Colouring random subgraphs

Boris Bukh, Michael Krivelevich, and Bhargav Narayanan

Abstract. We study several basic problems about colouring the p-random subgraph G_p of an arbitrary graph G, focusing primarily on the chromatic number and colouring number of G_p. In particular, we show that there exist infinitely many k-regular graphs G for which the colouring number (i.e., degeneracy) of $G_{1/2}$ is at most $k/3 + o(k)$ with high probability, thus disproving the natural prediction that such random graphs must have colouring number at least $k/2 - o(k)$.

1. Introduction

For a graph G and $p \in (0,1)$, let G_p denote the random subgraph of G obtained by randomly including each edge of G independently with probability p. Here, we shall study some basic questions about properly colouring the vertices of G_p for fixed $p \in (0,1)$. Studying typical coloring properties of a random subgraph of a graph with given parameters is a very natural setup in the wide context of random graphs. We are also motivated partly by the following old question of Erdős and Hajnal [11, 12] that remains frustratingly open: is it true that for every pair $t, g \in \mathbb{N}$, there exists a $k = k(t,g) \in \mathbb{N}$ such that any graph with chromatic number at least k contains a subgraph with chromatic number at least t and girth at least g? A natural step towards this question of Erdős and Hajnal — motivated by Erdős' randomised construction (see [3]) of graphs of large girth and chromatic number — is to study the colouring-related properties of the random subgraph G_p of an arbitrary graph G of large chromatic number.

Concretely, we shall focus on the two problems that we next describe. First, we study the following 'chromatic number problem': for $k \to \infty$, given an arbitrary graph G with chromatic number $\chi(G) = k$, what can we say (asymptotically) about the chromatic number $\chi(G_p)$ of the random graph G_p? Second, as a more approachable weakening of the chromatic number problem where we restrict our attention to *greedy* colourings, we also study the following 'colouring number problem': for $k \to \infty$, given an arbitrary graph G with minimum degree $\delta(G) = k$, what can we say (again, asymptotically) about the colouring number $\mathcal{C}(G_p)$ of the random graph G_p? Here and later, the colouring number $\mathcal{C}(G)$ of a graph G is the minimum integer k such that every subgraph G' of

Date: 22 November, 2023.

2010 *Mathematics Subject Classification.* Primary 05C80; Secondary 05C15.

G has a vertex of degree less than k. Equivalently, the t-core of G is the maximal subgraph of G in which all vertices have degree at least t; the coloring number is the largest k such that the $(k-1)$-core is non-empty.

Towards the first of our two primary questions, the main problem — popularised by the first author, but certainly natural enough to have been independently considered by other researchers — is the following.

Problem 1.1. *As $k \to \infty$, is it true that for any graph G with $\chi(G) = k$, we have*

$$\mathbb{E}[\chi(G_{1/2})] = \Omega(k/\log k)?$$

The lower bound of $k/\log k$ in Problem 1.1 is natural, and best possible if true; indeed, for the complete graph $G = K_k$, the classical result of Bollobás [7] pinning down the chromatic number of dense Erdős–Rényi random graphs asserts that $\chi(G_{1/2}) \sim k/(2\log_2 k)$ with high probability.

Problem 1.1 strikes us as a rather basic question; however, not much appears to be known, and the state of the art is as follows. First, for any graph G with $\chi(G) = k$, since $G_{1/2}$ and its complement (in G) have the same distribution, it follows from a simple product-colouring argument that $\mathbb{E}[\chi(G_{1/2})] \geq k^{1/2}$; a similar argument (using a random partition into r parts) shows that $\mathbb{E}[\chi(G_{1/r})] \geq k^{1/r}$ for any $r \in \mathbb{N}$. The argument in [1] shows that $\chi(G_{1/2}) = \Omega(k/\log n)$ holds with high probability, where n is the number of vertices of G. Finally, Mohar and Wu [16] have settled the fractional analogue of Problem 1.1 in the affirmative. Specifically, it was proven in [16] that if G has fractional chromatic number k, then with probability $1 - o_k(1)$ the fractional chromatic number of the random subgraph $G_{1/2}$ is at least $k/8\log_2(4k)$.

Our primary contribution towards the chromatic number problem is an extension of the work of Shinkar [17] studying 'large deviations' of $\chi(G_{1/2})$. Taken together, our results give bounds for the entire lower tail of $\chi(G_{1/2})$; the first and the third bounds in the result below are due to Shinkar, while our contribution here is a proof of the second bound.

Theorem 1.2. *For any graph G with $\chi(G) = k$, we have*

$$\mathbb{P}\left(\chi(G_{1/2}) \leq d\right) \leq \begin{cases} \exp(-\Omega((k^{1/2}-d)^2/k^{1/2})) & \text{for } k^{1/2}/2 \leq d \leq k^{1/2}, \\ \exp(-\Omega(k/d)) & \text{for } k^{1/3} \leq d \leq k^{1/2}/2, \text{ and} \\ \exp(-\Omega(k(k-d^3))/d^3) & \text{for } d \leq k^{1/3}. \end{cases}$$

Towards the second of our two primary questions, we raise the following problem.

Problem 1.3. *As $k \to \infty$, determine the largest $\mathfrak{D}(k)$ for which we have*

$$\mathbb{P}\left(\mathcal{C}(G_{1/2}) \geq \mathfrak{D}(k)\right) \geq 1/2$$

for all graphs G with $\delta(G) \geq k$.

In other words, Problem 1.3 asks the following: as $k \to \infty$, what is the best possible lower bound on (the probable value of) the colouring number $\mathcal{C}(G_{1/2})$ that holds for all graphs G with minimum degree $\delta(G) \geq k$? Problem 1.3 is the 'degree-analogue' of Problem 1.1, replacing proper colourings with (the more tractable) proper *greedy* colourings, and the chromatic number with the minimum degree. This is motivated in large part by the degree-analogue — due to Thomassen [18] and also wide open — of the aforementioned problem of Erdős–Hajnal [11, 12]: is it true that for every pair $t, g \in \mathbb{N}$, there exists a $k = k(t, g) \in \mathbb{N}$ such that any graph with average degree at least k contains a subgraph with average degree at least t and girth at least g?

Let us point out that Problem 1.3 and its variants arise naturally in some other (non-mathematical) contexts as well. First, several variants of Problem 1.3 have been studied by biologists, sociologists and theoretical computer scientists as models of 'cascading failures' in networks; see [6, 19] and the references therein, for example. Second, we note that Problem 1.3 can also be recast in the language of bootstrap percolation [9]. Bootstrap percolation on a graph G is a model — originating in statistical physics — for the spread of infection on G defined as follows: starting with an initially infected set of vertices A, infection spreads along the edges of G, where a vertex of G gets infected if the number of its (previously) infected neighbours in G exceeds a specified threshold, and A is said to percolate if all the vertices of G are eventually infected. There is by now a large body of (mathematical) work devoted to understanding the percolating sets for various graph families (see [4, 15], for example), and Problem 1.3 may also be rephrased in this language: given a graph G with $\delta(G) = k$, we are looking to understand for what $t = t(k)$ we can guarantee that the set A of vertices of degree at most t in $G_{1/2}$ percolates in bootstrap percolation on $G_{1/2}$ with the threshold $(\deg(v, G_{1/2}) - t)$ at each vertex v of $G_{1/2}$ (or in other words, for what $t = t(k)$ we can guarantee that the t-core of $G_{1/2}$ is non-empty).

It is clear from considering the complete graph $G = K_{k+1}$ that we have the upper bound $\mathfrak{D}(k) \leq k/2 + o(k)$. On the other hand, if an n-vertex graph G satisfies $\delta(G) = k$, then G has $\geq kn/2$ edges, and so Chernoff bound [3, Theorem A.1.1] implies that $\Pr[e(G_{1/2}) < kn/4 - \sqrt{kn/2}] < e^{-2} < 1/2$; using this and the well-known fact that any graph of average degree d contains a subgraph of minimum degree at least $d/2$, it follows that $\mathfrak{D}(k) \geq k/4 - o(k)$.

While the upper bound of $\mathfrak{D}(k) \leq k/2 + o(k)$ seems like the natural guess for the truth, the following result — our main contribution towards the colouring number problem, and our most significant result here — shows that this is not the case.

Theorem 1.4. *As $k \to \infty$, we have $\mathfrak{D}(k) \leq k/3 + o(k)$.*

In more detail, the proof of Theorem 1.4 shows (for all $k \in \mathbb{N}$ divisible by 3) that there exist large k-regular graphs G for which the t-core of $G_{1/2}$ is empty with high

probability for some $t = k/3 + o(k)$. Our next result, stated below, serves to illustrate some of the subtleties that arise in studying Problem 1.3.

Theorem 1.5. *For every $\alpha > 0$, there exists a $\beta > 0$ such that for infinitely many $k \in \mathbb{N}$, there exist arbitrarily large k-regular graphs G for which the following holds with high probability (as $k \to \infty$): any induced subgraph H of $G_{1/2}$ with $\delta(H) \geq k/4 + \alpha k$ satisfies $|V(H)|/|V(G)| = O((1-\beta)^{k^2})$.*

In other words, Theorem 1.5 asserts (for infinitely many $k \in \mathbb{N}$) that there exist large k-regular graphs G for which the t-core of $G_{1/2}$ is *just barely* non-empty (i.e., is very small relative to G) for any $t = k/4 + o(k)$. In the light of this, it seems clear to us that improving on the easy lower bound of $\mathfrak{D}(k) \geq k/4 + o(k)$ sketched above is likely to require some interesting ideas.

This paper is organised as follows. After covering some preliminaries in Section 2, we give the proof of Theorem 1.2 in Section 3 and the proofs of Theorem 1.4 and Theorem 1.5 in Section 4. Finally, we conclude in Section 5 with a discussion of open problems and directions for further work.

2. Preliminaries

We start by establishing some notation and collecting together some tools that we will rely on in the sequel.

Our graph theoretic notation is for the most part standard; we refer the reader to [8] for terms not defined here. That said, we remind the reader of a few standard notions that come up frequently in this paper.

First, recall that the *chromatic number* $\chi(G)$ of a graph G is the smallest number of colours needed to properly colour the vertices of G, i.e., to colour the vertices in such a way that no two adjacent vertices share the same colour.

Next, following Erdős and Hajnal [13], the *colouring number* $\mathcal{C}(G)$ of a graph G is the least number c for which there exists an ordering of the vertices of G in which each vertex has fewer than c neighbours preceding it in the ordering; this parameter — also (essentially) called the degeneracy or the core number — is the number of colours used by the natural greedy algorithm for properly colouring the vertices of G.

We shall also need two notions of graph boundaries: for a subset $S \subset V(G)$ of the vertices of a graph G, its *vertex boundary* ∂S consists of those vertices of G not in S that are adjacent to at least one vertex in S, and its *edge boundary* ∇S consists of those edges of G with one end in S and the other in $\overline{S}$. Overloading this notation slightly, for a subset $S \subset V(G)$ of the vertices of a directed graph G, its vertex boundary ∂S consists of those out-neighbours of S that are not in S, and similarly, its edge boundary ∇S consists of those edges of G directed from S to $\overline{S}$.

We need a standard bound for the number of connected components in a graph of given maximum degree; it can be found, e.g., in [5] (see Lemma 2 there), along with a proof.

Lemma 2.1. *For a graph G of maximum degree Δ, the number of connected, t-edge subgraphs of G containing a given vertex is less than $(e\Delta)^t$.*

Another fairly standard fact we utilize is a quantitative connection between eigenvalues and edge distribution in regular graphs. For a graph G its eigenvalues are those of its adjacency matrix $A(G)$. The following bound is due to Alon and Milman [2].

Lemma 2.2. *Let G be a d-regular graph on n vertices with the second largest eigenvalue λ. Then for every subset $S \subset V$, one has*

$$|\nabla S| \geq \frac{(d-\lambda)|S|(n-|S|)}{n}.$$

3. Chromatic number

First, following [1], we record (in slightly greater generality) a proof of the fact that for any n-vertex graph G with $\chi(G) = k$, we have $\chi(G_{1/2}) = \Omega(k/\log n)$ with high probability (as $n \to \infty$).

Proposition 3.1. *For any n-vertex graph G with $\chi(G) = k$ and any $0 < p < 1$, we have*

$$\chi(G_p) \geq \frac{pk}{2\log n}$$

with high probability (as $n \to \infty$).

Proof. The probability that there exists a set $V' \subset V(G)$ for which

(a) the induced subgraph $G[V']$ has minimum degree at least $2\log n/p$, and
(b) V' becomes an independent set in G_p,

is at most

$$\sum_{m=2\log n/p}^{n} \binom{n}{m}(1-p)^{(m\log n)/p} \leq \sum_{m=2\log n/p}^{n} \left(\frac{en}{m}\cdot\frac{1}{n}\right)^m = o(1).$$

As every t-chromatic graph contains a subgraph of minimum degree at least $t-1$ (a color-critical subgraph), the above implies that with high probability, any independent set V' in G_p induces a subgraph of G_p of chromatic number at most $(2\log n)/p$ in G. As G_p can be partitioned into $\chi(G_p)$ independent sets (by definition), the result follows. □

We now prove Theorem 1.2.

Proof of Theorem 1.2. As mentioned, our contribution is the second bound in the statement of the theorem. Let $G = (V, E)$ be a graph with $\chi(G) = k$. Our goal is to estimate from above the probability $\mathbb{P}(\chi(G_{1/2}) \leq d)$.

Fix an optimal coloring $V = V_1 \cup \ldots \cup V_k$ of G. Equipartition $[k] = I_1 \cup \ldots \cup I_s$ with $|I_j| \geq 2d^2$ and $s = \Theta(k/d^2)$. Set $G_j = G[\cup_{i \in I_j} V_i]$ for $1 \leq j \leq s$, and note that $\chi(G_j) = |I_j| \geq 2d^2$. If $\chi(G_{1/2}) \leq d$, then the chromatic numbers of all the random subgraphs $(G_j)_{1/2}$ are at most d. Observe crucially that these events are independent as the graphs G_j do not share any vertices, and thus edges. Hence

$$\mathbb{P}(\chi(G_{1/2}) \leq d) \leq \prod_{i=1}^{s} \mathbb{P}(\chi((G_i)_{1/2}) \leq d)\,.$$

Recall that $\chi(G_j) \geq 2d^2$. Hence, as explained in the introduction, $\mathbb{E}[\chi((G_j)_{1/2})] \geq \sqrt{2}d$. Using the first bound in the statement of the theorem (proved using Doob martingales as outlined by Shinkar [17]), we get $\mathbb{P}(\chi((G_i)_{1/2}) \leq d) \leq e^{-cd}$ for some absolute constant $c > 0$. It follows that

$$\mathbb{P}(\chi(G_{1/2}) \leq d) \leq \left(e^{-cd}\right)^s = \exp(-\Theta(k/d))\,,$$

as required. □

A twist on the above idea also provides a new and fairly simple proof of the third statement of Theorem 1.2 for the range $d = O(k^{1/4})$. Here is an outline. For G with an optimal coloring $V = V_1 \cup \ldots \cup V_k$, fix a collection of subsets $I_1, \ldots, I_s \subset [k]$ with $|I_j| \geq 2d^2$ and $s = \Theta(k^2/d^4)$ so that $|I_i \cap I_j| \leq 1$ for every $1 \leq i \neq j \leq s$; the existence of such a collection is a fairly standard fact in design theory. Let $G_j = G[\cup_{i \in I_j} V_i]$, $1 \leq j \leq s$. The events $A_j = \{\chi((G_j)_{1/2}) \leq d\}$ are again independent, and each happens with probability at most e^{-cd}. It follows that $\mathbb{P}(\chi(G_{1/2}) \leq d) \leq (e^{-cd})^s = \exp(-\Theta(k^2/d^3))$.

4. Colouring number

Our proofs of Theorem 1.4 and Theorem 1.5 rely on the existence of good expander graphs. Here, we make use of specific graphs that happen to be Ramanujan, but any family of sufficiently strong expanders should suffice.

We start with the proof of Theorem 1.4.

Proof of Theorem 1.4. Since $\mathfrak{D}(k)$ is a non-decreasing function of k, it suffices to only consider k that are divisible by 3. Given any small $0 < \alpha < 1/100$ and $k \in \mathbb{N}$ with $3 \mid k$, we shall construct a k-regular graph G (infinitely many, in fact) with the property that, for $t = k/3 + \alpha k$, the t-core of $G_{1/2}$ is empty with probability $1 - o(1)$ as $k \to \infty$; clearly, this suffices to prove the result.

We need the following well-known fact: there are positive constants c', c'' such that for infinitely many n there is a 3-regular graph H on n vertices without cycles shorter

than $c' \log n$, and with all eigenvalues λ_i but the first one $\lambda_1 = 3$ satisfying $\lambda_i \leq 3 - c''$. See [10], say, for a proof of this fact.

Choose H as above, and let G be a $(k/3)$-blow-up of H, i.e., G is obtained from H by replacing each vertex of H by an independent set of size $k/3$ — we call these sets (and interchangeably, the vertices of H) *super-vertices* — and by replacing each edge of H by a complete bipartite graph in G between the corresponding super-vertices.

In order to help the reader to grasp our argument, let us state that it implements and analyses the following bootstrap percolation-type process on H. First, we form a random subset R of *protected edges* of H, where an edge $e \in E(H)$ is declared protected independently and with probability $p = \exp\{-\Theta(k)\}$; protected edges correspond to complete bipartite graphs between the super-vertices of H in which in the random subgraph $G_{1/2}$ there is a vertex of degree at least $k/6 + \alpha k/2$; clearly the events corresponding to the edges of H becoming protected are independent for different edges of H, and happen each with probability exponentially small in k. Then a random vertex r of H is chosen; in the argument this will be a super-vertex of H all of whose incident edges get erased in the first round of deletions. Now, consider the following propagation process. We start with $V_0 = \{r\}$, and at each step update V_0 by adding to it all the vertices of H outside of V_0 that have at least two neighbours in V_0, or alternatively have at least one neighbor in V_0 and are not incident to any protected edge from R. We will prove that if H is a good expander with logarithmic girth, then typically, the above propagation process ends with V_0 consuming all the vertices of H; this corresponds to the random subgraph $G_{1/2}$ having an empty $(k/3 + \alpha k)$-core, as desired.

We say that a vertex of G *survives* or *lives* if it is present in the t-core of $G_{1/2}$, and that it *dies* otherwise; similarly, we say that a super-vertex of H *dies* if none of its constituent vertices survive, and that it *lives* or *survives* otherwise.

We shall, for technical reasons, construct $G_{1/2}$ by deleting the edges of G in two rounds: in the first round, each edge of G is independently sampled with probability $\alpha/3$, in the second round, each edge of G is independently sampled with probability $(1/2 - \alpha/3)/(1 - \alpha/3) \geq 1/2 - \alpha/2$, and finally, all the sampled edges are deleted to form $G_{1/2}$.

Since $n \geq (\alpha/3)^{-k^3}$, there is, with high probability *over the random deletions in the first round*, some super-vertex for which all the edges incident to it in G are deleted in the first round. Therefore, let us condition on the event that all the edges incident to some super-vertex are deleted in the first round; let r be any such super-vertex. Clearly, such an r dies. Let T_r be the connected set of dead super-vertices containing r; we claim that $|T_r| = n$ with high probability *over the random deletions in the second round*. Since the two rounds of deletions are independent, this claim clearly implies that the t-core of $G_{1/2}$ is empty with high probability.

The rest of the proof is devoted to the proof of the claim above, namely that for any fixed super-vertex r, conditional on r dying after the first round of deletions, the second round of deletions guarantee that $|T_r| = n$ with high probability. In what follows, we fix an arbitrary super-vertex r, abbreviate T_r by T, and write $\mathbb{P}_r$ for the probability over the random deletions in the second round, conditioned on r dying in the first round.

We now need slightly different arguments based on how large $m = |T|$ might be. Before we turn to this, we observe that

(1) if $m < n$, then since H is connected, the vertex boundary ∂T of T in H is both non-empty and necessarily contained in the set of surviving super-vertices, and
(2) for each surviving super-vertex $v \in \partial T$, there is at least one vertex $v^* \in V(G)$ contained in v that survives.

First, we handle the case where $1 \leq m < 99n/100$ by a union bound over the potential choices of T. Our task then is to bound, for all choices of T_0 with $|T_0| = m$, the probability $\mathbb{P}_r(T = T_0)$.

Consider any connected set T_0 of m super-vertices containing r. Due to our choice of H, Lemma 2.2, and since $|T_0| = m < 99n/100$, we know that $|\partial T_0| > c_1 m$ (for some universal $c_1 > 0$). Let $v_1, v_2, \ldots, v_\ell$ be a maximal independent set of surviving super-vertices in $H[\partial T_0]$, and note that since each vertex in $H[\partial T_0]$ has degree at most 2, we must have $\ell \geq |\partial T_0|/3 \geq c_1 m/3$. Next, note that if $T = T_0$, then there must exist vertices $v_1^* \in v_1, v_2^* \in v_2, \ldots, v_\ell^* \in v_\ell$ of G that also survive.

For any such choice of vertices $v_1^*, v_2^*, \ldots, v_\ell^*$, we shall now estimate the probability, over the second round of deletions, that these vertices survive. By virtue of how G is constructed from H, it is clear that v_i^* is not adjacent to any of the vertices $v_1^*, v_2^*, \ldots, v_{i-1}^*$ for all $1 \leq i \leq \ell$. This allows us to bound

$$\begin{aligned}
\mathbb{P}_r\left(v_1^*, \ldots, v_\ell^* \text{ survive} \mid T_0 \text{ dies}\right) &\leq \prod_{i=1}^{\ell} \mathbb{P}_r\left(\begin{array}{c}\text{at least } k/3 + \alpha k \text{ edges from } v_i^* \\ \text{to } \overline{T_0} \text{ survive the second round}\end{array}\right) \\
&\leq \prod_{i=1}^{\ell} \mathbb{P}\left(\mathrm{Binom}(2k/3, 1/2 + \alpha/2) \geq k/3 + \alpha k\right) \\
&\leq (1 + c_2)^{-k\ell}, \qquad (1)
\end{aligned}$$

where $c_2 > 0$ is a constant depending on α alone.

Using the fact that $\ell \geq c_1 m/3$, a union bound over all potential choices of $v_1^*, v_2^*, \ldots, v_\ell^*$ — of which there are at most $(k/3)^{2m}$ since $|\partial T_0| \leq 2|T_0| = 2m$ — yields the estimate

$$\mathbb{P}_r\left(T = T_0\right) \leq (k/3)^{2m}(1 + c_2)^{-c_1 km/3}.$$

Finally, the number of connected sets T_0 of size m that contain the fixed root r is, by Lemma 2.1, at most $(3e)^m$. Thus, it follows again from the union bound that

$$\mathbb{P}_r\left(1 \le |T| < 99n/100\right) \le \sum_{m=1}^{99n/100} \left(ek^2/3\right)^m (1+c_2)^{-c_1 km/3} = o(1), \tag{2}$$

with the last asymptotic estimate holding in the limit of $k \to \infty$.

Next, we deal with the possibility that $99n/100 \le m < n$. In this case, note that (by the definition of T), every super-vertex $v \in \overline{T}$ sends at most one edge to T, and hence has at least two neighbors in $\overline{T}$. Let S be a connected component in $H[\overline{T}]$ and put $s = |S|$; since S has minimum degree 2, it contains a cycle, and since H has girth at least $c' \log n$ (for some universal $c' > 0$), this implies that $0.01n \ge s = |S| \ge c' \log n$. Observe that since S is a connected component of $H[\overline{T}]$, it must be the case that $\partial S \subset T$. As $|\nabla S| \ge c''s$ for an absolute constant $c_3 > 0$, again due to our choice of H and Lemma 2.2. Since $|S| = s \le 0.01n$, and since each super-vertex of S has at most one neighbour outside S, we conclude that at least $c''s$ super-vertices in S have a neighbour in T, so $|\partial T \cap S| \ge c''s$. As before, we may find a set $I \subset \partial T \cap S$ of $c''s/3$ super-vertices that are independent in $H[\partial T]$. As we argued for (1), the probability of the super-vertices in I all surviving conditional on T dying is at most

$$\left((k/3)(1+c_2)^{-k}\right)^{c''s/3},$$

where $c_2 > 0$ is, exactly as before, a constant depending on α alone. Then, again invoking Lemma 2.1, by a union bound over the choice of a connected S in H of size s, and $I \subset S$ (which can be chosen in at most 2^s ways), we get

$$\mathbb{P}_r\left(99n/100 < |T| < n\right) \le \sum_{s=c' \log n}^{n/100} n(3e)^s 2^s \left((k/3)(1+c_2)^{-k}\right)^{c''s/3} = o(1), \tag{3}$$

with the last asymptotic estimate holding in the limit of $k \to \infty$. The desired claim, namely that $\mathbb{P}_r(|T| < n) = o(1)$, follows from (2) and (3), and the proof is complete. □

The bottleneck in the proof of Theorem 1.4 that we just saw comes from the tension between graph expansion and the impact of having dead neighbours. Specifically, instead of starting with 3-regular graphs and looking at the $(k/3 + \alpha k)$-core, if we started with ℓ-regular graphs and looked at the t-core, then having a dead super-neighbour would be a serious mortality risk only if $t > \frac{1}{2}(1 - 1/\ell)k$. Improving the argument would require using 2-regular expanders, which clearly do not exist.

To prove Theorem 1.5 we turn to *directed* expander graphs instead. There do exist directed expander graphs all whose in-degrees are equal to 2. A downside to this approach is that the gadgets we now use to form the super-vertices are more complex than mere independent sets. Consequently, these gadgets contain high-density subgraphs which have a non-neglible chance of surviving in the last phase of the deletion process

when $99n/100 \leq m < n$; this explains why a tiny number of vertices survive in Theorem 1.5.

We call a directed graph *d-regular* if the in-degree and out-degree of each vertex are d. The following lemma follows from a standard probabilistic construction.

Lemma 4.1. *For all sufficiently large n, there exist 2-regular directed graphs H on n vertices that, for every non-trivial subset $S \subset V(H)$ of vertices, satisfy*

$$|\partial S| \geq c_3 \min(|S|, n - |S|),$$

where $c_3 > 0$ is a universal constant.

Proof. A uniformly random 2-regular directed graph on n vertices has this property with high probability as $n \to \infty$. A proof of this fact is, at this point, a routine argument using the configuration model; see [14], for example, for a similar argument in the context of undirected graphs (that extends to the directed case as well). □

Proof of Theorem 1.5. Given any small $0 < \alpha < 1/16$, fix an integer $2/\alpha \leq s \leq 4/\alpha$, and let k be any large positive integer that is divisible by $2s$. We shall construct, for infinitely many $n \in \mathbb{N}$, a k-regular graph G on $nk(s+3)(1/2 - 1/2s)$ vertices with the property that, for

$$t = k/4 + 2k/s,$$

the t-core of $G_{1/2}$ has density at most $(1-\delta)^{k^2}$ with probability $1 - o(1)$ as $k \to \infty$, where $\delta > 0$ is a constant depending on α alone; clearly, this suffices to prove the result.

Let H be a 2-regular directed expander on $n \geq (\alpha/6)^{-k^3}$ vertices as promised by Lemma 4.1. To describe the blow-up process we use to construct G from H, we need to be able to distinguish the in-edges at each vertex of H; to that end, two-colour the edges of H (with colours red and blue, say) so that the two in-edges at each vertex are coloured differently. We then build G from H according to the procedure illustrated in (1) as follows; it is routine to verify that this construction indeed produces a k-regular graph.

(1) Replace each vertex v of H by a disjoint union of $s+3$ independent sets of size $k/2 - k/2s$ each; denote these independent sets by $I_1(v), \ldots, I_{s+3}(v)$.
(2) For each vertex v of H, place a complete bipartite graph between the sets $I_j(v)$ and $I_{j+1}(v)$ for each $1 \leq j \leq s+2$.
(3) For every *red* directed edge $u \to v$ in H and each $2 \leq j \leq s+2$, place an arbitrary $(k/2s)$-regular bipartite graph between the sets $I_j(u)$ and $I_1(v)$.
(4) For every *blue* directed edge $u \to v$ in H and each $2 \leq j \leq s+2$, place an arbitrary $(k/2s)$-regular bipartite graph between the sets $I_j(u)$ and $I_{s+3}(v)$.

To orient the reader, let us say that what follows is an analysis of the following bootstrap percolation-type process on H. We form a random subset R of the vertices

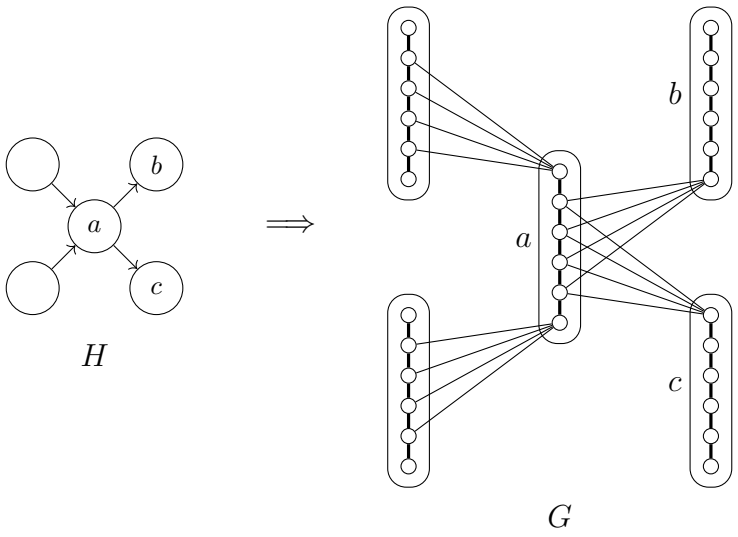

FIGURE 1. From H to G: thick edges are complete bipartite graphs of degree $k/2 - k/2s$, and thin edges are bipartite graphs of degree $k/2s$.

(namely, those termed 'resilient' in the sequel) by placing every vertex $v \in V(H)$ into R independently with probability $\exp\big(-\Theta(k^2)\big)$. Then, a random initial vertex $r \in V(H)$ is chosen, and at this point, we consider the following propagation process. We start with $V_0 = \{r\}$, and at each step, we update V_0 by adding to it every out-neighbour u of V_0 that satisfies $u \notin R$. The goal is to prove that typically V_0 grows to contain all but an exponentially small (in k^2) proportion of the vertices of H.

As in the proof of Theorem 1.4, we delete edges of G in two rounds: we sample the edges in the two rounds independently with probabilities $1/3s \geq \alpha/6$ and $(1/2 - 1/3s)/(1-1/3s) \geq 1/2 - 1/2s$ respectively, and then delete all the sampled edges. With the same notions of vertices and super-vertices *surviving* and *dying* as in the proof of Theorem 1.4, we assume that some super-vertex — we write r for such a super-vertex — dies in the first round with high probability; this is justified since $n \geq (\alpha/6)^{-k^3}$ is large enough to ensure this. We then write $\mathbb{P}_r$ to denote the probability over the random deletions in the second round, conditioned on r dying in the first round.

Note that the $s+3$ independent sets inside a super-vertex form a path; call a pair of adjacent independent sets $I_j(v)$ and $I_{j+1}(v)$ inside some super-vertex v *resilient* if either the set

$$\{v^* \in I_j(v) : \text{at least } k/4 \text{ edges from } v^* \text{ to } I_{j+1}(v) \text{ survive the second round}\}$$

or the set

$$\{v^* \in I_{j+1}(v) : \text{at least } k/4 \text{ edges from } v^* \text{ to } I_j(v) \text{ survive the second round}\}$$

has size at least k/s.

Call a super-vertex v *nearly dead* if each of the sets $I_j(v)$ for $2 \leq j \leq s+2$ contains fewer than k/s surviving vertices. Note that each dead super-vertex is also nearly dead, and that if v is nearly dead, then the vertices in the sets $I_3(v), I_4(v), \ldots, I_{s+1}(v)$ all die since each vertex therein is adjacent to fewer than

$$k/s + k/s + k/2s + k/2s < k/4$$

surviving vertices (with room to spare).

Writing T for the set of all nearly dead super-vertices that can be reached along a directed path starting at r in H, we observe the following.

Claim 4.2. *Each $v \in \partial T$ contains a resilient pair.*

Proof. Let $u \in T$ be a nearly dead in-neighbour of v, and consider the directed edge from u to v; without loss of generality, suppose that this edge is coloured red. Let $1 \leq \ell \leq s+3$ be the smallest index for which $I_\ell(v)$ contains at least k/s surviving vertices; by the definition of v not being nearly dead, such an ℓ exists and satisfies $\ell \leq s+2$.

Suppose for the sake of contradiction that the sets $I_\ell(v)$ and $I_{\ell+1}(v)$ do not form a resilient pair. Then $I_\ell(v)$ contains a surviving vertex v^* incident to fewer than $k/4$ edges into $I_{\ell+1}(v)$ that survive the second round of deletions; such a vertex v^* then has to be incident to more than $t - k/4 = 2k/s$ surviving vertices outside $I_{\ell+1}(v)$. We cannot have $\ell = 1$ because the sets $I_3(u), I_4(v), \ldots, I_{s+1}(u)$ are all dead, and the sets $I_2(u)$ and $I_{s+2}(u)$ contain fewer than k/s surviving vertices each, leaving v^* with fewer than $k/s + k/s = 2k/s$ surviving neighbours outside $I_{\ell+1}(v)$. We cannot have $2 \leq \ell \leq s+2$ either because such a v^* is adjacent to at most $k/2s$ vertices inside each of the two out-neighbours of v, and fewer than k/s vertices in $I_{\ell-1}(v)$ since this set contains fewer than k/s surviving vertices (by the minimality of ℓ), leaving v^* with fewer than $k/2s + k/2s + k/s = 2k/s$ surviving neighbours outside $I_{\ell+1}(v)$. Hence the sets $I_\ell(v)$ and $I_{\ell+1}(v)$ form a resilient pair, as desired. □

Call a super-vertex resilient if it contains a resilient pair, and let R be the set of resilient super-vertices. Since resilience of a super-vertex depends only on the edges of G inside the super-vertex, events of the form $\{v \in R\}$ are mutually independent for different super-vertices v. It is also clear that for each super-vertex v, we have

$$\begin{aligned}\mathbb{P}_r(v \in R) &\leq (s+2)\binom{k/2 - k/2s}{k/s}\mathbb{P}(\mathrm{Binom}(k/2 - k/2s, 1/2 + 1/2s) \geq k/4)^{k/s} \\ &\leq (1+c_4)^{-k^2}\end{aligned}$$

for some constant $c_4 > 0$ that depends on s (and thus α) alone.

This gives us a way to estimate the size of T: if $|T_0| \leq 99n/100$, then

$$\mathbb{P}_r(T = T_0) \leq \left((1+c_4)^{-k^2}\right)^{|\partial T_0|} \leq (1+c_4)^{-c_3|T_0|k^2/100},$$

from which it follows (as in the proof of Theorem 1.4) that

$$\mathbb{P}_r(|T| \leq 99n/100) \leq \sum_{m=1}^{99n/100} (3e)^m (1+c_4)^{-c_3 m k^2/100} = o(1)$$

as $k \to \infty$.

To see that the size of T must be very nearly n (and not just at least $99n/100$), we analyse its complement. By the definition of R and (4.2), we have $\partial T \subseteq R$. By Markov's inequality, $|R| < n(1+c_4/2)^{-k^2}$ with high probability as $k \to \infty$. By our choice of H satisfying Lemma 4.1, for every T with $n/2 \leq |T| \leq n - n(1+c_4/2)^{-k^2}/c_3$, we have $|\partial T| \geq n(1+c_4/2)^{-k^2}$. This tells us that

$$\left|\overline{T}\right|/n \leq (1+c_4/2)^{-k^2})/c_3 = (1-\beta)^{k^2},$$

with high probability as $k \to \infty$, where $\beta > 0$ is again a constant depending on α alone; this completes the proof. □

5. Conclusion

A large number of interesting open problems remain; below, we highlight a few that we find particularly appealing.

First, in the context of the chromatic number of problem, all the lower bounds on $\chi(G_{1/r})$ in terms of $\chi(G)$ that we currently have rely crucially on r being an integer. It would be very interesting to prove any lower bound for $\mathbb{E}[\chi(G_{0.499})]$ that is (asymptotically) better than $\chi(G)^{1/3}$.

Second, in the context of the colouring number of problem, we have been unable to prove any interesting lower bounds for $\mathfrak{D}(k)$. We would not be surprised if the truth is that $\mathfrak{D}(k) = k/3 + o(k)$ as $k \to \infty$, but even showing that $\mathfrak{D}(k) \geq k/3.99 + o(k)$ appears to be a challenging problem, as evidenced by Theorem 1.5. In fact, we (somewhat embarrassingly) do not know if there exists a function $f(k) \to \infty$ as $k \to \infty$ such that $\mathfrak{D}(k) \geq k/4 + f(k)$.

Acknowledgements

The first author was supported by NSF grants DMS-2154063 and DMS-1555149, and the third author was supported by NSF grants DMS-2237138 and CCF-1814409, as well as a Sloan Research Fellowship. Additionally, we also wish to thank Yury Person and Benny Sudakov for valuable discussions.

References

1. N. Alon, M. Krivelevich, and B. Sudakov, *Subgraphs with a large cochromatic number*, J. Graph Theory **25** (1997), 295–297. 2, 5
2. N. Alon and V. D. Milman, λ_1, *isoperimetric inequalities for graphs, and superconcentrators*, J. Combin. Theory Ser. B **38** (1985), 73–88. 5
3. N. Alon and J. H. Spencer, *The probabilistic method*, fourth ed., Wiley Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, 2016. 1, 3
4. J. Balogh and B. G. Pittel, *Bootstrap percolation on the random regular graph*, Random Structures Algorithms **30** (2007), 257–286. 3
5. A. Beveridge, A. Frieze, and C. McDiarmid, *Random minimum length spanning trees in regular graphs*, Combinatorica **18** (1998), 311–333. 5
6. L. Blume, D. Easley, J. Kleinberg, R. Kleinberg, and É. Tardos, *Which networks are least susceptible to cascading failures?*, 2011 IEEE 52nd Annual Symposium on Foundations of Computer Science—FOCS 2011, IEEE Computer Soc., Los Alamitos, CA, 2011, pp. 393–402. 3
7. B. Bollobás, *The chromatic number of random graphs*, Combinatorica **8** (1988), 49–55. 2
8. ———, *Modern graph theory*, Graduate Texts in Mathematics, vol. 184, Springer-Verlag, New York, 1998. 4
9. J. Chalupa, P. L. Leath, and G. R. Reich, *Bootstrap percolation on a Bethe lattice*, J. Phys. C **12** (1979), L31—-L35. 3
10. P. Chiu, *Cubic Ramanujan graphs*, Combinatorica **12** (1992), 275–285. 7
11. P. Erdős, *Problems and results in chromatic graph theory*, Proof Techniques in Graph Theory (Proc. Second Ann Arbor Graph Theory Conf., Ann Arbor, Mich., 1968), Academic Press, New York, 1969, pp. 27–35. 1, 3
12. ———, *Some unsolved problems in graph theory and combinatorial analysis*, Combinatorial Mathematics and its Applications (Proc. Conf., Oxford, 1969), Academic Press, London, 1971, pp. 97–109. 1, 3
13. P. Erdős and A. Hajnal, *On chromatic number of graphs and set-systems*, Acta Math. Acad. Sci. Hungar. **17** (1966), 61–99. 4
14. S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, Bull. Amer. Math. Soc. (N.S.) **43** (2006), 439–561. 10
15. S. Janson, T. Łuczak, T. Turova, and T. Vallier, *Bootstrap percolation on the random graph $G_{n,p}$*, Ann. Appl. Probab. **22** (2012), 1989–2047. 3
16. B. Mohar and H. Wu, *Fractional chromatic number of a random subgraph*, J. Graph Theory **95** (2020), 467–472. 2
17. I. Shinkar, *On coloring random subgraphs of a fixed graph*, arXiv:1612.04319, 2016. 2, 6

18. C. Thomassen, *Girth in graphs*, J. Combin. Theory Ser. B **35** (1983), 129–141. 3
19. D. J. Watts, *A simple model of global cascades on random networks*, Proc. Natl. Acad. Sci. USA **99** (2002), 5766–5771. 3

Department of Mathematics, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Email address: bbukh@math.cmu.edu

School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997801, Israel

Email address: krivelev@tauex.tau.ac.il

Department of Mathematics, Rutgers University, Piscataway, NJ 08854, USA

Email address: narayanan@math.rutgers.edu