

Rainbow Hamilton cycles in random graphs

Alan Frieze *

Po-Shen Loh †

Abstract

One of the most famous results in the theory of random graphs establishes that the threshold for Hamiltonicity in the Erdős-Rényi random graph $G_{n,p}$ is around $p \sim \frac{\log n + \log \log n}{n}$. Much research has been done to extend this to increasingly challenging random structures. In particular, a recent result by Frieze determined the asymptotic threshold for a loose Hamilton cycle in the random 3-uniform hypergraph by connecting 3-uniform hypergraphs to edge-colored graphs.

In this work, we consider that setting of edge-colored graphs, and prove a result which achieves the best possible first order constant. Specifically, when the edges of $G_{n,p}$ are randomly colored from a set of $(1 + o(1))n$ colors, with $p = \frac{(1+o(1))\log n}{n}$, we show that one can almost always find a Hamilton cycle which has the additional property that all edges are distinctly colored (rainbow).

1 Introduction

Hamilton cycles occupy a position of central importance in graph theory, and are the subject of countless results. In the context of random structures, much research has been done on many aspects of Hamiltonicity, in a variety of random structures. See, e.g., any of [3, 4, 5, 19, 24] concerning Erdős-Rényi random graphs and random regular graphs, any of [6, 14, 20, 21] regarding directed graphs, or any of the recent developments [9, 11, 12, 13] on uniform hypergraphs. In this paper we consider the existence of rainbow Hamilton cycles in edge-colored graphs. (A set S of edges is called *rainbow* if each edge of S has a different color.) There are two general types of results in this area: existence **whp**¹ under random coloring and guaranteed existence under adversarial coloring.

When considering adversarial (worst-case) coloring, the guaranteed existence of a rainbow structure is called an *Anti-Ramsey* property. Erdős, Nešetřil, and Rödl [10], Hahn and Thomassen [17] and Albert, Frieze, and Reed [1] (correction in Rue [25]) considered colorings of the edges of the complete graph K_n where no color is used more than k times. It was shown in [1] that if $k \leq n/64$, then there must be a rainbow Hamilton cycle. Cooper and Frieze [7] proved a random graph threshold for this property to hold in almost every graph in the space studied.

There is also a history of work on random coloring (see, e.g., any of [7, 8, 15, 16]), and it has recently become apparent that this random setting may be of substantial utility. Indeed, a result of Janson and Wormald [16] on rainbow Hamilton cycles in randomly edge-colored random regular

*Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, email: alan@random.math.cmu.edu. Research supported in part by NSF award DMS-0753472.

†Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, PA 15213, e-mail: plosh@cmu.edu.

¹A sequence of events \mathcal{E}_n is said to occur *with high probability* (**whp**) if $\lim_{n \rightarrow \infty} \Pr \mathcal{E}_n = 1$.

graphs played a central role in the recent determination of the threshold for loose Hamiltonicity in random 3-uniform hypergraphs by Frieze [11]. Roughly speaking, a hyperedge (triple of vertices) can be encoded by an ordinary edge (pair of vertices), together with a color. Hence, a random 3-uniform hypergraph gives rise to a randomly edge-colored random graph. We will discuss this further in Section 4.

Let us now focus on the random coloring situation, where we consider the following model. Let $G_{n,p,\kappa}$ denote a randomly colored random graph, constructed on the vertex set $[n]$ by taking each edge independently with probability p , and then independently coloring it with a random color from the set $[\kappa]$. We are interested in conditions on n, p, κ which imply that $G_{n,p,\kappa}$ contains a rainbow Hamilton cycle **whp**. The starting point for our present work is the following theorem of Cooper and Frieze.

Theorem. (See [8], Theorem 1.1.) *There exist constants K_1 and K_2 such that if $p > \frac{K_1 \log n}{n}$ and $\kappa > K_2 n$, then $G_{n,p,\kappa}$ contains a rainbow Hamilton cycle **whp**.*

The aim of this paper is to substantially strengthen the above result by proving the following theorem:

Theorem 1.1. *If $p = \frac{(1+\epsilon) \log n}{n}$ and $\kappa = (1+\theta)n$, where $\epsilon, \theta > \frac{125}{\sqrt{\log \log n}}$, then $G_{n,p,\kappa}$ contains a rainbow Hamilton cycle **whp**.*

To discuss the tightness of our main theorem, let us recall the threshold for Hamiltonicity in $G_{n,p}$, established by Komlós and Szemerédi [19]. We find that we must have $p > \frac{\log n + \log \log n + \omega(n)}{n}$ with $\omega(n) \rightarrow \infty$, or else the underlying uncolored $G_{n,p}$ will not even be Hamiltonian. We also need at least n colors to appear on the edges in order to have enough colors for a rainbow Hamilton cycle. Note that the earlier result came within a constant factor of both of these minimum requirements.

Our theorem drives both constants down to be best possible up to first order. We allow our error terms ϵ and θ to decrease slowly, although we do not expect our constraints on them to be optimal. Our discussion above shows that the trivial lower bound for ϵ is around $\frac{\log \log n}{\log n}$. Then, if $p \sim \frac{\log n}{n}$, we need at least $n + \Omega(n^{1/2})$ colors just to ensure that **whp** at least n distinct colors occur on the $m \sim \frac{1}{2}n \log n$ edges in the graph; hence, the trivial lower bound for θ is around $\frac{1}{\sqrt{n}}$. We leave further exploration to future work, and highlight a potential answer in our conclusion.

This paper is organized as follows. Section 2 provides an outline of our proof. The proofs of the main steps follow in the section thereafter. We conclude in Section 5 with some remarks and open problems. The following (standard) asymptotic notation will be utilized extensively. For two functions $f(n)$ and $g(n)$, we write $f(n) = o(g(n))$, $g(n) = \omega(f(n))$, or $f(n) \ll g(n)$ if $\lim_{n \rightarrow \infty} f(n)/g(n) = 0$, and $f(n) = O(g(n))$ or $g(n) = \Omega(f(n))$ if there exists a constant M such that $|f(n)| \leq M|g(n)|$ for all sufficiently large n . We also write $f(n) = \Theta(g(n))$ if both $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$ are satisfied. All logarithms will be in base $e \approx 2.718$.

2 Proof of Theorem 1.1: high level description

In our proofs below we will assume that ϵ, θ are sufficiently small so that various inequalities are true. In addition we will assume that $\epsilon, \theta > \frac{125}{\sqrt{\log \log n}}$.

We will implicitly assume throughout (when convenient) that they are sufficiently small. Our proof proceeds in three phases, so our parameters come in threes. Let us arbitrarily partition the

$\kappa = (1 + \theta)n$ colors into three disjoint groups $C_1 \cup C_2 \cup C_3$, with sizes

$$|C_1| = \theta_1 n, \quad |C_2| = (1 + \theta_2)n, \quad |C_3| = \theta_3 n.$$

We will analyze the random edge generation in three stages, so we define the probabilities

$$p_1 = \frac{\epsilon_1 \log n}{2n}, \quad p_2 = \frac{(1 + \epsilon_2) \log n}{2n}, \quad p_3 = \frac{\epsilon_3 \log n}{2n}.$$

The ϵ 's and θ 's are defined by the relations

$$\epsilon_1 = \epsilon_2 = \epsilon_3 = \frac{\epsilon}{3}, \quad \theta_1 = \theta_3 = \min \left\{ \frac{\theta}{3}, \frac{\epsilon_2}{4} \right\}, \quad \theta_2 = \theta - \theta_1 - \theta_3. \quad (1)$$

(We would have taken $\theta_1 = \theta_2 = \theta_3 = \frac{\theta}{3}$, except that Lemma 3.9 requires $\theta_1 + \theta_3 \leq \frac{\epsilon_2}{2}$.)

2.1 Underlying digraph model

It is more convenient for our entire argument to work with directed graphs, as this will allow us to conserve independence. Recall that $D_{n,p}$ is the model where each of the $n(n-1)$ possible directed edges appears independently with probability p . We generate a random colored undirected graph via the following procedure. First, we independently generate three digraphs $D_1^\circ = D_{n,p_1}$, $D_2^\circ = D_{n,p_2}$, and $D_3^\circ = D_{n,p_3}$, and color all of the directed edges independently and uniformly at random from the full set of colors.

We next use the D_i° to construct a colored undirected graph G , by taking the undirected edge uv if and only if at least one of \vec{uv} or \vec{vu} appear among the D_i° . The colors of the undirected edges are inherited from the colors of the directed edges, in the priority order $D_1^\circ, D_2^\circ, D_3^\circ$. Specifically, if \vec{uv} or \vec{vu} appear already in D_1° , then uv takes the color used in D_1° even if \vec{uv} or \vec{vu} appear again in D_3° , say. In the event that both \vec{uv} and \vec{vu} appear in D_1° , the color of \vec{uv} is used for uv with probability $1/2$, and the color of \vec{vu} is used otherwise. Similarly, if neither of \vec{uv} nor \vec{vu} appear in D_1° , but \vec{uv} appears in both D_2° and D_3° , the color used in D_2° takes precedence. It is clear that the resulting colored graph G has the same distribution as $G_{n,p,\kappa}$, with

$$p = 1 - (1 - p_1)^2(1 - p_2)^2(1 - p_3)^2 = (1 + \epsilon + O(\epsilon^2)) \frac{\log n}{n}.$$

2.2 Partitioning by color

In each of our three phases, we will use one group of edges and one group of colors. Since each D_i° contains edges colored from the entire set $C_1 \cup C_2 \cup C_3$, for each i we define $D_i \subset D_i^\circ$ to be the spanning subgraph consisting of all directed edges whose color is in C_i .

Our final undirected graph is generated by superimposing directed graphs and disregarding the directions. Consequently, we do not need to honor the directions when building Hamilton cycles. To account for this, we define three corresponding colored undirected graphs G_1, G_2 , and G_3 . These will be edge-disjoint, respecting priority.

The first, G_1 , is constructed as follows. For each pair of vertices u, v with $\vec{uv} \in D_1$ but $\vec{vu} \notin D_1^\circ$, place uv in G_1 in the same color as \vec{uv} . If both \vec{uv} and \vec{vu} are in D_1 , we still place the edge uv in G_1 , but randomly select either the color of \vec{uv} or of \vec{vu} . However, if $\vec{uv} \in D_1$ but $\vec{vu} \in D_1^\circ \setminus D_1$,

then uv is only placed in G_1 with probability $1/2$; if it is placed, it inherits the color of \vec{uv} . Note that this construction precisely captures all undirected edges arising from D_1° , using colors in C_1 .

We are less careful with G_2 , as our argument can afford to discard all edges that arise from multiply covered pairs. Specifically, we place $uv \in G_2$ if and only if $\vec{uv} \in D_2 \setminus D_1^\circ$ and $\vec{vu} \notin D_1^\circ \cup D_2^\circ$. As the pair $\{u, v\}$ is now spanned by only one directed edge in D_2° , the undirected edge uv inherits that unique color. We define G_3 similarly, placing $uv \in G_3$ if and only if $\vec{uv} \in D_3 \setminus (D_1^\circ \cup D_2^\circ)$ and $\vec{vu} \notin D_1^\circ \cup D_2^\circ \cup D_3^\circ$.

In this way, we create three edge-disjoint graphs G_i . By our observations in the previous section, we may now focus on finding an (undirected) rainbow Hamilton cycle in $G_1 \cup G_2 \cup G_3$. Importantly, note that in terms of generating colored undirected edges, the digraph D_1° has higher ‘‘priority’’ than D_2° or D_3° . So, for example, the generation of G_1 is not affected by the presence or absence of edges from D_2° or D_3° .

2.3 Main steps

We generally prefer to work with G_i and D_i instead of D_i° because we are guaranteed that the edge colors lie in the corresponding C_i . This allows us to build rainbow segments in separate stages, without worrying that we use the same color twice. Let $d_i^+(v)$ denote the out-degree of v in D_i . We now define a set S of vertices that need special treatment. We first let

$$S_0 = S_{0,1} \cup S_{0,2} \cup S_{0,3},$$

where

$$S_{0,1} = \left\{ v : d_1^+(v) \leq \frac{\epsilon_1 \theta_1}{20} \log n \right\} \quad (2)$$

$$S_{0,2} = \left\{ v : d_2^+(v) \leq \frac{1}{20} \log n \right\} \quad (3)$$

$$S_{0,3} = \left\{ v : d_3^+(v) \leq \frac{\epsilon_3 \theta_3}{20} \log n \right\}. \quad (4)$$

Also, define $\gamma = \min \left\{ \frac{1}{4}, \frac{1}{4} \epsilon_1 \theta_1, \frac{1}{4} \epsilon_3 \theta_3 \right\}$, and note that the constraints on ϵ, θ in Theorem 1.1 imply the bound

$$\gamma > \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{12} \cdot \left(\frac{125}{\sqrt{\log \log n}} \right)^2 > \frac{1}{\log \log n}. \quad (5)$$

Lemma 2.1. *With probability $1 - o(n^{-1})$, the set S_0 satisfies $|S_0| \leq \frac{1}{3} n^{1-\gamma}$.*

The vertices in S_0 are delicate because they have low degree. We also need to deal with vertices having several neighbors in S_0 . For this, we define a sequence of sets S_0, S_1, \dots, S_t in the following way. Having chosen S_t , if there is still a vertex $v \notin S_t$ with at least 4 out-neighbors in S_t (in any of the graphs D_1, D_2 , or D_3), we let $S_{t+1} = S_t \cup \{v\}$ and continue. Otherwise we stop at some value $t = T$ and take $S = S_T$.

Lemma 2.2. *With probability $1 - o(n^{-1})$, the set S contains at most $n^{1-\gamma}$ vertices.*

To take care of the dangerous vertices in S , we find a collection of vertex disjoint paths Q_1, Q_2, \dots, Q_s , $s \leq |S|$ such that (i) each path uses undirected edges in G_2 , (ii) all colors which

appear on these edges are distinct, (iii) all interior vertices of the paths are vertices of S , (iv) every vertex of S appears in this way, and (v) the endpoints of the paths are not in S . Let us say that these paths *cover* S .

Lemma 2.3. *The graph G_2 contains a collection Q_1, Q_2, \dots, Q_s of paths that cover S **whp**.*

The next step of our proof uses a random greedy algorithm to find a rainbow path of length close to n , avoiding all of the previously constructed Q_i .

Lemma 2.4. *The graph G_2 contains a rainbow path P of length $n' = n - \frac{n}{\sqrt[3]{\log n}}$ **whp**. Furthermore, P is entirely disjoint from all of the Q_i , and all colors used in P and the Q_i are distinct and from C_2 .*

Let U be the vertices outside P . (Note that U contains all of the paths Q_i .) In order to link the vertices of U into P , we split P into short segments, and use the edges of G_3 to splice U into the system of segments. We will later use the edges of G_1 to link the segments back together into a rainbow Hamilton cycle, so care must be taken to conserve independence. The following lemma merges all vertices of U into the collection of segments, and prepares us for the final stage of the proof. Here, $d_1^+(v; A)$ denotes the number of D_1 -edges from v to a set A . Let

$$L = \max \left\{ 10 \cdot e^{\frac{40}{\epsilon_3 \theta_3}}, \frac{7}{\theta_1} \right\}, \quad (6)$$

and note that our conditions on ϵ, θ in Theorem 1.1, together with (1), imply that $\epsilon_3 \theta_3 > \frac{1}{3} \cdot \frac{1}{12} \cdot \left(\frac{125}{\sqrt{\log \log n}} \right)^2 > \frac{434}{\log \log n}$, so we have

$$L < \max \left\{ 10 \cdot e^{\frac{40}{434} \log \log n}, 7 \cdot \frac{12 \sqrt{\log \log n}}{100} \right\} < \sqrt[10]{\log n}. \quad (7)$$

Lemma 2.5. *With probability $1 - o(1)$, the entire vertex set can be partitioned into segments I_1, \dots, I_r , with $r = (1 - o(1)) \frac{n}{L}$, such that the edges which appear in the segments all use different colors from $C_2 \cup C_3$. The segment endpoints are further partitioned into $A \cup B$, with each segment having one endpoint in A and one in B , such that every $a \in A$ has $d_1^+(a; B) \geq \frac{\epsilon_1 \theta_1}{200L} \log n$, and every $b \in B$ has $d_1^+(b; A) \geq \frac{\epsilon_1 \theta_1}{200L} \log n$. All of the numeric values $d_1^+(a; B)$ and $d_1^+(b; A)$ have already been revealed, but the locations of the corresponding edges are still independent and uniform over B and A , respectively.*

The final step links together the segments I_1, \dots, I_r using distinctly-colored edges from G_1 . For this, we create an auxiliary colored directed graph Γ , which has one vertex w_k for each segment I_k . There is a directed edge $\overrightarrow{w_j w_k} \in \Gamma$ if there is an edge $e \in G_1$ between the B -endpoint of I_j and the A -endpoint of I_k ; it inherits the color of e . Since all colors of edges in Γ are from C_1 , it therefore suffices to find a rainbow Hamilton directed cycle in Γ . We will find this by connecting Γ with a well-studied random directed graph model.

Definition 2.6. *The d -in, d -out random directed graph model $D_{d\text{-in}, d\text{-out}}$ is defined as follows. Each vertex independently chooses d out-neighbors and d in-neighbors uniformly at random, and all resulting directed edges are placed in the graph. Due to independence, it is possible that a vertex u selects v as an out-neighbor, and v also selects u as an in-neighbor. In that case, instead of placing two repeated edges \overrightarrow{uv} , place only one.*

Instead of proving Hamiltonicity from scratch, we apply the following theorem of Cooper and Frieze.

Theorem 2.7. (See [6], Theorem 1.) *The random graph $D_{2\text{-in},2\text{-out}}$ contains a directed Hamilton cycle **whp**.*

This result does not take colors into account, however. Fortunately, in equation (6), we define L to be large enough to allow us to select a subset of G_1 -edges which is itself already rainbow. The analysis of this procedure is the heart of the proof of the final step.

Lemma 2.8. *The colored directed graph Γ contains a rainbow directed Hamilton cycle **whp**.*

Since each directed edge of Γ corresponds to an undirected G_1 -edge from a B -endpoint of a segment to an A -endpoint of another segment, a directed Hamilton cycle in Γ corresponds to a Hamilton cycle linking all of the segments together. Lemma 2.8 establishes that it is possible to choose these linking edges as a rainbow set from C_1 . The edges within the segments were themselves colored from $C_2 \cup C_3$, so the result is indeed a rainbow Hamilton cycle in the original graph, as desired.

3 Proofs of intermediate lemmas

In the remainder of this paper, we prove the lemmas stated in the previous section. Although the first lemma is fairly standard, we provide all details, and use the opportunity to formally state several other well-known results which we apply again later.

3.1 Proof of Lemma 2.1

Our first lemma controls the number of vertices whose degrees in D_i are too small. Recall from Section 2.1 that the D_i° are independently generated. Their edges are then independently colored, and the edges of D_i° which receive colors from C_i are collected into D_i . (Priorities only take effect when we form the G_i in Section 2.2.) Therefore, the out-degrees $d_i^+(v)$ of vertices v in D_i are distributed as

$$\begin{aligned} d_1^+(v) &\sim \text{Bin}\left(n-1, p_1 \cdot \frac{\theta_1}{1+\theta_1+\theta_2+\theta_3}\right) \geq \text{Bin}\left(0.99n, \frac{0.49\epsilon_1\theta_1 \log n}{n}\right) \\ d_2^+(v) &\sim \text{Bin}\left(n-1, p_2 \cdot \frac{1+\theta_2}{1+\theta_1+\theta_2+\theta_3}\right) \geq \text{Bin}\left(0.99n, \frac{0.49 \log n}{n}\right) \\ d_3^+(v) &\sim \text{Bin}\left(n-1, p_3 \cdot \frac{\theta_3}{1+\theta_1+\theta_2+\theta_3}\right) \geq \text{Bin}\left(0.99n, \frac{0.49\epsilon_3\theta_3 \log n}{n}\right). \end{aligned}$$

Here and in the remainder, we will say that $\pi_1 \geq \pi_2$ for distributions π_1, π_2 if we can find a coupling of random variables (X, Y) such that $X \geq Y$ and $X \sim \pi_1, Y \sim \pi_2$.

Thus the expected size of S_0 satisfies

$$\mathbb{E}[|S_0|] \leq n(\rho_1 + \rho_2 + \rho_3),$$

where

$$\begin{aligned}\rho_1 &= \mathbb{P} \left[\text{Bin} \left(0.99n, \frac{0.49\epsilon_1\theta_1 \log n}{n} \right) \leq \frac{\epsilon_1\theta_1 \log n}{20} \right] \\ \rho_2 &= \mathbb{P} \left[\text{Bin} \left(0.99n, \frac{0.49 \log n}{n} \right) \leq \frac{\log n}{20} \right] \\ \rho_3 &= \mathbb{P} \left[\text{Bin} \left(0.99n, \frac{0.49\epsilon_3\theta_3 \log n}{n} \right) \leq \frac{\epsilon_3\theta_3 \log n}{20} \right].\end{aligned}$$

We will repeatedly use the following case of the Chernoff lower tail bound, which we prove with an appropriate explicit constant.

Lemma 3.1. *The following holds for all sufficiently large mq , where m is a positive integer and $0 < q < 1$ is a real number.*

$$\mathbb{P} \left[\text{Bin} (m, q) \leq \frac{1}{9}mq \right] < e^{-0.533mq}.$$

Proof. Calculation yields

$$\mathbb{P} \left[\text{Bin} (m, q) \leq \frac{1}{9}mq \right] = \sum_{k=0}^{mq/9} \binom{m}{k} q^k (1-q)^{m-k} < \sum_{k=0}^{mq/9} \left(\frac{emq}{k} \right)^k e^{-\frac{8}{9}mq}$$

where $\left(\frac{emq}{k}\right)^k = 1$ by convention for $k = 0$. The function $\left(\frac{C}{k}\right)^k = \exp\{k(\log C - \log k)\}$ is increasing in k in the range $0 < k < C/e$. Thus

$$\begin{aligned}\mathbb{P} \left[\text{Bin} (m, q) \leq \frac{1}{9}mq \right] &< \frac{2mq}{9} \cdot \left(\frac{emq}{mq/9} \right)^{mq/9} e^{-\frac{8}{9}mq} \\ &= \frac{2mq}{9} \cdot (9e)^{mq/9} e^{-\frac{8}{9}mq} \\ &= e^{mq(\frac{1}{9} \log 9e - \frac{8}{9} + o(1))} \\ &< e^{-0.533mq},\end{aligned}$$

as claimed. □

Returning to the proof of Lemma 2.1, we observe that since $\frac{1}{20} < \frac{1}{9} \cdot 0.99 \cdot 0.49$, a direct application of Lemma 3.1 now gives

$$\begin{aligned}\rho_2 &< \mathbb{P} \left[\text{Bin} \left(0.99n, \frac{0.49 \log n}{n} \right) \leq \frac{1}{9} \cdot 0.99 \cdot 0.49 \log n \right] \\ &< e^{-0.533 \cdot 0.99 \cdot 0.49 \log n} \\ &< n^{-0.258}.\end{aligned}$$

A similar argument establishes that $\rho_1 < n^{-0.258\epsilon_1\theta_1}$ and $\rho_3 < n^{-0.258\epsilon_3\theta_3}$. This proves that $\mathbb{E}[|S_0|] = o(n^{1-\gamma})$, where we recall our definition $\gamma = \min\{\frac{1}{4}, \frac{1}{4}\epsilon_2\theta_2, \frac{1}{4}\epsilon_3\theta_3\}$. We complete the proof of the lemma by showing that $|S_0|$ is concentrated around its mean. For this, we use the Hoeffding-Azuma martingale tail inequality applied to the vertex exposure martingale (see, e.g., [2]). Recall that a martingale is a sequence X_0, X_1, \dots of random variables such that each conditional expectation $\mathbb{E}[X_{t+1} | X_0, \dots, X_t]$ is precisely X_t .

Theorem 3.2. *Let X_0, \dots, X_n be a martingale, with bounded differences $|X_{i+1} - X_i| \leq C$. Then for any $\lambda \geq 0$,*

$$\mathbb{P}[X_n \geq X_0 + \lambda] \leq \exp\left\{-\frac{\lambda^2}{2C^2n}\right\}.$$

Here we consider $|S_0|$ to be a function of Y_1, Y_2, \dots, Y_n where Y_k denotes the set of edges $\vec{jk}, \vec{kj} \in D_1^\circ \cup D_2^\circ \cup D_3^\circ$, $j < k$. The sequence $X_t = \mathbb{E}[|S_0| \mid Y_1, \dots, Y_t]$ is called the vertex-exposure martingale. There is a slight problem in that the worst-case Lipschitz value for changing a single Y_k can be too large, while the average case is good. There are various ways of dealing with this. We will make a small change in $D^\circ = D_1^\circ \cup D_2^\circ \cup D_3^\circ$. Let \hat{D}° be obtained from D° by reducing every degree below $5 \log n$. We do this in vertex order $v = 1, 2, \dots, n$ and delete edges incident with v in descending numerical order. We can show that this usually has no effect on D° .

Lemma 3.3. *With probability $1 - o(n^{-1})$, every vertex in $G_{n,p}$ with $p < \frac{1.1 \log n}{n}$ has degree at most $5 \log n$.*

Proof. The probability that a single vertex has degree at least $5 \log n$ is

$$\begin{aligned} \mathbb{P}[\text{Bin}(n-1, p) \geq 5 \log n] &\leq \binom{n}{5 \log n} \left(\frac{1.1 \log n}{n}\right)^{5 \log n} \\ &\leq \left(\frac{en}{5 \log n} \cdot \frac{1.1 \log n}{n}\right)^{5 \log n} \\ &= \left(\frac{1.1e}{5}\right)^{5 \log n} \\ &= n^{-2.57}, \end{aligned}$$

so a union bound over all vertices gives the result. \square

Therefore, $\mathbb{P}[\hat{D}^\circ = D^\circ] = 1 - o(n^{-1})$, and so if we let $\hat{Z} = |\hat{S}_0|$ be the size of the corresponding set evaluated in \hat{D}° , we obtain $\mathbb{E}[\hat{Z}] = \mathbb{E}[|S_0|] + o(1) = o(n^{1-\gamma})$. Furthermore, changing a Y_k can only change \hat{Z} by at most $15 \log n$. So, we have

$$\mathbb{P}\left[\hat{Z}_i \geq \mathbb{E}[\hat{Z}_i] + \frac{1}{4}n^{1-\gamma}\right] \leq \exp\left\{-\frac{n^{2-2\gamma}/16}{2(15 \log n)^2n}\right\} < o(n^{-1}),$$

completing the proof of Lemma 2.1. \square

3.2 Proof of Lemma 2.2

We use the following standard estimate to control the densities of small sets.

Lemma 3.4. *With probability $1 - o(n^{-1})$, in $D_{n,p}$ with $p < \frac{\log n}{n}$, every set S of fewer than $\frac{4}{e^4} \cdot \frac{n}{\log^2 n}$ vertices satisfies $e(S) < 2|S|$. Here, $e(S)$ is the number of directed edges spanned by S .*

Proof. Fix a positive integer $s < \frac{4}{e^4} \cdot \frac{n}{\log^2 n}$, and consider sets of size s . We may assume that $s \geq 2$, because a single vertex cannot induce any edges. The expected number of sets S with $|S| = s$ and

$e(S) \geq 2s$ is at most

$$\begin{aligned} \binom{n}{s} \cdot \binom{s^2}{2s} \left(\frac{\log n}{n}\right)^{2s} &\leq \left(\frac{en}{s}\right)^s \cdot \left(\frac{es^2}{2s}\right)^{2s} \left(\frac{\log n}{n}\right)^{2s} \\ &= \left(s \cdot \frac{e^3 \log^2 n}{4n}\right)^s. \end{aligned}$$

It remains to show that when this bound is summed over all $2 \leq s < \frac{4}{e^4} \cdot \frac{n}{\log^2 n}$, the result is still $o(n^{-1})$. Indeed, for each $2 \leq s \leq 2 \log n$, the bound is at most $O\left(\frac{\log^6 n}{n^2}\right)$, so the total contribution from that part is only $O\left(\frac{\log^7 n}{n^2}\right) = o(n^{-1})$. On the other hand, for each $2 \log n < s < \frac{4}{e^4} \cdot \frac{n}{\log^2 n}$, the bound is at most

$$\left(\frac{4}{e^4} \cdot \frac{n}{\log^2 n} \cdot \frac{e^3 \log^2 n}{4n}\right)^{2 \log n} = \left(\frac{1}{e}\right)^{2 \log n} = n^{-2}.$$

Thus the total contribution from $2 \log n < s < \frac{4}{e^4} \cdot \frac{n}{\log^2 n}$ is at most $o(n^{-1})$, as desired. \square

We are now ready to bound the size of the set S which was created by repeatedly absorbing vertices with many neighbors in S_0 .

Proof of Lemma 2.2. We actually prove a stronger statement, which we will need for Lemma 3.7. Suppose we have an initial S_0 satisfying $|S_0| < \frac{1}{3}n^{1-\gamma}$, as ensured by Lemma 2.1. Consider a sequence S'_0, S'_1, S'_2, \dots where S'_{t+1} is obtained from S'_t by adding a vertex $v \notin S'_t$ for which $d_i^+(v; S'_t) \geq 3$ for some i . Note that when this process stops, the final set S' will contain the set S which by our definition is obtained by adding vertices with degree at least 4 into previous S_t .

So, suppose that this process continues for so long that some $|S'_t|$ reaches $n^{1-\gamma} = o\left(\frac{n}{\log^2 n}\right)$. Note that $t \geq \frac{2}{3}n^{1-\gamma}$. Since each step introduces at least 3 edges, we must have $e(S'_t) \geq 3t \geq 2n^{1-\gamma} = 2|S'_t|$. Lemma 3.4 implies that can only happen with probability $o(n^{-1})$. (By construction, $D_1 \cup D_2 \cup D_3$ is an instance of $D_{n,q}$ for some $q < \frac{\log n}{n}$.) \square

3.3 Proof of Lemma 2.3

In this section, we show that for each vertex $v \in S$, we can find a disjoint G_2 -path Q containing v which starts and ends outside S . We also need all colors appearing on these edges to be different. Since we are working in a regime where degrees can be very small, we need to accommodate the most delicate vertices first. Specifically, let $S_{0,0}$ be the set of all vertices with $d_2(v) \leq \frac{1}{10} \log n$, where $d_2(v)$ is the degree of v in G_2 . Although $S_{0,0}$ will typically not be entirely contained within $S_{0,2}$, we can show that it is usually quite small.

Lemma 3.5. *We have $|S_{0,0}| < n^{0.48}$ whp.*

Proof. By construction, $G_2 \sim G_{n,q_2}$, where

$$q_2 = 2p_2(1-p_2) \cdot \frac{1+\theta_2}{1+\theta_1+\theta_2+\theta_3} \cdot (1-p_1)^2, \quad (8)$$

because the first factor is the probability that exactly one of \vec{uv} or \vec{vu} appears in D_2^o , the second factor is the probability that it receives a color from C_2 , and the third factor is the probability that

neither \vec{uv} nor \vec{vu} appear in D_1° . Hence for a fixed vertex v , its relevant degree in G_2 is distributed as

$$d_2(v) \sim \text{Bin}(n-1, q_2) \geq \text{Bin}\left(0.99n, \frac{0.99 \log n}{n}\right).$$

Since $\frac{1}{10} < \frac{1}{9} \cdot 0.99 \cdot 0.99$, Lemma 3.1 implies that

$$\begin{aligned} \mathbb{P}\left[d_2(v) \leq \frac{1}{10} \log n\right] &< \mathbb{P}\left[\text{Bin}\left(0.99n, \frac{0.99 \log n}{n}\right) \leq \frac{1}{9} \cdot 0.99 \cdot 0.99 \log n\right] \\ &< e^{-0.533 \cdot 0.99 \cdot 0.99 \log n} \\ &< n^{-0.522}. \end{aligned} \tag{9}$$

Therefore, $\mathbb{E}[|S_{0,0}|] < n \cdot n^{-0.522}$, and Markov's inequality yields the desired result. \square

We have shown that vertices of $S_{0,0}$ are few in number. Our next result shows that they are also scattered far apart. This will help us when we construct the covering paths, by preventing paths from colliding.

Lemma 3.6. *Let $\text{dist}_2(v, w)$ denote the distance between v and w in G_2 . Then, **whp**, every pair $v, w \in S_{0,0}$ satisfies $\text{dist}_2(v, w) \geq 5$.*

Proof. Recall that $G_2 \sim G_{n, q_2}$ with q_2 defined as in (8). Consider a fixed pair of vertices v, w . For a fixed sequence of $k \leq 4$ intermediate vertices x_1, x_2, \dots, x_k , let us bound the probability q that v, w both have $d_2 \leq \frac{1}{10} \log n$, and all the edges $vx_1, x_1x_2, x_2x_3, \dots, x_kw$ appear in G_2 . First expose the edges $vx_1, x_1x_2, \dots, x_kw$, and then expose the edges between v and $[n] \setminus \{v, x_1, \dots, x_k, w\}$, and between w and that set.

This gives the following bound on our probability q :

$$q \leq \left(\frac{1.01 \log n}{n}\right)^{k+1} \cdot \mathbb{P}\left[\text{Bin}\left(n-2-k, \frac{0.99 \log n}{n}\right) \leq \frac{\log n}{10}\right]^2 \tag{10}$$

A calculation analogous to (9) bounds the Binomial probability by $n^{-0.52}$, so taking a union bound over all $O(n^k)$ choices for the x_i , for all $0 \leq k \leq 4$, we find that for fixed v, w , the probability that $\text{dist}_2(v, w) < 5$ is at most

$$\sum_{k=0}^4 O(n^k) \cdot \left(\frac{1.01 \log n}{n}\right)^{k+1} \cdot (n^{-0.52})^2 < n^{-2.04+o(1)}.$$

Therefore, a final union bound over the $O(n^2)$ choices for v, w completes the proof. \square

The previous result will help us cover vertices in $S_{0,0}$ with G_2 -paths. However, the objective of this section is to cover all vertices of S . Although the analogue of Lemma 3.6 does not hold for S , it is still possible to prove that S is sparsely connected to the rest of the graph. Recall from (5) that $\gamma = \min\left\{\frac{1}{4}, \frac{1}{4}\epsilon_1\theta_1, \frac{1}{4}\epsilon_3\theta_3\right\} > \frac{1}{\log \log n}$.

Lemma 3.7. *With respect to edges of G_2 , every vertex v is adjacent to at most $\frac{2}{\gamma}$ vertices in S **whp**. (This applies whether or not v itself is in S .)*

Proof. Fix a vertex v . Let S' be the set obtained by constructing the analogous sequences to $S'_{0,1}, S'_{0,2}, S'_{0,3}, S'_1, \dots$ on the graph induced by $[n] \setminus \{v\}$, where $S'_{0,i}, i = 1, 2, 3$ is defined as in (2)–(4), but the S'_{t+1} are obtained by adding vertices with at least 3 (not 4) D_i -out-neighbors in S'_t . Clearly, S' contains $S \setminus \{v\}$, because the effect of ignoring v is compensated for by using 3 instead of 4. The advantage of using S' instead of S is that S' can be generated without exposing any edges incident to v . As we will take a final union bound over the n choices of v , it therefore suffices to show that with probability $1 - o(n^{-1})$, the particular vertex v has at most $\frac{2}{\gamma}$ neighbors in S' .

For this, we expose all edges of $D_1^\circ \cup D_2^\circ \cup D_3^\circ$ that are spanned by $[n] \setminus \{v\}$. Recall that our proof of Lemma 2.2 already absorbed vertices with 3 out-neighbors (instead of 4), so we have $|S'| \leq n^{1-\gamma}$ with probability $1 - o(n^{-1})$. It remains to control the number of edges between v and S' , so we now expose all edges of $D_1^\circ \cup D_2^\circ \cup D_3^\circ$. The G_2 -edges there appear independently with probability q_2 as defined in (8), so the probability that at least $\frac{2}{\gamma}$ edges appear is at most

$$\begin{aligned} \mathbb{P} \left[\text{Bin} \left(n^{1-\gamma}, q_2 \right) \geq \frac{2}{\gamma} \right] &\leq \binom{n^{1-\gamma}}{2/\gamma} \left(\frac{1.01 \log n}{n} \right)^{2/\gamma} \\ &\leq \left(\frac{e \cdot n^{1-\gamma}}{2/\gamma} \cdot \frac{1.01 \log n}{n} \right)^{2/\gamma} \\ &< \left(\frac{2\gamma \log n}{n^\gamma} \right)^{2/\gamma} \\ &< \frac{(2 \log n)^{2 \log \log n}}{n^2} = o(n^{-1}). \end{aligned} \tag{11}$$

Taking a final union bound over all initial choices for v completes the proof. \square

We will cover each vertex $v \in S$ with a G_2 -path by joining two G_2 -paths of length up to 2, each originating from v . It is therefore convenient to extend the previous result by one further iteration.

Corollary 3.8. *With respect to edges of G_2 , every vertex v is within distance two of at most $\left(\frac{2}{\gamma}\right)^2$ vertices in S **whp**. (This applies whether or not v itself is in S .)*

Proof. Fix a vertex v . Construct S' in the same way as in the proof of Lemma 3.7, exposing only edges spanned by $[n] \setminus \{v\}$. Using only those exposed edges, let $T \subset [n] \setminus \{v\}$ be the set of all vertices in S' or adjacent to S' via edges from G_2 . By Lemma 3.3, the maximum degree of $G_2 \setminus \{v\}$ is at most $5 \log n$ with probability $1 - o(n^{-1})$, so $|T| \leq n^{1-\gamma} \cdot 5 \log n$. A similar calculation to (11) then shows that with probability $1 - o(n^{-1})$, v has at most $\frac{2}{\gamma}$ neighbors in T .

Taking a union bound over all v , and combining this with Lemma 3.7, we conclude that **whp**, every vertex has at most $\frac{2}{\gamma}$ neighbors in $S \cup N(S)$, and each of them has at most $\frac{2}{\gamma}$ neighbors in S . This implies the result. \square

We are now ready to start covering the vertices of S with disjoint rainbow G_2 -paths. The most delicate vertices are those in $S_{0,0}$, because by definition all other vertices already have G_2 -degree at least $\frac{1}{10} \log n$. Naturally, we take care of $S_{0,0}$ first.

Lemma 3.9. *The colored graph G_2 contains a rainbow collection $Q_1, Q_2, \dots, Q_{s'}$ of disjoint paths that cover $S \cap S_{0,0}$ **whp**.*

Proof. We condition on the high-probability events in Lemmas 3.5, 3.6, and 3.7, and use a greedy algorithm to cover each $v \in S \cap S_{0,0}$ with a path of length 2, 3, or 4. Recall that $G_2 \sim G_{n,q_2}$, where q_2 was specified in (8). We can bound q_2 by

$$q_2 > (1 + \epsilon_2) \frac{\log n}{n} \left(1 - \frac{\log n}{n}\right) \cdot (1 - \theta_1 - \theta_3) \cdot \left(1 - \frac{\log n}{n}\right)^2.$$

Since equation (1) ensures that $\theta_1 + \theta_3 \leq \frac{\epsilon_2}{2}$, and our conditions on Theorem 1.1 force $\epsilon_2 > \frac{1}{3} \cdot \frac{125}{\sqrt{\log \log n}} \gg \frac{\log \log n}{\log n}$, the minimum degree in G_2 is at least two **whp**, see e.g. [3]. Condition on this as well.

Now consider a vertex $v \in S \cap S_{0,0}$, and let x_1 and x_2 be two of its neighbors. If both x_i are already outside S , then we use $x_1 v x_2$ to cover v . Otherwise, suppose that x_1 is still in S . Since we conditioned on vertices in $S_{0,0}$ being separated by distances of at least 5 (Lemma 3.6), x_1 cannot be in $S_{0,0}$, so it has at least $\frac{1}{10} \log n$ G_2 -neighbors. These cannot all be in S , because we conditioned on the fact that every vertex has fewer than $\frac{2}{\gamma} < 2 \log \log n$ neighbors in S (Lemma 3.7). So, we can pick one, say y_1 , such that $y_1 x_1 v$ is a path from outside S to v . A similar argument allows us to continue the path from v to a vertex outside S in at most two steps. Therefore, there is a collection of paths of length 2–4 covering each vertex in $S \cap S_{0,0}$. They are all disjoint, since we conditioned on vertices of $S_{0,0}$ being separated by distances of at least 5.

At this point, we have exposed all G_2 -edges spanned by S and its neighbors, but the only thing we have revealed about their colors is that they are all in C_2 . Now expose the precise colors on all edges of these paths. Since we conditioned on $|S_{0,0}| < n^{0.48}$ (Lemma 3.5), the total number of edges involved is at most $4 \cdot n^{0.48} < n^{0.49}$. The number of colors in C_2 is $(1 + \theta_2)n$, so by a simple union bound the probability that some pair of edges receives the same color in C_2 is at most

$$\binom{n^{0.49}}{2} \frac{1}{(1 + \theta_2)n} = o(1).$$

Therefore, the covering paths form a rainbow set **whp**, as desired. \square

We have now covered the most dangerous vertices of S . The remainder of this section provides an argument which covers all other vertices in S .

Proof of Lemma 2.3. Condition on the high-probability events of Lemmas 2.2, 3.7, 3.9, and Corollary 3.8. We have already covered all vertices in $S \cap S_{0,0}$ with disjoint rainbow paths of lengths up to four (Lemma 3.9). We cover the rest of the vertices in $S \setminus S_{0,0}$ with paths of length two, using a simple iterative greedy algorithm. Indeed, suppose that we are to cover a given vertex $v \in S \setminus S_{0,0}$. Since it is not in $S_{0,0}$, it has G_2 -degree at least $\frac{1}{10} \log n$, and at most $\frac{2}{\gamma}$ of these neighbors can be within S (Lemma 3.7).

Furthermore, we can show that at most $2\left(\frac{2}{\gamma}\right)^2$ of v 's neighbors outside S can already have been used by covering paths. Indeed, for each neighbor $w \notin S$ of v that was used by a previous covering path, we could identify a vertex $x \in S$ adjacent to w which was part of that covering path. Importantly, x is within distance two of v , so the collection of all x obtainable in this way is of size at most $\left(\frac{2}{\gamma}\right)^2$, as we conditioned on Corollary 3.8. Since every covering path uses exactly two vertices outside S , the total number of such w is at most $2\left(\frac{2}{\gamma}\right)^2$. Putting everything together, we conclude that the number of usable G_2 -edges emanating from v is at least

$$\frac{1}{10} \log n - \frac{2}{\gamma} - 2 \left(\frac{2}{\gamma}\right)^2 > \frac{1}{11} \log n.$$

Expose the colors (necessarily from C_2) which appear on these G_2 -edges. Of the total of $(1 + \theta_2)n$ available, we only need to avoid at most $4|S|$ which have already been used on previous covering paths. Since we conditioned on $|S| \leq n^{1-\gamma}$ (Lemma 2.2), this is at most $4n^{1-\gamma}$ colors to avoid. We only need to have two new colors to appear among this collection in order to add a new rainbow path of length two covering v . Taking another union bound, we find that the probability that at most one new color appears is at most

$$(1 + \theta_2)n \cdot \left(\frac{4n^{1-\gamma} + 1}{(1 + \theta_2)n} \right)^{\frac{1}{11} \log n} = o(n^{-1}).$$

Here, the first factor of $(1 + \theta_2)n$ corresponds to the number of ways to choose the new color to add (or none at all). Since we only run our algorithm for $o(n)$ iterations (once per vertex in $S \setminus S_{0,0}$), we conclude that **whp** we can cover all vertices of S with disjoint rainbow G_2 -paths. \square

3.4 Proof of Lemma 2.4

In this section, we construct a rainbow G_2 -path which contains most of the vertices of the graph, but avoids all covering paths from the previous section. In order to carefully track the independence and exposure of edges, recall from Section 2.2 that G_2 is deterministically constructed from the random directed graphs D_1° , D_2° , and D_3° . Let us consider the generation of the D_i° to be as follows. The probability that the directed edge \vec{vw} appears in D_1 is $p_1 \cdot \frac{\theta_1}{1+\theta_1+\theta_2+\theta_3}$, so we expose each D_1 -out-degree $d_1^+(v)$ by independently sampling from the $\text{Bin}\left(n-1, p_1 \cdot \frac{\theta_1}{1+\theta_1+\theta_2+\theta_3}\right)$ distribution. Importantly, we do not reveal the locations of the out-neighbors. Similarly, for D_2 and D_3 , we expose all out-degrees $d_2^+(v)$ and $d_3^+(v)$, each sampled from the appropriate Binomial distribution. By Lemma 3.3, all $d_i^+(v) \leq 5 \log n$ **whp**; we condition on this.

Note that from this information, we can later fully generate (say) D_1 and D_1° as follows. At each vertex v , we independently choose $d_1^+(v)$ out-neighbors uniformly at random. This will determine all D_1 -edges. Next, for every edge which is not part of D_1 , independently sample it to be part of $D_1^\circ \setminus D_1$ with probability $p_1 \left(1 - \frac{\theta_1}{1+\theta_1+\theta_2+\theta_3}\right)$. This will determine all D_1° edges, and a similar system will determine all edges of D_2° and D_3° .

Returning to the situation where only the $d_i^+(v)$ have been exposed, we then construct the $S_{0,i}$ by collecting all vertices whose $d_i^+(v)$ are too small, and build the sequence $S_0, S_1, S_2, \dots, S_t$. In each iteration of that process, we go over all vertices which are not yet in the current S_t . At each v , we expose all D_i -edges incident to S_t . For this section, we will only care about the D_2 -out-edges from $v \notin S$ (initially counted by $d_2^+(v)$) that are *not* consumed in this process. Fortunately, at each exposure stage, there is a clear distribution on the number of these out-edges that are consumed toward S_t , and this will only affect the number, not the location, of the out-edges which are not consumed.

Therefore, after this procedure terminates, we will have a final set S , and the set of revealed (directed) edges is precisely those edges spanned by S , together with all those between S and $V_1 = [n] \setminus S$. This set of revealed edges is exactly what is required to construct the covering paths Q_1, \dots, Q_s in Lemma 2.3. Within V_1 , the precise locations of the edges are not yet revealed. Instead, for each vertex $v \in V_1$, there is now a number $d_2^*(v)$, corresponding to the number of D_2 -out-edges from v to vertices outside S .

We now make two crucial observations. First, the distributions of where these endpoints lie are still independent and uniform over V_1 . Second, every $d_2^*(v) \leq 5 \log n$ and $d_2^*(v) \geq \frac{1}{20} \log n - 3 > \frac{1}{21} \log n$, because if there were 4 out-edges from v to S , then v should have been absorbed into S during the process.

This abundance of independence makes it easy to analyze a simple method for finding a long path, based on a greedy algorithm with backtracking. (This procedure is similar to that used in [23] by Fernandez de la Vega.) Indeed, the most straightforward attempt would be to start building a path, and at each iteration expose the out-edges of the final endpoint, as well as their colors. If there is an option which keeps the path rainbow, we would follow that edge, and repeat. If not, then we should backtrack to the latest vertex in the path which still has an option for extension.

We formalize this in the following algorithm. Particularly dangerous vertices will be coded by the color red (not related to the colors of the edges in the $G_{n,p,\kappa}$). Let $V_2 \subset V_1$ be the set of all vertices which are not involved in the covering paths Q_i . We will find a long G_2 -path within V_2 which avoids all of the covering paths.

Algorithm.

1. Initially, let all vertices of V_2 be uncolored, and select an arbitrary vertex $v \in V_2$ to use as the initial path $P_0 = \{v\}$. Let $U_0 = V_2 \setminus \{v\}$. This is the set of “untouched” vertices. Let $R_0 = \emptyset$. This will count the “red” vertices.
2. Now suppose we are at time t . If $|U_t| < \frac{n}{2\sqrt[3]{\log n}}$, terminate the algorithm.
3. If the final endpoint v of P_t is not red, then expose the first $\frac{1}{2}d_2^*(v)$ of v 's D_2 -out-neighbors. If none of them lies in U_t , via an edge color not yet used by P_t or any of the covering paths Q_i , then color v red, setting $U_{t+1} = U_t$, $P_{t+1} = P_t$, and $R_{t+1} = R_t \cup \{v\}$.
 Otherwise, arbitrarily choose one of the suitable out-neighbors $w \in U_t$. Set $U_{t+1} = U_t \setminus \{w\}$. Expose whether $\overrightarrow{vw} \in D_1^\circ$, $\overleftarrow{vw} \in D_1^\circ$, or $\overleftarrow{vw} \in D_2^\circ$. If none of those three directed edges are present, then add w to the path, setting $P_{t+1} = P_t \cup \{w\}$ and $R_{t+1} = R_t$. Otherwise, color both v and w red, and set $P_{t+1} = P_t$ and $R_{t+1} = R_t \cup \{v, w\}$.

4. If the final endpoint v of P_t is red, then expose the second $\frac{1}{2}d_2^*(v)$ of v 's D_2 -out-neighbors. First suppose that none of them lies in U_t , via an edge color not yet used by P_t or any of the covering paths Q_i . In this case, find the last vertex v' of P_t which is not red, color it red, and make it the new terminus of the path. That is, set $U_{t+1} = U_t$, let P_{t+1} be P_t up to v' , and set $R_{t+1} = R_t \cup \{v'\}$. If v' did not exist (i.e., all vertices of P_t were already red), then instead let v' be an arbitrary vertex of U_t and restart the path, setting $P_{t+1} = \{v'\}$, $R_{t+1} = R_t$, $U_{t+1} = U_t \setminus \{v'\}$.

On the other hand, if v has a suitable out-neighbor $w \in U_t$, then set $U_{t+1} = U_t \setminus \{w\}$. Expose whether $\overrightarrow{vw} \in D_1^\circ$, $\overleftarrow{vw} \in D_1^\circ$, or $\overleftarrow{vw} \in D_2^\circ$. If none of those three directed edges are present, then add w to the path, setting $P_{t+1} = P_t \cup \{w\}$ and $R_{t+1} = R_t$. Otherwise, color w red, find the last vertex v' of P_t which is not red, and follow the remainder of the first paragraph of this step.

The key observation is that the final path P_T contains every non-red vertex which lies in $V_2 \setminus U_T$.

Since Lemmas 2.2 and 2.3 imply that $|V_2| \geq n - 3|S| \geq n - 3n^{1-\gamma}$, and we run until $U_T < \frac{n}{2\sqrt[3]{\log n}}$, Lemma 2.4 therefore follows from the following bound and the fact that $n^\gamma > n^{1/\log \log n} \gg \sqrt[3]{\log n}$.

Lemma 3.10. *The final number of red vertices is at most $n \cdot e^{-\frac{1}{300}\sqrt[3]{\log n}}$ **whp**.*

Proof. The color red is applied in only two situations. The first is when we expose whether any of $\vec{vw} \in D_1^\circ$, $\vec{wv} \in D_1^\circ$, or $\vec{vw} \in D_2^\circ$ hold. To expose whether $\vec{vw} \in D_1^\circ$, we reveal whether $\vec{vw} \in D_1$, using the previously exposed value of $d_1^*(v)$, which we already conditioned on being at most $5 \log n$. Since v 's D_1 -out-neighbors are uniform, the probability that $\vec{vw} \in D_1$ is at most $\frac{5 \log n}{(1-o(1))n}$. If it is not in D_1 , the probability that it is in $D_1^\circ \setminus D_1$ is bounded by $\frac{\log n}{n}$ by the description at the beginning of Section 3.4. The analysis for the other two cases are similar, so a union bound gives that the chance that any of $\vec{vw} \in D_1^\circ$, $\vec{wv} \in D_1^\circ$, or $\vec{vw} \in D_2^\circ$ hold is at most $3 \cdot \frac{7 \log n}{n}$. Note that this occurs at most n times, because each instance reduces the size of U_t by 1. Hence the expected number of red vertices of this type is at most $O(\log n)$, which is of much smaller order than $ne^{-\Theta(\sqrt[3]{\log n})}$.

The other situation in which red is applied comes immediately after the failed exposure of some $k = \frac{1}{2}d_2^*(v) > \frac{1}{42} \log n$ D_2 -out-neighbors, in either of Steps 3 or 4. Failure means that all k of them either fell outside U_t , or had edge colors already used in P_t or some covering path Q_i . Step 2 controls $|U_t| \geq \frac{n}{2\sqrt[3]{\log n}}$, and the total number of colors used in P_t or any covering path Q_i is at most $n - |U_t|$, out of the $(1 + \theta_2)n$ available. Further note that because of our order of exposure, there is a set T of size at most $3 \log n$ such that v 's D_2 -out-neighbors are uniformly distributed over $V_1 \setminus T$. This is because we have exposed whether \vec{vu} was a D_2 -edge, for the predecessor u of v along P_t , which eliminates one vertex, and we may also have already exposed the first half of v 's D_2 -out-neighbors in a prior round, which could consume up to $\frac{1}{2} \cdot 5 \log n$ vertices. Therefore, the chance that a given out-neighbor exposure is successful (i.e., lands inside U_t , via one of the $\geq |U_t|$ unused colors), is at least

$$\frac{|U_t \setminus T|}{|V_1 \setminus T|} \cdot \frac{|U_t|}{(1 + \theta_2)n} \geq \left(\frac{n}{2\sqrt[3]{\log n}} \cdot \frac{1}{(1 - o(1))n} \right) \cdot \left(\frac{n}{2\sqrt[3]{\log n}} \cdot \frac{1}{(1 + \theta_2)n} \right) > \frac{1}{5(\log n)^{2/3}}.$$

We conclude that the chance that all $k \geq \frac{1}{42} \log n$ fail is at most

$$(1 + o(1)) \left(1 - \frac{1}{5(\log n)^{2/3}} \right)^{\frac{1}{42} \log n} < e^{-\frac{1}{210}\sqrt[3]{\log n}}.$$

Since we will not perform this experiment more than twice for each of the n vertices, linearity of expectation and Markov's inequality imply that **whp**, the final total number of red vertices is at most $n \cdot e^{-\frac{1}{300}\sqrt[3]{\log n}}$, as desired. \square

3.5 Proof of Lemma 2.5

At this point, we have a rainbow G_2 -path P of length $n' \geq n - \frac{n}{\sqrt[3]{\log n}}$, which is disjoint from the paths Q_i which cover S . Recall from (6) and (7) that we defined $L = \max \{10e^{40/(\epsilon_3\theta_3)}, \frac{7}{\theta_1}\} < \sqrt[10]{\log n}$. Split P into $r = \frac{n'}{L}$ segments of length L , as in Figure 1. If n' is not divisible by L , we may discard the remainder of P , because $L < \sqrt[10]{\log n}$.

Partition the $2r$ endpoints into two sets $A_1 \cup B_1$ so that each segment has one endpoint in each set, but there are no vertices $a \in A_1$ and $b \in B_1$ which are consecutive along P . By possibly discarding the final segment (which will only cost an additional $L < \sqrt[10]{\log n}$), we may ensure that the initial and final endpoints are both in A_1 .

The reason for our unusual partition is as follows. In our construction thus far, we already needed to expose the locations of some D_1° -edges, since they had priority over the D_2° . In certain locations, we have revealed that there are no D_1° -edges. In particular, between every consecutive pair of vertices u, v on the path P , we found a D_2 -edge, and confirmed the absence of any D_1° -edges.

Fortunately, our construction did not expose any D_1° -edges between non-consecutive vertices of the path P . In particular, if we now wished, for any vertex $a \in A_1$, we could expose the number N of its D_1 -out-neighbors that lie in B_1 ; then, the distribution of these N out-neighbors would be uniform over B_1 . This uniformity is crucial, and would not hold, for example, if some vertex of B_1 were consecutive with a along P .

The proof of Lemma 2.5 breaks into the following steps. Recall that $d_1^+(v; T)$ denotes the number of D_1 -edges from a vertex v to a subset T of vertices. Say that v is T -good if $d_1^+(v; T) \geq \frac{\epsilon_1 \theta_1}{181L} \log n$; call it T -bad otherwise.

Step 1. For every vertex $v \in P \setminus (A_1 \cup B_1)$, as well as for the initial and final endpoints of P , expose the value of $d_1^+(v; B_1)$. We show that **whp**, the initial and final endpoints of P are B_1 -good, and at most $n \cdot e^{-\sqrt{\log n}}$ vertices of $P \setminus (A_1 \cup B_1)$ are B_1 -bad.

Step 2. We will describe below a procedure that will allow us to absorb all remaining vertices and covering paths into a new system of segments, using G_3 -edges that are aligned with B_1 -good vertices. (See Figure 2.) This removes some segment endpoints, while adding other new endpoints. Let $A_2 \cup B_2$ be the new partition of endpoints. Crucially, $B_2 = B_1$, while $|A_2| = |A_1|$ by losing up to $\frac{2n}{\sqrt[3]{\log n}}$ vertices, and then adding back the same number. Importantly, every new vertex in $A_2 \setminus A_1$ will be B_2 -good.

Step 3. We will then show that the system of segments can be grouped into several blocks of consecutive segments, in the sense that between successive segments in the same block, there is an original edge of P . (See Figure 3.) Also, the initial and final endpoints of each block are always of type A , and are all B_2 -good, and each block will contain at least $\frac{\sqrt[3]{\log n}}{L^2}$ segments.

Step 4. For every vertex $a \in A_2$, expose the value of $d_1^+(a; B_2)$, and for every vertex $b \in B_2$, expose the value of $d_1^+(b; A_2)$. We show that **whp**, at most $n \cdot e^{-\sqrt{\log n}}$ vertices of B_2 are A_2 -bad, and every block contains a string of four consecutive segments, each of whose A_2 -endpoints and B_2 -endpoints are B_2 -good and A_2 -good, respectively.

Step 5. For each consecutive pair of segments along the same block (from Step 3) which has either an A_2 -endpoint which is B_2 -bad or a B_2 -endpoint which is A_2 -bad, merge them, together

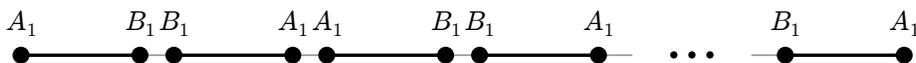


Figure 1: The long path P , divided into consecutive segments of length L . Endpoints of successive segments are adjacent via original edges of P . The set of segment endpoints has been partitioned into $A_1 \cup B_1$. Note that endpoints that are adjacent via an edge of P are always assigned to the same set.

with a neighboring segment in order to maintain parity between A 's and B 's. (See Figure 4; the final claim of Step 4 ensures there is a way to accomplish this without running out of segments.) Let $A_3 \cup B_3$ be the final partition of segment endpoints after the merging. We show that **whp**, all $a \in A_3$ have $d_1^+(a; B_3) \geq \frac{\epsilon_1 \theta_1}{200L} \log n$, and all $b \in B_3$ have $d_1^+(b; A_3) \geq \frac{\epsilon_1 \theta_1}{200L} \log n$. Furthermore, if we were to expose the D_1 -edges between A_3 and B_3 , then each vertex $a \in A_3$ would independently sample $d_1^+(a; B_3)$ uniformly random neighbors in B_3 , and similarly for $b \in B_3$.

This will complete the proof because the final number of segments is $|A_3| = (1 - o(1)) \frac{n}{L}$.

3.5.1 Step 1

By construction, $|B_1| = (1 - o(1)) \frac{n}{L}$. Now consider an arbitrary vertex $v \in P \setminus B_1$. We have only exposed the numeric value of $d_1^*(v)$ thus far in our construction, and not where the D_1 -out-neighbors are. So let us now expose the numeric value of $d_1^+(v; B_1)$, but again, not precisely where the out-endpoints are. As we observed in the beginning of Section 3.4, we have $d_1^*(v) \geq \frac{\epsilon_1 \theta_1}{20} \log n - 3$. Our work in the previous section consumes up to one D_1 -out-edge at each vertex $v \in P$, when we reveal whether $\vec{v\hat{w}} \in D_1^\circ$ in the third step of the algorithm. Therefore, $d_1^+(v; B_1)$ stochastically dominates $\text{Bin}\left(\frac{\epsilon_1 \theta_1}{20} \log n - 4, \frac{0.999}{L}\right)$. Hence we can use Lemma 3.1 to bound the probability that $d_1^+(v; B_1)$ is too small.

$$\begin{aligned} \mathbb{P}\left[d_1^+(v; B_1) < \frac{1}{9} \cdot \frac{0.998}{L} \cdot \frac{\epsilon_1 \theta_1}{20} \log n\right] &< e^{-0.533 \cdot \frac{0.998}{L} \cdot \frac{\epsilon_1 \theta_1}{20} \log n} \\ \mathbb{P}\left[d_1^+(v; B_1) < \frac{\epsilon_1 \theta_1}{181L} \log n\right] &= o(e^{-\frac{\epsilon_1 \theta_1}{40L} \log n}) \\ &= o(e^{-(\log n)^{8/9}}), \end{aligned}$$

since $\epsilon_1, \theta_1 = \Omega\left(\frac{1}{\sqrt{\log \log n}}\right)$ and $L < \sqrt[10]{\log n}$. The expected number of such vertices in P is at most n times this probability. Applying Markov's inequality, we conclude that **whp**, the number of B_1 -bad vertices in $P \setminus (A_1 \cup B_1)$ is at most $n \cdot e^{-\sqrt{\log n}}$, with room to spare. This also shows that the initial and final endpoints of P are B_1 -good **whp**.

3.5.2 Step 2

At this point, our entire vertex set is partitioned as follows. We have a collection of rainbow segments I_1, \dots, I_r , each of length exactly L . These already consume at least $n - \frac{n}{\sqrt[3]{\log n}}$ vertices. Since we discarded the remainder of P , as well as possibly the final segment, we have $r \geq \frac{1}{L} \left(n - \frac{n}{\sqrt[3]{\log n}}\right) - 2$. A separate collection of rainbow paths Q_1, \dots, Q_s covers all vertices of S . There are also some remaining vertices. In this section, we will use G_3 -edges to absorb the latter two classes into the rainbow segments.

Since we will not use any further G_2 -edges, but edges from D_2° take precedence over those from D_3 , we also now expose all edges in D_2° . Lemma 3.3 ensures that **whp**, no vertex is incident to more than $5 \log n$ edges of D_2° . Condition on this outcome. Note that by construction, we have not exposed the locations of any D_3° -edges between vertices outside S , although vertices outside S may have up to three exposed D_3 -neighbors located in S .

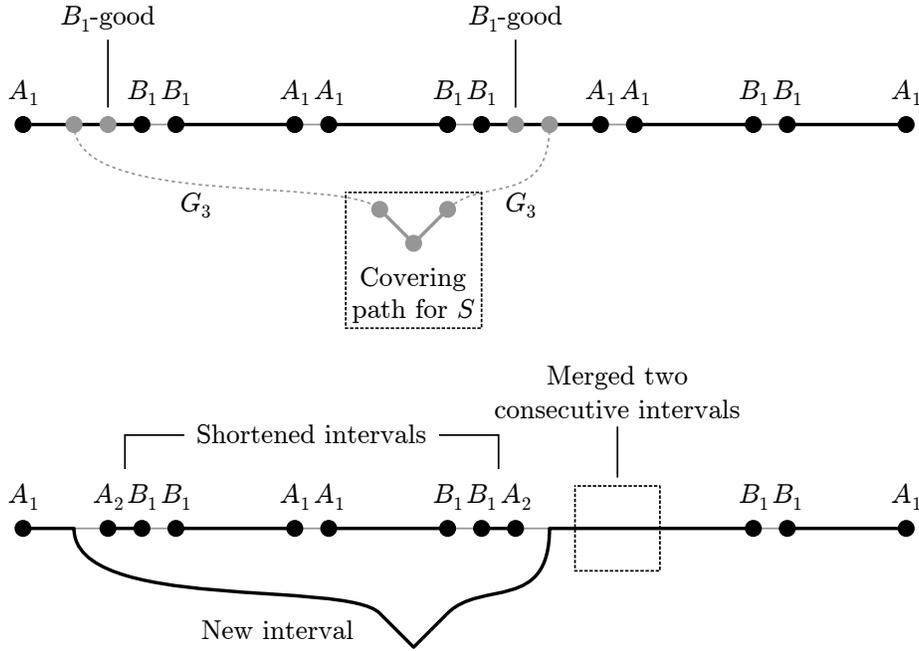


Figure 2: A covering path of S is absorbed into the system of segments using G_3 -edges. Note that the resulting endpoint partition still has one A -endpoint and one B -endpoint in every segment. Importantly, the direction of the splicing is such that all new endpoints are of type- A , as indicated by the A_2 -vertices. This is why we continue the new segment rightward, through to the next B -endpoint.

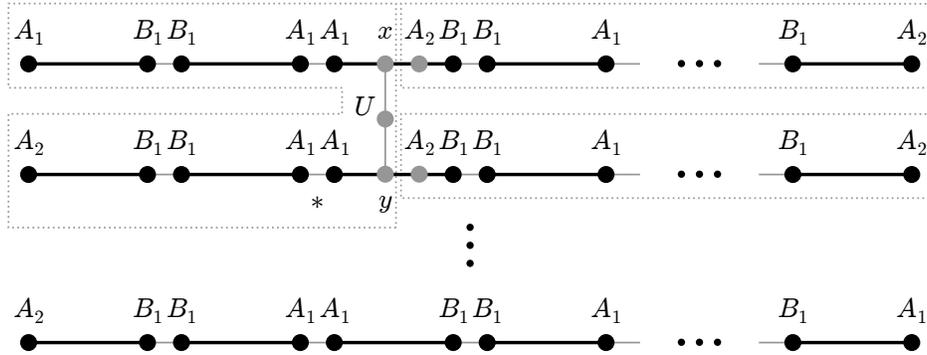


Figure 3: Evolution of block partition during absorption. Each horizontal row represents an original block, within which the successive segments have their endpoints connected by edges of the original path P . The vertical gray path from x to y represents the absorption of a new vertex into the collection of segments, involving two different blocks. This operation cuts the two edges between x, y and their adjacent A_2 -vertices, and adds back the P -edge marked by the asterisk. Afterward, the segments can be re-partitioned into new blocks (see the gray dotted lines), with all initial and final endpoints in each block of type- A , and B_1 -good.

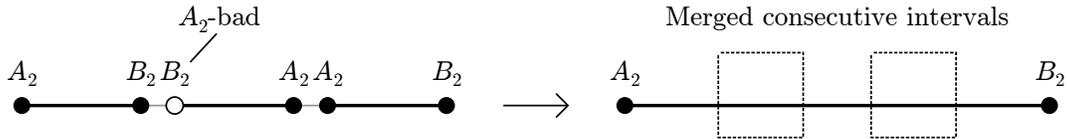


Figure 4: Original edges of P are used to merge consecutive segments in the same block, so that a bad endpoint can be eliminated.

We now use a simple greedy algorithm to absorb all residual paths and vertices into our collection of segments. In each step, we find a pair of G_3 -edges linking either a new Q_i or a new vertex to two distinct segments I_x and I_y , using two new colors from C_3 . We will ensure that throughout the process, all segments I_w used in this way are separated by at least $\frac{\sqrt[3]{\log n}}{L^2}$ full segments I_z along P .

The specific procedure is as follows. Suppose we have already linked in t paths or vertices, and are considering the next path or vertex to link in. Suppose it is a path Q_i (the vertex case can be treated in an analogous way). Let u, v be the endpoints of Q_i . We need to find vertices x, y in distinct segments I_x and I_y such that (i) according to P , I_x and I_y are separated by at least $\frac{\sqrt[3]{\log n}}{L^2}$ full segments from each other, and from all I_z previously used in this stage, (ii) x and y are separated from the endpoints of I_x and I_y by at least two edges of P , and (iii) if $x' \in I_x$ is the vertex adjacent to x in the direction of the B_1 -endpoint of I_x , and $y' \in I_y$ is the vertex adjacent to y in the direction of the B_1 -endpoint of I_y , then both x' and y' are B_1 -good. We choose the direction of the B_1 -endpoint because x' and y' will become the new A_2 -endpoints of shortened segments; see Figure 2 for an illustration.

So, let F be the set of vertices in the segments which fail properties (ii) or (iii). By Step 1, the dominant term arises from the endpoints because $n \cdot e^{-\sqrt{\log n}} \ll \frac{n}{\sqrt[10]{\log n}} < \frac{n}{L}$, so $|F| < \frac{5n}{L}$. Also let T be the set of vertices contained in segments that are at least $\frac{\sqrt[3]{\log n}}{L^2}$ full segments away from any segments which have previously been touched by this algorithm. Since we observed at the beginning of this section that the total number of segments was at least $\frac{1}{L}(n - \frac{n}{\sqrt[3]{\log n}}) - 2$, we have

$$|T| \geq \left(\frac{1}{L} \left(n - \frac{n}{\sqrt[3]{\log n}} \right) - 2 - \left(\frac{2\sqrt[3]{\log n}}{L^2} \right) \cdot (2t) \right) L > n - \frac{4.5n}{L},$$

since $t < \frac{1.1n}{\sqrt[3]{\log n}}$ and $L < \sqrt[10]{\log n}$. Let us now find a newly-colored G_3 -edge from u to a vertex of $T \setminus F$. Note that the number of vertices outside $T \setminus F$ is at most $\frac{9.5n}{L}$.

We have not yet exposed the specific locations of the D_3 -neighbors of u , but only know (since $u \notin S$) that $d_3^+(u) \geq \frac{\epsilon_3 \theta_3}{20} \log n$, and up to three of those D_3 -out-neighbors lie within S . Consider what happens when we expose the location w of one of u 's D_3 -out-neighbors which is outside S . This will produce a useful G_3 -edge uw if (i) w lands in $T \setminus B$, (ii) neither \vec{uw} or \vec{wu} appeared in D_1° or D_2° , (iii) \vec{wu} does not appear in D_3° , and (iv) the color of the edge is new.

Let us bound the probability that w fails any of these properties. We may consider (i)-(iii) together, since we showed that at most $\frac{9.5n}{L}$ vertices were outside $T \setminus B$, and we conditioned on u being incident to at most $5 \log n$ edges of D_2° . After sampling the location of w , we expose whether \vec{uw} or \vec{wu} appear in D_1° ; by the same argument as used in Lemma 3.10, the probability of each is at most $\frac{7 \log n}{n} = o(\frac{1}{L})$. So, the probability of failing either (i) or (ii) is at most $\frac{9.6}{L}$. We conditioned at the beginning of Section 3.4 on $d_3^+(w) \leq 5 \log n$, so when we expose whether \vec{wu} is in D_3° , we again fail only with probability at most $\frac{7 \log n}{n} = o(\frac{1}{L})$. Finally, when we expose the color of the new edge, we know that it will be in C_3 , so the probability that it is a previously used color is at most

$$(2t) \cdot \frac{1}{\theta_3 n} < \frac{2n}{\sqrt[3]{\log n}} \cdot \frac{1}{\theta_3 n} = \frac{2}{\theta_3 \sqrt[3]{\log n}} = o\left(\frac{1}{L}\right).$$

Therefore, the probability that all of the at least $\frac{\epsilon_3 \theta_3}{20} \log n - 3$ D_3 -out-edges of u fail is at most

$$\left(\frac{10}{L}\right)^{\frac{\epsilon_3 \theta_3}{20} \log n - 3} = \left(\frac{L}{10}\right)^3 n^{-\frac{\epsilon_3 \theta_3}{20} \log \frac{L}{10}} \leq \left(\frac{L}{10}\right)^3 n^{-2},$$

where we used the definition of L in (6) for the final bound. A similar calculation works for v , and for the separate case when we incorporate a new vertex into the segments. Therefore, taking a union bound over the $o(n)$ iterations in linking vertices and paths, we conclude that our procedure completes successfully **whp**. \square

3.5.3 Step 3

Step 1 established that **whp**, the initial and final vertices of the long path P are both B_1 -good, so the original system of segments can be arranged as a single block, with successive segments linked by edges of P . We now prove by induction that after the absorption of each path or vertex in Step 2, the collection of segments can be re-partitioned into blocks of segments, such that within each block, consecutive segments have their endpoints linked via P , and the initial and final endpoints in each block are of type- A , and B_1 -good.

There are two cases, depending on whether the absorption involves two segments in the same block (as in Figure 2), or in different blocks (as in Figure 3). If the segments are in the same block, then we can easily divide that block into two blocks satisfying the condition. Indeed, in Figure 2, one of the new blocks is the string of segments between the vertices indicated by A_2 in the diagram, and the other new block is the complement. This works because within each of the two new blocks, every edge between successive segments was an edge between successive segments of the original block, hence in P . Also, of the four initial/final endpoints among the two new blocks, two of them were the initial/final endpoints of the original block, and the other two were identified as B_1 -good vertices, now in A_2 . Therefore, the new block partition satisfies the requirements.

On the other hand, if the absorption involves two segments from different blocks, then one can re-partition the two blocks into three new blocks, as illustrated in Figure 3. A similar analysis to above then completes the argument. Clearly, since the intervals chosen for the splicing are separated by at least $\frac{\sqrt[3]{\log n}}{L^2}$ segments, it also easily follows that the resulting blocks contain at least that many segments as well.

3.5.4 Step 4

We have not yet revealed anything about the D_1 -out-neighbors of any vertices in $B_2 = B_1$; the only thing we know is that they had $d_1^+ \geq \frac{\epsilon_1 \theta_1}{20} \log n$. For each vertex $b \in B_1$, let us now expose the numeric value of $d_1^+(b; A_2)$, but again, not precisely where the out-endpoints are. Since our absorption procedure maintained $|A_2| = |A_1| = (1 - o(1))\frac{n}{L}$, the same argument that we used for Step 1 now establishes the bound for Step 4, with plenty of room to spare.

It remains to show that every block contains a string of 4 consecutive segments, all 8 of whose endpoints are good. Consider an arbitrary block. Step 3 ensures that this block has at least $\frac{\sqrt[3]{\log n}}{L^2}$ segments, which constitute at least $\frac{\sqrt[3]{\log n}}{4L^2}$ disjoint strings of 4 consecutive segments. By the argument in Step 1, the probability that a particular endpoint is bad is $o(e^{-(\log n)^{8/9}})$, so the probability that a particular string of 4 segments has a bad endpoint is at most 8 times that bound, which is also $o(e^{-(\log n)^{8/9}})$. Therefore, the probability that all of these disjoint strings are bad in this block is at most

$$\left(e^{-(\log n)^{8/9}} \right)^{\frac{\sqrt[3]{\log n}}{4L^2}} < e^{-(\log n)^{1.02}} = o(n^{-1}),$$

where we used $L < \sqrt[10]{\log n}$. Taking a union bound over the $o(n)$ blocks, we find that **whp**, every block contains a string of 4 consecutive segments, all of whose endpoints are good.

3.5.5 Step 5

By Steps 1 and 4, the total number of merges which occur in Step 5 is well below $O(n \cdot e^{-\sqrt{\log n}}) = o(n/L)$. It is therefore easy to see that the merging can be done so that there are $(1 - o(1))n/L$ segments at the end. Also, since Step 4 ensures that every block contains a string of 4 well-behaved segments, the merging is possible even though the procedure absorbs neighboring segments in order to preserve parity between A 's and B 's. To see this, start from the endpoint at the very beginning of the block, and move toward the string of 4 well-behaved segments. When a bad endpoint is encountered, merge the two adjacent segments, as well as the next segment in the direction that we are traveling. The endpoint of the resulting segment is then the other endpoint of the merged neighboring segment. If that is also bad, then merge in the next segment after that as well, and even the following segment to maintain parity. (In this example, we have now merged four segments into one.) Keep proceeding in this way until the string of 4 well-behaved segments is reached. Since all endpoints of those 4 segments are good, at worst this procedure will merge in the first two of those well-behaved segments before stopping at good endpoints. Then, start from the very last endpoint of the entire block, and move backwards toward the 4 well-behaved segments, following the same procedure. Again, at worst the last two of the well-behaved segments will be merged by this procedure. Importantly, the two procedures will not interact with each other, due to the sufficiently long well-behaved region in the middle.

Since $B_3 \subset B_2 = B_1$ and $A_3 \subset A_2$, for every $a \in A_3$ we can independently sample $d_1^+(a; B_3)$ using only the value of $d_1^+(a; B_1)$. Indeed, since $a \in A_3$ was B_1 -good, it had $d_1^+(a; B_1) \geq \frac{\epsilon_1 \theta_1}{181L} \log n$. We will only have $d_1^+(a; B_3) < \frac{\epsilon_1 \theta_1}{200L} \log n$ if at least $\frac{\epsilon_1 \theta_1}{181L} \log n - \frac{\epsilon_1 \theta_1}{200L} \log n > \frac{\epsilon_1 \theta_1}{2000L} \log n$ of those out-neighbors land in $B_1 \setminus B_3$ as opposed to B_3 . Let

$$q = \frac{|B_1 \setminus B_3|}{|B_1|} = O\left(\frac{n \cdot e^{-\sqrt{\log n}}}{n/L}\right) = e^{-(1-o(1))\sqrt{\log n}}.$$

Note that

$$\begin{aligned} \mathbb{P}\left[\text{Bin}\left(\frac{\epsilon_1 \theta_1}{181L} \log n, q\right) \geq \frac{\epsilon_1 \theta_1}{2000L} \log n\right] &\leq \left(\frac{\epsilon_1 \theta_1}{181L} \log n\right) q^{\frac{\epsilon_1 \theta_1}{2000L} \log n} \\ &\leq (12eq)^{\frac{\epsilon_1 \theta_1}{2000L} \log n} \\ &= e^{-(1-o(1))\sqrt{\log n} \cdot \frac{\epsilon_1 \theta_1}{2000L} \log n} \\ &< e^{-(\log n)^{4/3}} \\ &= o(n^{-1}). \end{aligned}$$

Therefore, a final union bound establishes that **whp**, every $a \in A_3$ has $d_1^+(a; B_3) \geq \frac{\epsilon_1 \theta_1}{200L} \log n$.

A similar argument establishes the bound for $d_1^+(b; A_3)$, for $b \in B_3$, because we had $A_3 \subset A_2$, and had only previously exposed the value of $d_1^+(b; A_2)$. The last claim in Step 5 is clear from our order of exposure.

3.6 Proof of Lemma 2.8

In this final stage of the proof, we use G_1 to link together the endpoints of the system of segments I_1, \dots, I_r created by Lemma 2.5. As described in the overview (Section 2.3), we construct an auxiliary directed graph Γ . Importantly, no D_1° -edges have been revealed between the endpoint sets A and B , so we may now specify a model for the random r -vertex digraph Γ .

Indeed, consider a vertex $w_k \in \Gamma$, $1 \leq k \leq r$, and let a, b be the A - and B -endpoints of the corresponding segment I_k . We first generate a set E_1 of Γ -edges by sending exactly $d_1^+(b; A)$ directed edges out of w_k , and exactly $d_1^+(a; B)$ directed edges into w_k . This is analogous to the d -in, d -out model, except that not all degrees are equal. Some directed edges will be generated twice and for these edges we keep only one copy in E_1 . Let $F_1 \subseteq E_1$ be these edges that are generated exactly twice in the algorithm. Color every edge of E_1 independently from C_1 . Finally, generate a random subset $F_2 \subset E_1 \setminus F_1$ by independently sampling each edge of $E_1 \setminus F_1$ with probability $\frac{1}{2} p_1 \cdot \left(\frac{1+\theta_2+\theta_3}{1+\theta_1+\theta_2+\theta_3} \right)$. Let $E_1 \setminus F_2$ be the final edge set of Γ .

The reason for the removal of F_2 is that some of the initially-generated edges of Γ will find conflicts once D_1° is generated. Indeed, every edge $\overrightarrow{w_j w_k}$ that we have placed in Γ corresponds to an edge $\overrightarrow{ba} \in D_1$ or $\overrightarrow{ab} \in D_1$ (or both), for some $b \in B$, $a \in A$. When both do not occur, and only $\overrightarrow{ba} \in D_1$ (say), then we need to expose whether $\overrightarrow{ab} \in D_1^\circ \setminus D_1$; if it is in $D_1^\circ \setminus D_1$, it removes $\overrightarrow{w_j w_k}$ from Γ with probability $1/2$.

To simplify notation, let $\delta^+(w_k)$ and $\delta^-(w_k)$ be the numbers of out- and in-edges that are generated at w_k to build the initial edge set E_1 . They correspond to $d_1^+(b; A)$ and $d_1^+(a; B)$ above, and have therefore been revealed by our previous exposures. Importantly, we have the bounds

$$\frac{\epsilon_1 \theta_1}{200L} \log n \leq \delta^\pm(w_k) \leq 5 \log n.$$

It is more convenient to restrict our attention to a smaller subset $E_2 \subset E_1$ which is itself already rainbow; then, every ordinary directed Hamilton cycle will automatically be rainbow. For this, we expose at every vertex w_k what the colors of the $\delta^+(w_k)$ out-edges and $\delta^-(w_k)$ in-edges will be, but not their locations.

Lemma 3.11. *Suppose that $\delta^\pm(w_k) \geq \frac{\epsilon_1 \theta_1}{200L} \log n$ for all $1 \leq k \leq r$. Then **whp**, it is possible to select 3 out-edges O_k and 3 in-edges I_k from each w_k so that all $6r$ selected colors are distinct. Furthermore, the endpoints of the edges in the sets O_k, I_k are independent of w_k and the endpoints of O_l, I_l , $l \neq k$.*

Proof. Construct an auxiliary bipartite graph H with vertex partition $W \cup C_1$, where $W = \{w_1^+, w_1^-, \dots, w_r^+, w_r^-\}$. Place an edge between w_k^+ and c if one of w_k 's $\delta^+(w_k)$ out-edges has color c . Edges between w_k^- and c are defined with respect to w_k 's in-edge colors. The desired result is a perfect 1-to-3 matching in H . For this, we apply the 1-to-3 version of Hall's theorem: we must show that for every $X \subset W$, we have $|N(X)| \geq 3|X|$, where $N(X)$ is the union of the H -neighborhoods of all vertices of X .

This follows from a standard union bound. Indeed, fix an integer $1 \leq x \leq 2r$, and consider an arbitrary pair of subsets $X \subset W$ and $Y \subset C$, with $|X| = x$ and $|Y| = 3x$. The probability that $N(X) \subset Y$ (in H) is at most

$$\left[\left(\frac{3x}{\theta_1 n} \right)^{\frac{\epsilon_1 \theta_1}{200L} \log n} \right]^x.$$

The innermost term is the probability that a random color from C_1 is in Y . The exponents come from the fact that each vertex of $w_k \in X$ samples at least $\delta^\pm(w_k) \geq \frac{\epsilon_1 \theta_1}{200L} \log n$ colors for its out- and in-neighbors.

Multiplying this bound by the number of ways there are to select X and Y , and using $r \leq \frac{n}{L}$, we find that the probability of failure for a fixed x is at most

$$\begin{aligned} \binom{2r}{x} \binom{\theta_1 n}{3x} \left[\left(\frac{3x}{\theta_1 n} \right)^{\frac{\epsilon_1 \theta_1}{200L} \log n} \right]^x &\leq \left(\frac{2er}{x} \right)^x \left(\frac{\theta_1 n}{3x} \right)^{3x} \left[\left(\frac{3x}{\theta_1 n} \right)^{\frac{\epsilon_1 \theta_1}{200L} \log n} \right]^x \\ &= \left[\left(\frac{2en}{Lx} \right) \left(\frac{\theta_1 n}{3x} \right)^3 \left(\frac{3x}{\theta_1 n} \right)^{\frac{\epsilon_1 \theta_1}{200L} \log n} \right]^x. \end{aligned}$$

We will sum this over all $1 \leq x \leq 2r$. The outer exponent allows us to bound this by a decreasing geometric series, so it suffices to show that the interior of the square bracket is uniformly $o(1)$ for all $1 \leq x \leq 2r$. Indeed, observe that the exponent of x inside the bracket is $\frac{\epsilon_1 \theta_1}{200L} \log n - 4 > 0$, so it is maximized at $x = 2r \leq \frac{2n}{L}$. Yet

$$\left(\frac{2en}{L(2n/L)} \right) \left(\frac{\theta_1 n}{3(2n/L)} \right)^3 \left(\frac{3(2n/L)}{\theta_1 n} \right)^{\frac{\epsilon_1 \theta_1}{200L} \log n} = (e) \left(\frac{6}{\theta_1 L} \right)^{\frac{\epsilon_1 \theta_1}{200L} \log n - 3}$$

Since (6) ensures that $L \geq \frac{7}{\theta_1}$, we have $\frac{6}{\theta_1 L} \leq \frac{6}{7}$. Yet the exponent $\frac{\epsilon_1 \theta_1}{200L} \log n - 3$ tends to infinity as n grows, so we indeed obtain a uniform upper bound of $o(1)$ for all $1 \leq x \leq 2r$. Therefore, **whp**, every subset $X \subset W$ has $|N(X)| > 3|X|$, and the 1-to-3 version of Hall's theorem establishes the desired result.

Observe that our bipartite graph H does not depend on the neighbors of each w_k , only the colors of the edges of the vertices incident with w_k . Hence the output of the construction is also independent of the neighbors of each w_k . \square

Recall from the beginning of this section that the final edge set of Γ is $E_1 \setminus F_2$. Let E_2 be the set of $6r$ edges selected by Lemma 3.11. This corresponds to a copy D_3 of $D_{3\text{-in},3\text{-out}}$. Unfortunately, in our model we still need to expose the locations of the remaining $\delta^\pm(w_k) - 3$ remaining in- and out-edges at every vertex w_k , where we can assume that $\delta^\pm(w_k) \leq 5 \log n$. It is still possible that an edge in E_2 may be generated again in this stage. That edge would then have $1/2$ probability of receiving the color of the new copy, which would not be in our specially constructed rainbow set. To account for this, let F_3 be the set of edges generated by exposing these *remaining* in- and out-edges. These edges will be deleted from D_3 if they duplicate an edge of D_3 .

Here the construction is $D_{3\text{-in},3\text{-out}}$ plus some more randomly chosen edges, which can be thought of as a sub-graph of $D_4 = D_{(5 \log n)\text{-in},(5 \log n)\text{-out}}$, where some D_4 edges cause their duplication in D_3 to be deleted. The only dependencies between D_3 and D_4 are that if a vertex v chooses w as an out-neighbor in D_3 , then w will not be chosen again as an out-neighbor by v in D_4 . We handle these dependencies by allowing extra edges:

- (i) We embed D_4 in $D_5 = D_{(6 \log n)\text{-in},(6 \log n)\text{-out}}$ where we ignore D_3 in the construction of D_5 .
- (ii) We then simplify matters by embedding D_5 in a copy D_6 of $D_{r, \frac{120 \log n}{r}}$.

For (ii) we observe a simple coupling which realizes $D_{(6 \log n)\text{-in},(6 \log n)\text{-out}}$ as a subgraph of $D_{r, \frac{120 \log n}{r}}$ **whp**. For this, consider the following system for generating a random directed graph. For every ordered pair of vertices (u, v) , generate two independent Bernoulli random variables $I_{u,v}^+$ and $I_{u,v}^-$, each with probability parameter $\frac{60 \log n}{r}$. Create the directed edge \vec{uv} , if and only if at least one of $I_{u,v}^+$ or $I_{v,u}^-$ took the value 1. For each vertex u , let $D^+(u)$ be the number of other vertices v for which $I_{u,v}^+ = 1$, and let $D^-(u)$ be the number of $I_{u,v}^- = 1$. These are all distributed as $\text{Bin}\left(r-1, \frac{60 \log n}{r}\right)$. Lemma 3.1 establishes that for fixed u , the probability that $D^+(u) < 6 \log n$ is at most $e^{-0.533 \cdot (60 - o(1)) \log n}$, and similarly for $D^-(u)$. Therefore, a union bound establishes that **whp**, all $D^+(u), D^-(u) \geq 6 \log n$. Conditioning on the values of $D^+(u)$ and $D^-(u)$, we see that when the indicators are revealed, this indeed contains $D_{(6 \log n)\text{-in},(6 \log n)\text{-out}}$. The result then follows by recalling that $r = (1 - o(1)) \frac{n}{L}$, and taking the union of $D_{r, \frac{120L \log n}{n}}$ with another independent $D_{r, \frac{\log n}{n}}$ to cover F_2 .

There are now two ways in which an edge e of D_3 can be deleted: either from F_2 or from D_6 . The event $e \in F_2$ is independent of all events involving other edges.

It suffices to find a directed Hamilton cycle in what is left of D_3 without regard to color. At each vertex u , some subset of its 6 edges are in $F_2 \cup D_6$. These choices for possible deletion are independent. Furthermore, the probability that more than one of the 6 in- or out-edges is deleted is at most

$$\binom{6}{2} \left(\frac{1}{2} p_1 \cdot \left(\frac{1 + \theta_2 + \theta_3}{1 + \theta_1 + \theta_2 + \theta_3} \right) + \frac{K \log n}{r} \right)^2 = o(n^{-1}),$$

so a union bound establishes that **whp**, $E_2 \setminus (F_2 \cup D_6)$ contains a copy of $D_{2\text{-in},2\text{-out}}$. This is known to be Hamiltonian **whp** by Theorem 2.7 from Section 2.3. This completes the proof of the final lemma in our proof of Theorem 1.1, and so our entire proof is complete. \square

4 Concluding remarks

Our main contribution, part (b) of Theorem 1.1, sharpens the earlier result of Cooper and Frieze [8] to achieve optimal first-order asymptotics. As we mentioned in the introduction, we suspect that our result can be further sharpened within the $o(1)$ term. We do not push to optimize our bounds on ϵ, θ because it is not clear that incremental improvements upon our current approach will be sufficiently interesting. Instead, we would be more interested in determining whether one can extend the celebrated “hitting time” result of Bollobás [4], which states that one can typically find a Hamilton cycle in the random graph process as soon as the minimum degree reaches two.

Question 4.1. *Consider the edge-colored random graph process in which e_1, e_2, \dots, e_N , $N = \binom{n}{2}$ is a random permutation of the edges of K_n . The graph G_m is defined as $([n], \{e_1, e_2, \dots, e_m\})$. Let each edge receive a random color from a set C of size at least n . Then **whp**, the first time that a rainbow Hamilton cycle appears is precisely the same as the first time that the minimum degree of G_m is at least two **and** at least n colors have appeared.*

Although this may be out of reach at the moment, another natural challenge is to settle the problem in either the case when the number of edges is just sufficient for an ordinary Hamilton cycle, or in the case when the number of colors is, say, exactly n . The following argument comes tantalizingly close to proving that $Kn \log n$ edges are sufficient when we have exactly n colors.

Let $H_{n,p;3}$ denote the random 3-uniform hypergraph where each potential hyperedge appears independently with probability p . In this object, a *loose Hamilton cycle* is a permutation of the vertices (v_1, \dots, v_n) such that $\{v_1, v_2, v_3\}, \{v_3, v_4, v_5\}, \dots, \{v_{n-1}, v_n, v_1\}$ all appear as hyperedges (n must be even).

Theorem 4.2. (See [11], Theorem 1.) *There is an absolute constant K such that if $p > \frac{K \log n}{n^2}$, then for all n divisible by 4, the random hypergraph $H_{n,p;3}$ on the vertex set $\{x_1, \dots, x_{n/2}, y_1, \dots, y_{n/2}\}$ contains a loose Hamilton cycle **whp**. Furthermore, one can find such a cycle of the special form $(x_{\sigma(1)}, y_{\tau(1)}, x_{\sigma(2)}, y_{\tau(2)}, \dots, x_{\sigma(n/2)}, y_{\tau(n/2)})$, for some permutations $\sigma, \tau \in S_{n/2}$.*

As mentioned in the introduction, this theorem was proven by connecting loose Hamilton cycles in random 3-uniform hypergraphs with rainbow Hamilton cycles in randomly edge-colored random graphs, and applying a result of Janson and Wormald [16]. We will now consider the reverse connection.

Let K be the constant in Theorem 4.2. We are given an even integer n , and a graph $G \sim G_{n,p}$ with $p > \frac{K \log n}{n}$ on vertex set $\{v_1, \dots, v_n\}$, whose edges are randomly colored from the set $\{c_1, \dots, c_n\}$. Construct an auxiliary 3-uniform hypergraph H with vertex set $\{v_1, \dots, v_n, c_1, \dots, c_n\}$, by taking the hyperedge $\{v_i, v_j, c_k\}$ whenever the edge $v_i v_j$ appears in G , with color c_k . Note that every such hyperedge appears independently with probability $\frac{p}{n} > \frac{K \log n}{n^2}$, since $v_i v_j$ appears in G with probability p , and receives color c_k with probability $\frac{1}{n}$. There is a small difference between the distribution of H and $H_{n,p;3}$. In the latter we can have edges $\{v, w, c_1\}$ and $\{v, w, c_2\}$ where $c_1 \neq c_2$, whereas this cannot happen in H . So we cannot apply Theorem 4.2 directly to argue that H has a loose Hamilton cycle **whp**, of the form $(v_{\sigma(1)}, c_{\tau(1)}, v_{\sigma(2)}, c_{\tau(2)}, \dots, c_{\tau(n/2)})$. This would correspond to the Hamilton cycle $(v_{\sigma(1)}, \dots, v_{\sigma(n/2)})$ in G , with edges colored $c_{\tau(1)}, \dots, c_{\tau(n/2)}$, hence rainbow.

To fix this unfortunate situation, we must re-prove Theorem 4.2, changing the model to disallow edges $\{v, w, c_1\}$ and $\{v, w, c_2\}$ to occur. This would, we think involve amending the proof of a special case of the main result of Johansson, Kahn and Vu [18]. This can probably be done, but we don't feel that it is right to merely say so.

References

- [1] M. Albert, A.M. Frieze and B. Reed, Multicoloured Hamilton cycles, *Electronic Journal of Combinatorics* **2** (1995), R10.
- [2] N. Alon and J. Spencer, **The Probabilistic Method**, 3rd ed., Wiley, New York, 2007.
- [3] B. Bollobás, **Random Graphs**, 2nd ed., Cambridge University Press, 2001.
- [4] B. Bollobás, The evolution of sparse graphs, in: **Graph Theory and Combinatorics**, Proc. Cambridge Combinatorial Conf. in honour of Paul Erdős (B. Bollobás, ed.), Academic Press, 1984, 35–57.
- [5] B. Bollobás and A. Frieze, On matchings and Hamiltonian cycles in random graphs, in: **Random Graphs '83 (Poznan, 1983)**, North-Holland Math. Stud., 118, North-Holland, Amsterdam (1985), 23–46.

- [6] C. Cooper and A.M. Frieze, Hamilton cycles in random graphs and directed graphs, *Random Structures and Algorithms* **16** (2000), 369–401.
- [7] C. Cooper and A.M. Frieze, Multicoloured Hamilton cycles in random graphs: an anti-Ramsey threshold, *Electronic Journal of Combinatorics* **2** (1995), R19.
- [8] C. Cooper and A.M. Frieze, Multi-coloured Hamilton cycles in random edge-coloured graphs, *Combinatorics, Probability and Computing* **11** (2002), 129–134.
- [9] A. Dudek and A.M. Frieze, Loose Hamilton cycles in random k -uniform hypergraphs, *Electronic Journal of Combinatorics* **18** (2011), P48.
- [10] P. Erdős, J. Nešetřil, and V. Rödl, Some problems related to partitions of edges of a graph, in: **Graphs and Other Combinatorial Topics**, Teubner, Leipzig (1983), 54–63.
- [11] A.M. Frieze, Loose Hamilton cycles in random 3-uniform hypergraphs, *Electronic Journal of Combinatorics* **17** (2010), N28.
- [12] A.M. Frieze and M. Krivelevich, Packing Hamilton cycles in random and pseudo-random hypergraphs, submitted.
- [13] A.M. Frieze, M. Krivelevich, and P. Loh, Packing tight Hamilton cycles in 3-uniform hypergraphs, *Random Structures and Algorithms*, to appear.
- [14] A.M. Frieze, An algorithm for finding Hamilton cycles in random digraphs, *Journal of Algorithms* **9** (1988), 181–204.
- [15] A.M. Frieze and B. D. McKay, Multicoloured trees in random graphs, *Random Structures and Algorithms* **5** (1994), 45–56.
- [16] S. Janson and N. Wormald, Rainbow Hamilton cycles in random regular graphs, *Random Structures Algorithms* **30** (2007), 35–49.
- [17] G. Hahn and C. Thomassen, Path and cycle sub-Ramsey numbers, and an edge colouring conjecture, *Discrete Mathematics*. **62** (1986), 29–33.
- [18] A. Johansson, J. Kahn and V. Vu, Factors in random graphs, *Random Structures and Algorithms* **33** (2008) 35-49.
- [19] J. Komlós and E. Szemerédi, Limit distribution for the existence of Hamiltonian cycles in a random graph, *Discrete Mathematics* **43** (1983), 55–63.
- [20] C.J.H. McDiarmid, General percolation and random graphs, *Adv. Appl. Probab.* **13** (1981), 40–60.
- [21] C.J.H. McDiarmid, General first-passage percolation, *Adv. Appl. Probab.* **15** (1983), 149–161.
- [22] B.D. McKay and N.C. Wormald, Asymptotic enumeration by degree sequence of graphs with degree $o(n^{1/2})$, *Combinatorica* **11** (1991), 369–382.

- [23] W. Fernandez de la Vega, Long paths in random graphs, *Studia Math. Sci. Hungar.* (1979), 335–340.
- [24] R.W. Robinson and N.C. Wormald, Almost all regular graphs are Hamiltonian, *Random Structures and Algorithms* **5** (1994), 363–374.
- [25] R. Rue, Comment on [1]. *Electronic Journal of Combinatorics* **2** (1995).