\bar{n}]$. Then $(1/(k+1))\bar{f}$ is Lipschitz for $\bar{G}$, and so

$$\mathrm{Var}(\bar{f}) \leqslant (k+1)^2\,\mathrm{spread}(\bar{G}).$$

Next, let $\bar{\mu} = (1/\bar{n})\sum_i \mu_i$; let $h(i) = \mu_i - \bar{\mu}$ for each $i \in [\bar{n}]$; and let the random variable X take values in $[\bar{n}]$ with $\mathbb{P}(X = i) = |V_i|/n$. Observe that $\mathbb{E}[\bar{f}(X)] = \mu$. Then

$$\begin{aligned}\sum_i \frac{|V_i|}{n}(\mu_i-\mu)^2 = \mathrm{Var}\,\bar{f}(X) \;&=\; \mathrm{Var}\,h(X) \;\leqslant\; \mathbb{E}[h(X)^2] \\ &\leqslant \sum_i \frac{k+1}{n}(\mu_i-\bar{\mu})^2 \;\leqslant\; \frac{k+1}{k}\,\mathrm{Var}(h) \\ &= \frac{k+1}{k}\,\mathrm{Var}(\bar{f}) \;\leqslant\; \frac{(k+1)^3}{k}\,\mathrm{spread}(\bar{G}).\end{aligned}$$

Now (6.1) and the above bounds let us complete the proof. □

With the above lemma in hand, we may quickly complete the reduction to the case $c = 2$. Consider $0 < c < 2$. Fix a positive integer k with $k^2c > 2(k+1)$. Observe that given two disjoint k-subsets of $[n]$, the probability that there is an edge in $G_{n,c/n}$ between the sets is $1-(1-c/n)^{k^2} = k^2c/n + O(1/n^2)$. Assume that w.h.p. $\mathrm{spread}(R_{n,2/n}) \leqslant b$ for some constant b.

Consider a large n, partition the vertex set of C_n into paths of k or $k+1$ vertices (which we can always do once $n \geqslant k(k-1)$), and from $R_{n,c/n}$ form the corresponding contracted graph as in Lemma 6.1. Call the contracted graph $R_{\bar{n}}$. Then $R_{\bar{n}}$ contains a deterministic Hamilton cycle arising from the cycle C_n, and edges not in the cycle appear independently, each with probability at least $2(k+1)/n \geqslant 2/\bar{n}$. Thus w.h.p. $\mathrm{spread}(R_{\bar{n}}) \leqslant b$

https://doi.org/10.1017/S0963548314000248 Published online by Cambridge University Press

by the assumption above. Hence, by Lemma 6.1, w.h.p.

$$\mathrm{spread}(R_{n,c/n}) \leqslant \frac{(k+1)^3}{k} b + \frac{k^2}{4}.$$

This completes the reduction to the case $c = 2$.

Step 2: Joining the other vertices to the giant component H of $G_{n,2/n}$. Let us think of $R_{n,c/n}$ as generated by starting with $G_{n,c/n}$ on vertex set $[n]$, picking an independent uniform random Hamilton cycle C in the complete graph on $[n]$, and adding the edges of C if they are not already present. We shall see that adding some edges of C to the edges of H w.h.p. yields a well-behaved weak α-decorated expander G^+ on $[n]$ (for a suitable fixed value of $\alpha > 0$).

Condition on H being a fixed well-behaved α_0-decorated expander (for some fixed $\alpha_0 > 0$), fix a corresponding subgraph F, and let $D_1, \dots, D_\nu$ be the decorations. As usual, let $n' := v(H)$. We further assume (using Lemma 3.3 and Remark 3.4) that H satisfies (P1) and that (P3) holds in the stronger version $v_i(H) \leqslant e^{-2i} v(H)$ for all $i \geqslant C_6$.

Discard all edges outside H other than those from the random cycle C, which we take as oriented. The vertices in $V(H)$ divide the remaining vertices into n' paths: for each vertex w in $V(H)$ let Q_w be the maximal path of vertices outside H ending at w with $X_w \geqslant 0$ vertices (not counting w). We attach the path Q_w at w for each w in H, forming the graph G^+. If w is in $V(F)$ then we have one new decoration attached at w (if $X_w > 0$). If w is in decoration D_i then we add X_w vertices and edges to D_i (and no extra edge to $E(D_i, F)$).

The properties (DE1), (DE3), (DE4), (P2), (P3) are easily seen to hold for G^+ (for a suitable value of $\alpha > 0$). We must check that also (DE2′) holds.

What is the distribution of $(X_w : w \in V(H))$? We may assume without loss of generality that vertex n is in $V(H)$. Think of the vertices in $[n]$ as white. Re-colour vertex n black. Choose a uniformly random subset S of $[n-1]$ of size $n'-1$, and re-colour these elements black. Let $\tilde{X}_1$ be the number of white elements before the first black one, and for $i = 2, \dots, n'$ let $\tilde{X}_i$ be the number of white elements between the $(i-1)$th black vertex and the ith. The distribution of $(X_w : w \in V(H))$ is the same as that of $(\tilde{X}_1, \dots, \tilde{X}_{n'})$. Thus for each list $k_1, \dots, k_{n'}$ of non-negative integers with $\sum_i k_i = n - n'$ we have

$$\mathbb{P}(X_i = k_i \text{ for each } i) = \binom{n-1}{n'-1}^{-1}.$$

It follows (see for example [6]) that the family $(X_w : w \in V(H))$ is negatively associated. Also $X_w \leqslant_s \tilde{X}$ for each w, where $\tilde{X}$ is geometric with parameter $p' = (n'-1)/(n-1)$ (and mean $1/p' - 1$), and $\leqslant_s$ means stochastic ordering (*i.e.*, there exists a coupling such that $X_w \leqslant X$). But $p' \geqslant p := \gamma/3$ for n sufficiently large by (P1). Then $X_w \leqslant_s X$, where X is geometric with parameter p (note that this value p is fixed).

Let $\mathcal{A} = (A_i : i \in I)$ be the partition of $V(H)$ into the vertex sets $V(D)$ of the decorations D of H together with the singletons $\{w\}$ for each $w \in V(F)$. Thus $|I| = \nu + v(F)$. For each i, let $D_i^+ := \bigcup \{Q_w : w \in A_i\}$ be the (possibly empty) union of the paths Q_w attached to A_i.

Let $Y_i := v(D_i^+) = \sum_{w \in A_i} X_w$ for each $i \in I$. The family $(Y_i : i \in I)$ is negatively associated, since it is formed by taking sums of disjoint members of $(X_w : w \in V(H))$: see [11].

Hence, letting $M_X(t) := \mathbb{E}\, e^{tX}$ denote the moment generating function, $M_{Y_i}(t) \leqslant M_X(t)^{|A_i|}$. Also, the family $(\mathbf{1}[Y_i \geqslant j] : i \in I)$ is negatively associated for each j. Let

$$I_j := \{i \in I : |A_i| \leqslant (p/2)j\},$$

and

$$Z_j := \sum_{i \in I} \mathbf{1}[Y_i \geqslant j] \quad \text{and} \quad Z'_j := \sum_{i \in I_j} \mathbf{1}[Y_i \geqslant j].$$

Note that for each $j \geqslant 2/p$,

$$Z_j - Z'_j \leqslant |\{i \in I : |A_i| > (p/2)j\}| \leqslant e(H)\, e^{-\alpha_0 (p/2) j}, \tag{6.2}$$

by (DE2′) for H.

Observe that if $f(t) = e^{-t} M_X(t)^{p/2}$, then

$$\frac{d}{dt}(\log f(t))\,|_{t=0} = -1 + \frac{p}{2}\frac{M'_X(0)}{M_X(0)} = -1 + \frac{p}{2}\mathbb{E}\,X = -\frac{1+p}{2} < 0,$$

and so there exists $t_0 > 0$ such that $f(t_0) < 1$. Let $\alpha = -\log f(t_0)$ so $\alpha > 0$ and $f(t_0) = e^{-\alpha}$. For each $i \in I_j$, by Markov's inequality

$$\begin{aligned} \mathbb{E}[\mathbf{1}[Y_i \geqslant j]] = \mathbb{P}(Y_i \geqslant j) &\leqslant e^{-t_0 j} M_{Y_i}(t_0) \\ &\leqslant e^{-t_0 j} M_X(t_0)^{|A_i|} \leqslant (e^{-t_0} M_X(t_0)^{p/2})^j = e^{-\alpha j}. \end{aligned} \tag{6.3}$$

Since the family $(\mathbf{1}[Y_i \geqslant j] : i \in I)$ is negatively associated,

$$M_{Z'_j}(t) \leqslant \prod_{i \in I_j} M_{\mathbf{1}[Y_i \geqslant j]}(t) \leqslant M^{|I_j|}_{\mathrm{Be}(e^{-\alpha j})}(t) = M_{\mathrm{Bi}(|I_j|, e^{-\alpha j})}(t)$$

for each $t \geqslant 0$. Hence, the usual Chernoff estimates for the upper tail for $\mathrm{Bi}(|I_j|, e^{-\alpha j})$ apply to Z'_j too, and thus

$$\mathbb{P}\big(Z'_j \geqslant 2|I|e^{-(\alpha/3)j}\big) \leqslant \exp\big(-\tfrac{1}{3}|I|e^{-(\alpha/3)j}\big); \tag{6.4}$$

see, for example, Corollary 2.4 (and its proof: see Theorems 2.8 and 2.10) of [10]. For $j \leqslant (2/\alpha)\log|I|$, we have $|I|e^{-(\alpha/3)j} \geqslant |I|^{1/3} \geqslant (c_{10}n)^{1/3}$, and thus

$$\mathbb{P}\big(Z'_j \geqslant 2|I|e^{-(\alpha/3)j}\big) \leqslant \exp\big(-c_{11}n^{1/3}\big). \tag{6.5}$$

For $j > (2/\alpha)\log|I|$, we use Markov's inequality and (6.3), which yield

$$\mathbb{P}(Z'_j > 0) \leqslant \mathbb{E}\,Z'_j \leqslant |I|e^{-\alpha j}. \tag{6.6}$$

Summing (6.5) or (6.6) for $j \geqslant 0$ yields

$$\begin{aligned} &\mathbb{P}\big(Z'_j > 2|I|e^{-(\alpha/3)j} \text{ for some } j \geqslant 0\big) \\ &\quad\leqslant \sum_{j \leqslant (2/\alpha)\log|I|} \exp\big(-c_{11}n^{1/3}\big) + \sum_{j > (2/\alpha)\log|I|} |I|e^{-\alpha j} \\ &\quad= o(1) + O(|I|^{-1}) \to 0 \end{aligned}$$

as $n \to \infty$, since $|I| \geqslant c_{12}n$. Hence

$$\mathbb{P}(Z'_j < 2|I|e^{-(\alpha/3)j} \text{ for each } j = 0, 1, \ldots) \to 1 \quad \text{as } n \to \infty.$$

This result together with (6.2) shows that, for some fixed $\alpha' > 0$,

$$\mathbb{P}(Z_j < 3e(H)e^{-\alpha' j} \text{ for each } j = 0, 1, \ldots) \to 1 \quad \text{as } n \to \infty,$$

which by (DE2′) for H easily implies that (DE2′) holds for G^+ w.h.p.

Hence, w.h.p. G^+ is a well-behaved weak α-decorated expander, and we may use Lemma 3.6 to see that then (3.1) holds for G^+, and consequently for $R_{n,c/n}$, which completes the proof of Theorem 1.6.

In Lemma 3.2, we may insist that the giant component satisfies (DE2) rather than just (DE2′): see [2]. Using this result, it is not hard to adapt the above proof to deduce that w.h.p. G^+ satisfies (DE2) rather than just (DE2′).

To set Theorem 1.6 in context, note that for any K there exists a constant $c_0 > 0$ such that if $0 < c \leqslant c_0$ then w.h.p. $\text{spread}(R_{n,c/n}) > K$. Indeed, we have the following result.

Proposition 6.2. *For any K there exists a constant $\varepsilon > 0$ such that the following holds. Let G_n be formed from the cycle C_n by adding at most εn edges. Then* $\text{spread}(G_n) \geqslant K$ *for n sufficiently large.* □

Proof. Let $t := \lceil \sqrt{6K} \rceil$, and assume that $0 < \varepsilon \leqslant 1/16t$. We shall show that $\text{spread}(G_n) \geqslant t^2/6 \geqslant K$ if n is sufficiently large.

We first define a Lipschitz function for C_n. It is convenient to let the vertex set of C_n be $V = \{0, 1, \ldots, n-1\}$. Divide V into $\lfloor n/4t \rfloor$ sections $\{0, \ldots, 4t-1\}$, $\{4t, \ldots, 8t-1\}$, ... plus a 'remainder' (possibly). If $i \in V$ satisfies $i \equiv j \pmod{4t}$ where $0 \leqslant j < 4t$, we set $f(i) = j$ if $0 \leqslant j < t$, $f(i) = 2t - j$ if $t \leqslant j < 3t$ and $f(i) = j - 4t$ if $3t \leqslant j < 4t$. (Thus, on the section $\{0, \ldots, 4t-1\}$, f increases from 0 to t, then decreases from t to $-t$ and then increases to -1, always taking unit steps.) Observe that

$$\sum_{j=0}^{4t-1} f(j)^2 = 4 \sum_{j=0}^{t-1} j^2 + 2t^2 = \frac{4}{3} t \left(t^2 + \frac{1}{2} \right).$$

Now re-set $f(v)$ to 0 for each v in the 'remainder' (that is, for $4\lfloor n/4t \rfloor \leqslant v \leqslant n$), and for each v in a section which contains a vertex of degree > 2 in G_n. Then f is a Lipschitz function for G_n, and f is unchanged on at least

$$\frac{n}{4t} - 1 - 2\varepsilon n \geqslant \frac{n}{8t} - 1$$

sections. Now $\sum_{v \in V} f(v) = 0$, and

$$\sum_{v \in V} f(v)^2 \geqslant \left(\frac{n}{8t} - 1 \right) \frac{4}{3} t \left(t^2 + \frac{1}{2} \right) \geqslant n \cdot t^2/6$$

for n sufficiently large. Then $\text{Var}(f) \geqslant t^2/6$, and so $\text{spread}(G_n) \geqslant t^2/6 \geqslant K$, as required. □

https://doi.org/10.1017/S0963548314000248 Published online by Cambridge University Press

7. $\mathbf{K}_n$ with random edge lengths

In this section we prove Theorem 1.7. Given $\alpha > 0$, following [2], we say that a family $\mathcal{A} = (a_i : i \in I)$ of non-negative numbers has an *α-exponential tail* if

$$\frac{|\{i \in I : a_i \geqslant j\}|}{|I|} \leqslant 2e^{-\alpha j} \quad \text{for all } j \geqslant 0.$$

We need two lemmas.

Lemma 7.1. *For each $\lambda > 0$ there is an $\alpha > 0$ such that the following holds. Consider $K_{n,n}$ with independent edge lengths X_e, where X_e is exponentially distributed with parameter λ/n (and thus mean n/λ). Then w.h.p. there is a perfect matching such that the edge lengths have an α-exponential tail.*

Proof. This follows from the result of Walkup [20] that, if each vertex independently and uniformly picks arcs to two vertices in the other part, then w.h.p. there is a perfect matching using only such arcs (ignoring orientations). (With minimal changes, we could allow each vertex to pick three arcs instead of two, and then the corresponding weakened version of Walkup's result follows directly from Hall's Theorem: see [20].)

Replace each edge of $K_{n,n}$ with a pair of oppositely directed arcs. Let the arcs e have independent edge lengths X'_e, each exponentially distributed with parameter $\lambda/2n$. For each edge e of $K_{n,n}$, we may assume that X_e is the minimum of X'_{e1} and X'_{e2}, where $e1$ and $e2$ are the two arcs arising from orienting e. Let S be the set of $4n$ arcs formed from the two shortest arcs leaving each vertex.

By Walkup's result [20], there is w.h.p. a perfect matching using only arcs in S. Let $Z_j = |\{e \in S : X'_e \geqslant j\}|$. It will suffice to show (by changing α) that

$$\mathbb{P}\big(Z_j < 16ne^{-\alpha j/3} \text{ for each } j\big) \to 1 \quad \text{as } n \to \infty. \tag{7.1}$$

Let Y_n be the second smallest of n independent random variables $\tilde{X}_1, \dots, \tilde{X}_n$ which are each exponentially distributed with parameter $\lambda/2n$. Let $p = \mathbb{P}(\tilde{X}_1 < j) = 1 - e^{-\lambda j/2n}$, and observe that $p \leqslant \lambda j/2n$. Then

$$\begin{aligned}\mathbb{P}(Y_n \geqslant j) = \mathbb{P}(\mathrm{Bin}(n,p) \leqslant 1) &= (1-p)^n + np(1-p)^{n-1} \\ &\leqslant \left(1 + \frac{1}{2}\lambda j e^{\lambda j/2n}\right) \cdot e^{-\lambda j/2}.\end{aligned}$$

Thus there is a constant $\alpha > 0$ such that

$$\mathbb{P}(Y_n \geqslant j) \leqslant 2e^{-\alpha j} \quad \text{for each } j \geqslant 0. \tag{7.2}$$

Let $\tilde{Y}_1, \dots, \tilde{Y}_{2n}$ be independent, each distributed like Y_n. Let

$$\tilde{Z}_j := \sum_{i=1}^{2n} \mathbf{1}[\tilde{Y}_i \geqslant j],$$

and note that $Z_j \leqslant_s 2\tilde{Z}_j$ (recall that $\leqslant_s$ denotes stochastic domination). Then, by (7.2),

$$\tilde{Z}_j \leqslant_s \mathrm{Bin}(2n, 1 \wedge 2e^{-\alpha j}).$$

The remainder of the proof is quite similar to the end of the proof of Theorem 1.6. By a Chernoff estimate,

$$\mathbb{P}(\tilde{Z}_j \geqslant 8ne^{-\alpha j/3}) \leqslant \exp\big(-\tfrac{1}{6} \cdot 8ne^{-\alpha j/3}\big).$$

When $j \leqslant (2/\alpha)\log n$ this is $\leqslant \exp(-n^{1/3})$. For larger j we simply use Markov's inequality,

$$\mathbb{P}(\tilde{Z}_j > 0) \leqslant \mathbb{E}\,\tilde{Z}_j \leqslant 4ne^{-\alpha j},$$

and thus

$$\sum_{j \geqslant (2/\alpha)\log n} \mathbb{P}(\tilde{Z}_j > 0) \leqslant \frac{4}{1-e^{-\alpha}} n^{-1}.$$

It follows that

$$\mathbb{P}\big(\tilde{Z}_j \geqslant 8ne^{-\alpha j/3} \text{ for some } j \geqslant 0\big) \leqslant \sum_{j=0}^{\infty} \mathbb{P}(\tilde{Z}_j \geqslant 8ne^{-\alpha j/3}) \to 0$$

as $n \to \infty$, and (7.1) follows, completing the proof. □

Lemma 7.2. *Fix $0 < \gamma < 1$. Fix $\lambda > 0$. Then there is an $\alpha > 0$ such that the following holds. Consider a complete bipartite graph $K_{a,b}$, with parts A of size a and B of size b, where $\gamma n \leqslant a, b \leqslant n$. Let the edges e have independent lengths $\ell(e)$, each exponentially distributed with parameter λ/n. Then w.h.p. there is a set of edges $S = \{uw\} \subset A \times B$ such that:*

(a) $|\{u \in A : uw \in S\}| = 1$ *for each* $w \in B$,
(b) $|\{w \in B : uw \in S\}| \leqslant \lceil b/a \rceil$ *for each* $u \in A$,
(c) *the family* $(\ell(uw) : uw \in S)$ *has* α*-exponential tails.*

Proof. By considering adding at most a vertices to B, we see that it suffices to consider the case $a|b$. But now we see that it suffices to assume that $a = b$, and so the result follows from the last lemma. □

Now we are ready to prove Theorem 1.7. If X is uniform on $[n]$ and Y is exponentially distributed with parameter $1/n$, then $X \leqslant_s 1 + Y$; thus we may assume that edge lengths are i.i.d., each distributed like $1 + Y$. Next, replace each edge e by a blue copy e_B and a red copy e_R, and give these copies i.i.d. edge lengths, each distributed like $1 + Y'$, where Y' is exponentially distributed with parameter $1/2n$. We may assume that the length $\ell(e)$ of e is the smaller of the lengths of e_R and e_B. Note that if

$$b = b(n) = -2n\log\left(1 - \frac{2}{n}\right)$$

then $b \sim 4$ and

$$\mathbb{P}(Y' \leqslant b) = 1 - e^{-b/2n} = \frac{2}{n}.$$

Thus, by keeping only blue edges with an appropriate length $b + 1 \leqslant 6$ (for large n), we may generate a random graph $G_{n,2/n}$.

https://doi.org/10.1017/S0963548314000248 Published online by Cambridge University Press

For some $\alpha_1 > 0$, w.h.p. there is in this random graph a giant component H and a subgraph F showing that H is a well-behaved weak α_1-decorated expander: see Lemma 3.5. Condition on there being such an H and F, and fix them. We also assume that (P1) holds, *i.e.*, $n' > \gamma n/2$: see Lemma 3.3. Thus, $v(F) \geqslant \alpha_1 n' \geqslant c_{13} n$. List the decorations as $D_1, \dots, D_\nu$. Let $W = [n] \setminus V(H)$.

Now we use the red edges. By Lemma 7.2 applied to the red edges between $V(F)$ and W, there is a set S of red edges $\{uw\} \subset V(F) \times W$ such that:

(a) $|\{u \in V(F) : uw \in S\}| = 1$ for each $w \in W$,
(b) $|\{w \in W : uw \in S\}| \leqslant \lceil |W|/|F| \rceil \leqslant 1/\alpha$ for each $u \in V(F)$,
(c) the family $(\ell(uw) : uw \in S)$ has α_2-exponential tails (for a suitable $\alpha_2 > 0$).

Let G be the graph on $[n]$ with edge set $E(H) \cup S$. We still have the subgraph F and decorations $D_1, \dots, D_\nu$, but now for each edge $uw \in S$ we have a new one-vertex decoration $\{w\}$ decorating $u \in V(F)$. For $i = 1, \dots, \nu$ let $\tilde{v}(D_i) = v(D_i)$, and for each $w \in W$ let $\tilde{v}(\{w\}) = \ell(uw)$ $(\geqslant 1)$, where $uw \in S$. Now use D_i to refer to any of the $\nu + |W|$ decorations of G. Then G is a well-behaved weak α_3-decorated expander, for a suitable α_3, except that in condition (DE2$'$), $v(D_i)$ is replaced by $\tilde{v}(D_i)$.

To show that each Lipschitz function for G then satisfies inequality (3.1), which implies Theorem 1.7, we follow the proof of Lemma 3.6. We need no changes until just after inequality (3.5) when the proof starts to deal with decorations. From there until inequality (3.8), replace each appearance of v by $\tilde{v}$. Now the proof works just as before.

8. Open problems

We saw in the preceding sections that high degree is precisely what is needed to force the spread of the random regular graph $G(n,d)$ to be close to $1/4$. We believe that a corresponding result should hold for the giant component $H_{n,c/n}$ of $G_{n,c/n}$.

Problem 8.1. Is it the case that for each $\varepsilon > 0$ there exists c_0 such that, for each $c \geqslant c_0$, w.h.p. spread$(H_{n,c/n}) < 1/4 + \varepsilon$?

If Lemma 3.2 holds uniformly (in the sense that for any $c > 1$, $\alpha = \alpha(c)$ can be chosen such that the conclusions of the theorem hold in $H_{n,c'/n}$ for all $c' \geqslant c$, with this value of α), then the proof of Theorem 3.1 can be modified to yield an affirmative answer to the above question. This uniformity seems very likely to hold, but does not follow immediately from the proof of Lemma 3.2 given in [2].

There is a natural similar question for the 'small world' random graph $R_{n,c/n}$, to complete the picture described in Theorem 1.6 and Proposition 6.2.

Problem 8.2. Is it the case that for each $\varepsilon > 0$ there exists c_1 such that, for each $c \geqslant c_1$, w.h.p. spread$(R_{n,c/n}) < 1/4 + \varepsilon$?

Theorem 1.2 suggests that the spread of $G(n,d)$ might converge (in probability) to a constant, and similarly for Theorem 1.3 and $H_{n,c/n}$.

Problem 8.3. Do there exist constants α_d for each $d \geqslant 3$ and β_c for each $c > 1$ such that $\text{spread}(G(n,d)) \xrightarrow{\text{p}} \alpha_d$ and $\text{spread}(H_{n,c/n}) \xrightarrow{\text{p}} \beta_c$ as $n \to \infty$?

We know that if the constants α_d exist then they are (weakly) decreasing in d and tend to $1/4$ as $d \to \infty$. It seems likely that the analogous result should hold for $G_{n,c/n}$.

Problem 8.4. If the constants β_c exist, are they decreasing in c, and do they tend to $1/4$ as $c \to \infty$? If so, how quickly?

A related problem is to determine the rate with which the spread of $G(n,d)$ approaches $1/4$ as d becomes large.

Problem 8.5. What is the best possible dependence of d_0 on ε in Theorem 1.5.

Again, there are natural similar questions for $R_{n,c/n}$.

For $R_{n,c/n}$, we can also ask about the constant $C_3(c)$ in Theorem 1.6 as $c \to 0$. Proposition 6.2 shows that $C_3(c) \to \infty$ as $c \to 0$. The proof of Theorem 1.6 in Section 6 yields, through the argument in Step 1 of the proof with, for example, $k = \lceil 3/c \rceil$, that we can take $C_3(c) = O(c^{-2})$ as $c \to 0$. We conjecture that this is best possible, in analogy with Theorem 1.4 for $G_{n,c/n}$.

Problem 8.6. Is the optimal $C_3(c) = \Theta(c^{-2})$ as $c \to 0$?

In the small worlds model $R_{n,c/n}$ we start with the deterministic cycle C_n and add edges independently with probability c/n. Suppose that we start instead with a deterministic graph G_n on $[n]$: let us denote the corresponding random graph by $R(G_n, c/n)$, so $R_{n,c/n}$ is $R(\mathsf{C}_n, c/n)$. For example a popular small worlds model takes G_n as a power C_n^r of C_n, where two vertices are adjacent in C_n^r if they are at distance at most r in C_n.

We may adapt the proof of Theorem 1.6 to show the same result when G_n is the n-vertex path P_n; that is, there is a constant $C_{18} = C_{18}(c) > 0$ such that w.h.p. $\text{spread}(R(\mathsf{P}_n, c/n)) \leqslant C_{18}$. (Indeed, we could take G_n as C_n less any set of edges which are at distance at least $C_{19} \log n$ apart, for a sufficiently large constant C_{19} depending on c. For we could think of these edges as simply being coloured red, and w.h.p. no two red edges appear on the cycle in the same path Q_w between vertices in $V(H)$. If there is a red edge in Q_w, we of course join the part before the red edge to H by an edge from the first vertex in Q_w to its predecessor in the cycle.)

It seems likely that starting with the path P_n is the worst case, which leads to the following problem.

Problem 8.7. Is it the case that w.h.p. $\text{spread}(R(G_n, c/n)) \leqslant C_{20}(c)$ for every sequence G_n of connected graphs on $[n]$ and every $c > 0$?

https://doi.org/10.1017/S0963548314000248 Published online by Cambridge University Press

Acknowledgements

This work was started during the Oxford–Princeton workshop on Combinatorics in Oxford, June 2008, continued during visits of CM and SJ to the CRM in Montreal in September and October 2008, and finally completed following a workshop in Bellairs in March 2012.

References

[1] Alon, N., Boppana, R. and Spencer, J. (1998) An asymptotic isoperimetric inequality. *Geom. Funct. Anal.* **8** 411–436.
[2] Benjamini, I., Kozma, G. and Wormald, N. C. (2014) The mixing time of the giant component of a random graph. *Random Struct. Alg.* doi:10.1002/rsa.20539.
[3] Bobkov, S. G., Houdré, C. and Tetali, P. (2006) The subgaussian constant and concentration inequalities. *Israel J. Math.* **156** 255–283.
[4] Bollobás, B. (2001) *Random Graphs*, second edition, Vol. 73 of *Cambridge Studies in Advanced Mathematics*, Cambridge University Press.
[5] Ding, J., Kim, J. H., Lubetzky, E. and Peres, Y. (2010) Diameters in supercritical random graphs via first passage percolation. *Combin. Probab. Comput.* **19** 729–751.
[6] Dubhashi, D. and Ranjan, D. (1998) Balls and bins: A study in negative dependence. *Random Struct. Alg.* **13** 99–124.
[7] Durrett, R. (2010) *Random Graph Dynamics*, Cambridge University Press.
[8] Janson, S. and Luczak, M. J. (2007) A simple solution to the k-core problem. *Random Struct. Alg.* **30** 50–62.
[9] Janson, S., Knuth, D. E., Łuczak, T. and Pittel, B. (1993) The birth of the giant component. *Random Struct. Alg.* **4** 231–358.
[10] Janson, S., Łuczak, T. and Rucinski, A. (2000) *Random Graphs*, Wiley.
[11] Joag-Dev, K. and Proschan, F. (1983) Negative association of random variables, with applications. *Ann. Statist.* **11** 286–295.
[12] Kolchin, V. F. (1986) *Random Mappings*, Translation Series in Mathematics and Engineering, Optimization Software Inc., New York. Translated from the Russian.
[13] Łuczak, T. (1990) Component behavior near the critical point of the random graph process. *Random Struct. Alg.* **1** 287–310.
[14] Łuczak, T. (1991) Size and connectivity of the k-core of a random graph. *Discrete Math.* **91** 61–68.
[15] Łuczak, T. (1991) Cycles in a random graph near the critical point. *Random Struct. Alg.* **2** 421–439.
[16] Newman, M. E. J. and Watts, D. J. (1999) Renormalization group analysis of the small-world network model. *Phys. Lett. A* **263** 341–346.
[17] Pavlov, Y. L. (1977) The asymptotic distribution of maximum tree size in a random forest (in Russian). *Teor. Verojatnost. i Primenen.* **22** 523–533. English translation in *Theoret. Probab. Appl.* **22** (1977) 509–520.
[18] Pavlov, Y. L. (1996) *Random Forests* (in Russian), Karelian Centre Russian Academy of Sciences, Petrozavodsk. English translation (2000), VSP, Zeist.
[19] Riordan, O. and Wormald, N. C. (2010) The diameter of sparse random graphs. *Combin. Probab. Comput.* **19** 835–926.
[20] Walkup, D. W. (1980) Matchings in random regular bipartite digraphs. *Discrete Math.* **31** 59–64.
[21] Watts, D. J. and Strogatz, S. H. (1998) Collective dynamics of ‘small-world’ networks. *Nature* **393** 440–442.