Rainbow spanning trees in randomly coloured G_{k-out}

Deepak Bal,[*] Alan Frieze,[†] and Paweł Prałat[‡]

November 2, 2023

Abstract

Given a graph $G = (V, E)$ on n vertices and an assignment of colours to its edges, a set of edges $S \subseteq E$ is said to be rainbow if edges from S have pairwise different colours assigned to them. In this paper, we investigate rainbow spanning trees in randomly coloured random G_{k-out} graphs.

1   Introduction

Let $G = (V, E)$ be a graph in which the edges are coloured. A set $S \subseteq E$ is said to be *rainbow coloured* if each edge of S is in a different colour. There is by now a large body of research on the existence of rainbow structures in randomly coloured random graphs. Let us highlight a few selected results. Frieze and McKay [14] and Bal, Bennett, Frieze and Prałat [1] studied the existence of rainbow spanning trees in $G_{n,m}$, the classical Erdős–Rényi random graph process. Cooper and Frieze [6], Frieze and Loh [13] and Ferber and Krivelevich [10] studied the existence of rainbow Hamilton cycles in $G_{n,m}$. Janson and Wormald [15] studied the existence of rainbow Hamilton cycles in random regular graphs. Finally, Bal, Bennett, Pérez-Giménez and Prałat [2], investigated rainbow perfect matchings and Hamilton cycles in random geometric graphs. Of the most popular random graph models, what is missing here is the random multigraph G_{k-out}. The aim of this paper is to initiate the study of these problems in the context of a randomly coloured G_{k-out} graphs.

All asymptotics throughout are as $n \to \infty$ (we emphasize that the notations $o(\cdot)$ and $O(\cdot)$ refer to functions of n, not necessarily positive, whose growth is bounded). We say that an event in

---

[*]Department of Mathematics, Montclair State University, Montclair NJ 07043, USA; e-mail: `deepak.bal@montclair.edu`

[†]Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh PA 15213, USA; e-mail: `frieze@cmu.edu`; research supported in part by NSF grant DMS1952285.

[‡]Department of Mathematics, Toronto Metropolitan University, Toronto, ON, Canada; e-mail: `pralat@torontomu.ca`; research supported in part by NSERC grant; part of this work was done while the author was visiting the Simons Institute for the Theory of Computing.

a probability space holds *with high probability* (or *w.h.p.*) if the probability that it holds tends to one as $n \to \infty$. We often write G_{k-out} when we mean a graph drawn from the distribution G_{k-out}.

The random graph $G_{k-out} = G_{k-out}(n)$ is defined as follows. It has vertex set $[n] := \{1, \dots, n\}$ and each vertex $i \in [n]$ independently chooses k random distinct neighbours from $[n] \setminus \{i\}$, so that each of the $\binom{n-1}{k}$ sets is equally likely to be chosen. It was shown by Fenner and Frieze [8] that G_{k-out} is k-connected w.h.p. for $k \geq 2$. It was shown by Frieze [11] that G_{2-out} has a perfect matching w.h.p., and by Bohman and Frieze [3] that G_{3-out} is Hamiltonian w.h.p. All of the above results are sharp. For more details we direct the reader to Chapter 18 in [12].

We define the randomly coloured graph $G_{k,q} = G_{k,q}(n)$ (not to be confused with $G_{n,m}$) as follows: the underlying graph on n vertices is G_{k-out} and (i) there is a set Q of q colours, (ii) each colour appears $\rho := \lfloor kn/q \rfloor$ or $\rho + 1$ times (there are $kn - q\rho$ *popular* colours that appear $\rho + 1$ times, the remaining colours are *unpopular* and appear ρ times; note that if q divides kn, then all colours are unpopular), (iii) kn colours, including repetitions, are randomly assigned to the kn edges of G_{k-out}. Finally, let us note that, without loss of generality, we may assume that $q \leq kn$. Indeed, if the number of colours is more than kn, then some colours are not used at all and the problem is equivalent to the one with $q = kn$.

In this paper we investigate spanning trees. We will prove the following theorem.

Theorem 1. *If $k \geq 2$ and $q \geq n - 1$, then $G_{k,q}$ has a Rainbow Spanning Tree (**RST**) w.h.p.*

The result is best possible. Trivially, if $q \leq n - 2$, then there are not enough colours to create a rainbow tree. If $k = 1$, then $G_{k,q}$ is disconnected w.h.p. [8].

2 Preliminaries

2.1 Colour Monotonicity

In our problem, we randomly colour kn edges of $G_{k,q}$ with q colours and we aim to create a rainbow structure. Recall that there are $q_2 = kn - q\rho$ popular colours that are present $\rho + 1$ times and $q_1 = q - q_2$ unpopular colours that are present ρ times.

It is natural to expect that the more colours are available, the easier it is to achieve our goal. We prove this monotonicity property in the following, slightly broader, context. Suppose that we are given a finite set X and a set of colours C where $|C| = |X|$. (In our application, X is the set of kn edges of $G_{k,q}$ and C is the set of kn colours: q colours from set Q, including repetitions.) We also have two distinct partitions of C: $\mathcal{C} = \{C_1, \dots, C_q\}$ and $\widehat{\mathcal{C}} = \{\widehat{C}_1, \dots, \widehat{C}_{q+1}\}$, for some positive integer $q \leq |X| - 1$. (In our application, each part corresponds to a colour from set Q. Partitions $\mathcal{C}$ and $\widehat{\mathcal{C}}$ correspond to colourings with q and $q + 1$ colours respectively.) Let $\rho = \lfloor |X|/q \rfloor$, $\widehat{\rho} = \lfloor |X|/(q + 1) \rfloor$, $q_2 = |X| - q\rho$, and $q_1 = q - q_2$. Suppose that $|C_i| = \rho$ for $1 \leq i \leq q_1$ and $|C_i| = \rho + 1$ for $q_1 + 1 \leq i \leq q$, that is, there are q_1 parts in $\mathcal{C}$ of size ρ and q_2 parts of size $\rho + 1$.

We are given a collection of sets $X_1, X_2, \ldots$, each set X_i is a subset of X. (In our application, the X_i are the edges of spanning trees, perfect matchings, Hamilton cycles, etc.) Our goal is to create at least one rainbow set from this collection. Let us consider a random colouring of X via a random bijection from C to X. In order to show that the probability that some X_i is rainbow in a random colouring with $q + 1$ colours is at least the corresponding probability when elements of X are coloured with q colours, we need to couple the two partitions. In order to do that we need to consider two cases.

Case 1: $\widehat{\rho} = \rho$. Partition $\widehat{\mathcal{C}}$ is obtained from $\mathcal{C}$ by choosing ρ parts in $\mathcal{C}$ of size $\rho + 1$ and replacing them with $\rho + 1$ parts of size ρ (see Figure 1). We couple the two colourings by first randomly mapping the $|C| - \rho(\rho + 1)$ colours from the parts that are the same in both partitions, and conditioning on the result. If some rainbow X_i is created, then it is clearly present in both colourings. Otherwise, it is easy to see that partition $\widehat{\mathcal{C}}$ is at least as likely to complete a rainbow colouring. Indeed, suppose that some X_i has s elements that are not coloured yet; we may assume that $1 \leq s \leq \rho$ as, otherwise, such a set cannot be rainbow via the first partition (it could be rainbow via the second partition if $s = \rho + 1$). The probability that partition $\mathcal{C}$ completes a rainbow colouring is equal to $\prod_{i=1}^{s-1} \frac{\rho(\rho+1)-i(\rho+1)}{\rho(\rho+1)-i}$ that is at most the corresponding probability for partition $\widehat{\mathcal{C}}$, namely, $\prod_{i=1}^{s-1} \frac{\rho(\rho+1)-i\rho}{\rho(\rho+1)-i}$.
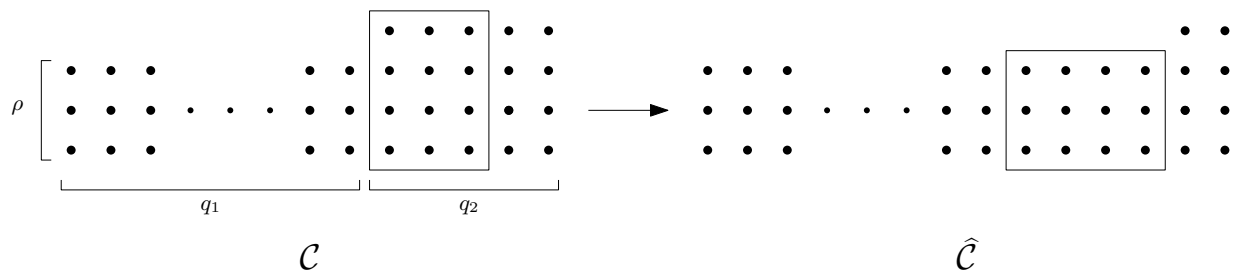

Figure 1: The coupling of the two colourings in Case 1.

Case 2: $\widehat{\rho} = \rho - 1$. As before, we start with partition $\mathcal{C}$ but this time we take all q_2 parts of size $\rho + 1$ (possibly $q_2 = 0$ so we might not have any) and we choose $\rho - 1 - q_2$ parts of size ρ. To create partition $\widehat{\mathcal{C}}$, we replace them with q_2 parts of size ρ and $\rho - q_2$ parts of size $\rho - 1$. As in the other case, either we create some rainbow X_i or partition $\widehat{\mathcal{C}}$ is at least as likely to complete a rainbow colouring of some set X_i.

The above coupling allows us to concentrate on the minimum number of colours. In particular, in the proof of Theorem 1, without loss of generality, we may assume that $q = n - 1$.

2.2 Degree Monotonicity

Based on Section 2.1, in order to prove Theorem 1 we may assume that $q = |Q| = n - 1$. In this setup, there are k popular colours present $k + 1$ times in $G_{k,q}$ and $q - k = n - 1 - k$ unpopular colours present k times in $G_{k,q}$. Exactly one out of $k + 1$ copies of each popular colour is called *special*. Similarly, there are $k + 1$ popular colours present $k + 2$ times in $G_{k+1,q}$, $q - (k+1) = n - 2 - k$ unpopular colours present $k + 1$ times, and there are $k + 1$ special copies of popular colours (one special copy of each).

It seems reasonable to expect that if $G_{k,q}$ has a particular rainbow structure w.h.p., then so does $G_{k+1,q}$. Let Γ_{k+1} be the bipartite graph with vertex sets $[n]$ and $Q' = Q \cup \{q\}$, where q is a "dummy" vertex that will be associated with special copies of popular colours. For each of the $(k+1)$ edges chosen by $v \in [n]$ in $G_{k+1,q}$, we observe a colour c of that edge without exposing its other endpoint. If a copy of colour c is non-special, then we add an edge between v and $c \in Q$ in Γ_{k+1}; otherwise, we add an edge between v and the "dummy" vertex q. Note that we need the extra vertex q to make Γ_{k+1} regular.

Recall that $(k+1)n$ colours, including repetitions, are randomly assigned to the $(k+1)n$ edges of $G_{k+1,q}$. Hence, Γ_{k+1} is distributed as a random $(k+1)$-regular bipartite (multi)graph (where multiple edges are allowed but not loops). Indeed, it fits the bipartite configuration model of Bollobás [4]. Each vertex could be replaced by a distinct set of $(k+1)$ points, each colour $c \in Q$ naturally corresponds to $(k+1)$ points associated with non-special copies of that colour, and the "dummy" vertex q corresponds to $(k+1)$ points associated with special copies of popular colours. Then, we randomly pair these points to get the colouring of the edges of $G_{k+1,q}$. Note that Γ_{k+1} contains no information about the actual vertex choices of edges in $G_{k+1,q}$, only their colour. Informally, we can think of each edge of Γ_{k+1} being associated with a box containing a random vertex from $[n]$. We do not need to open these boxes for what is next.

We will use the fact that Γ_{k+1} is contiguous to the sum of $(k+1)$ independent random perfect matchings—see the survey on random regular graphs by Wormald [18]. If we delete one of these matchings, then we obtain a random k-regular bipartite graph contiguous to Γ_k. Arguing as before, we observe that Γ_k may be viewed as a random assignment of kn colours, including repetitions, to the kn edges of $G_{k,q}$. "Opening" the boxes on each edge of Γ_k gives us $G_{k,q}$, which by assumption has a required rainbow structure w.h.p. So, using the above coupling we conclude that $G_{k+1,q}$ also has a rainbow structure w.h.p.

3 Rainbow Spanning Trees

In view of the results in Sections 2.1 and 2.2, we will assume that $k = 2$ and $q = n - 1$ for this section.

To establish the existence of a **RST** to prove Theorem 1, we will use the result of Edmonds [7] on the matroid intersection problem. A finite *matroid M* is a pair $(E, \mathcal{I})$, where E is a finite set (called the *ground set*) and $\mathcal{I}$ is a family of subsets of E (called the *independent sets*) with the following properties:

- $\emptyset \in \mathcal{I}$,

- for each $A' \subseteq A \subseteq E$, if $A \in \mathcal{I}$, then $A' \in \mathcal{I}$ (*hereditary property*),

- if A and B are two independent sets of $\mathcal{I}$ and A has more elements than B, then there exists an element in A that when added to B gives a larger independent set than B (*augmentation property*).

A maximal independent set (that is, an independent set which becomes dependent on adding any element of E) is called a *basis* for the matroid. An observation, directly analogous to the one of bases in linear algebra, is that any two bases of a matroid M have the same number of elements. This number is called the *rank* of M. For more details on matroids see, for example, [16].

In this scenario, M_1, M_2 are matroids over a common ground set E with rank functions r_1, r_2, respectively. Edmonds' general theorem shows that

$$\max(|I| : I \text{ is independent in both matroids}) = \min_{\substack{E_1 \cup E_2 = E \\ E_1 \cap E_2 = \emptyset}} (r_1(E_1) + r_2(E_2)), \qquad (1)$$

where $r_i(E_i)$ is the rank of the matroid induced by E_i.

In our application, the common ground set E is the set of coloured multi-edges of $G_{k,q}$. M_1 is the cycle matroid of the graph G_{k-out}; that is, $S \subseteq E$ is independent in M_1 if S induces a graph with no cycle (colours are ignored, two parallel edges are considered to be a cycle of length 2). Hence, for every $S \subseteq E$ we have $r_1(S) = n - \kappa(S)$, where $\kappa(S)$ is the number of components of the graph $G_S = ([n], S)$ induced by S. M_2 is the partition matroid associated with the colours; that is, $S \subseteq E$ is independent in M_2 if S has no two edges in the same colour. This time, for every $S \subseteq E$ we have that $r_2(S)$ is the number of distinct colours occurring in S. We use Edmonds' theorem to get the following useful observation that has been used a number of times in related contexts. In temporal order, it was used by Fenner and Frieze [9], Frieze and McKay [14] and by Suzuki [17].

Lemma 2. *Let $G = (V, E)$ be a multigraph in which each edge is coloured with a colour from a set Q. A necessary and sufficient condition for the existence of a **RST** is that*

$$\kappa(C_I) \le |Q| + 1 - |I| \qquad \qquad \text{for all } I \subseteq Q, \qquad (2)$$

where $C_I \subseteq E$ is the set of edges of colour from set I.

Proof. Clearly, G has a rainbow spanning tree if and only if G contains a set S of coloured edges of size $|V| - 1$ such that S is independent both in M_1 (S induces a spanning tree) and in M_2 (S is rainbow). Since no set of size at least $|V|$ is independent in M_1, the necessary and sufficient condition is that the right side of (1) is at least $|V| - 1$. Hence, the desired condition is that for every partition of the edge set E into E_1 and E_2 we have $r_1(E_1) + r_2(E_2) \ge |V| - 1$.

Let us fix a partition of E into E_1 and E_2. Let $J \subseteq Q$ be the set of colours occurring in E_2, $E_2' \subseteq E$ be the set of edges coloured with a colour from J, and $E_1' = E \setminus E_2'$. Clearly, (E_1', E_2') is also a partition of E, $E_2 \subseteq E_2'$ and so $E_1' \subseteq E_1$, and $r_2(E_2) = r_2(E_2') = |J|$. Moreover, since $E_1' \subseteq E_1$, $r_1(E_1') \le r_1(E_1)$ and so

$$r_1(E_1) + r_2(E_2) \ge r_1(E_1') + r_2(E_2').$$

Therefore, without loss of generality, we may restrict ourselves to sets E_2 containing all edges of colour from some set $J \subseteq Q$ and then take $I = Q \setminus J$ (that is, $E_2 = C_J$ and $E_1 = C_I$). The condition to verify is the following:

$$|V| - 1 \le r_1(E_1) + r_2(E_2) = (|V| - \kappa(C_I)) + (|Q| - |I|)$$

which is equivalent to (2). The proof of the lemma is finished. $\qquad \square$

Recall that $k = 2$, $q = n - 1$, and so $\rho = k = 2$. For a given $\ell \in [q] \cup \{0\}$, we define the event

$$\mathcal{A}_\ell = \{\exists I \subseteq Q, |I| = \ell : \kappa(C_I) \geq q - |I| + 2\}.$$

Trivially, $\mathcal{A}_0$ cannot occur. Since each colour of Q is used at least once, $\mathcal{A}_1$ cannot occur as well: if $|I| = 1$, then $\kappa(C_I) \leq n - 1$ but $q - |I| + 2 \geq n$. With some additional work we can eliminate $\mathcal{A}_2$ and $\mathcal{A}_3$ as well. Note that in G_{k-out}, the probability that a vertex v chooses u as a neighbour and vice versa is $O(1/n^2)$. Hence by linearity of expectation, the expected number of multiple edges in G_{k-out} is $O(1)$. The probability that both edges in a multiple edge receive the same (fixed) colour in $G_{k,q}$ is $O(1/q^2)$ and so the expected number of monochromatic multiple edges in $G_{k,q}$ is $O(1/q) = o(1)$. It follows from the first moment method that w.h.p. there are no monochromatic multiple edges. Since we aim for a statement that holds w.h.p., we may assume that this property is satisfied. We get that if $|I| = 2$, then $\kappa(C_I) \leq n - 2$ but $q - |I| + 2 \geq n - 1$. Similarly, the expected number of triples of colours such that there are two multiple edges that use only these colours is $O(1/q) = o(1)$. Hence, we may assume that for any I of size 3, $|C_I| \geq 5$ and so $\kappa(C_I) \leq n - 3$ whereas $q - |I| + 2 \geq n - 2$. Finally, note that $\mathcal{A}_q$ cannot occur w.h.p. since G_{k-out} is connected w.h.p.

We know that if there is no **RST**, then $\mathcal{A}_\ell$ occurs for some $\ell \in [4, q - 1]$. Let us concentrate on a minimal ℓ, corresponding set I, and let $S = C_I$. Let us start with the following simple but useful observation.

Claim 1. G_S *has no bridges.*

Proof. If there is a bridge in G_S, then we simply remove it and all edges of the same colour. The number of components increases by at least one and the number of colours decreases by one. Clearly, $\mathcal{A}_{\ell-1}$ occurs, contradicting the minimality of ℓ. $\qquad\square$

Recall that $\kappa(S) \geq q - \ell + 2 = n - \ell + 1$. Suppose then that G_S has i isolated vertices and $n - \ell + x - i$ non-trivial components for some integer $x \geq 1$. Let T be the set of vertices in the non-trivial components of G_S. By Claim 1, G_S has no bridges. Since non-trivial components without bridges have at least three vertices,

$$i + 3(n - \ell + x - i) \leq n \qquad (3)$$

which gives us that

$$|T| = n - i \leq \frac{3}{2}(\ell - x) \leq \frac{3}{2}(\ell - 1).$$

Let $\mathcal{B}_\ell$ denote the event

$$\{\exists I \subseteq Q, |I| = \ell, T \subseteq [n] : \quad t = |T| \leq 3(\ell - 1)/2,$$
$$\text{all edges coloured with } I \text{ are contained in } T,$$
$$\text{there are } u \geq \max\{\rho\ell, t\} \ I\text{-coloured edges}\}.$$

(Here $u \geq t$ because we are dealing with bridgeless components and $u \geq \rho\ell$ as each colour appears at least ρ times.) Clearly,

$$\mathcal{A}_\ell \subseteq \mathcal{B}_\ell \qquad \text{for } \ell \geq 4, \qquad (4)$$

and we will bound the probability of $\mathcal{B}_\ell$ to deal with small values of ℓ, that is, for $\ell \leq n/20$.

3.1 $4 \le \ell \le n/20$

Recall that $k = 2$, $q = n - 1$, and $\rho = \lfloor kn/q \rfloor = \lfloor kn/(n-1) \rfloor = 2$. There are $q_1 = q - k = n - 1 - k = n - 3$ unpopular colours that appear $k = 2$ times and $q_2 = k = 2$ popular colours that appear $k + 1 = 3$ times.

Note that

$$\mathbb{P}(\mathcal{B}_\ell) \le \sum_{t=1}^{3(\ell-1)/2} \binom{n}{t} \sum_{\ell_1 + \ell_2 = \ell} \binom{q_1}{\ell_1} \binom{q_2}{\ell_2} \binom{tk}{k\ell + \ell_2}$$
$$\times \left(\frac{t}{n} \right)^{k\ell + \ell_2} \frac{(k\ell + \ell_2)!}{(kn)(kn - 1) \cdots (kn - (k\ell + \ell_2 - 1))},$$

where the second sum is taken over all integers $0 \le \ell_1 \le q_1$, $0 \le \ell_2 \le q_2$ such that $\ell_1 + \ell_2 = \ell$. Indeed, in order to estimate $\mathbb{P}(\mathcal{B}_\ell)$, we consider all possibilities for $t = |T| \le 3(\ell-1)/2$ (the first sum) and all sets of size t (the term $\binom{n}{t}$). We independently consider sets of colours $I \subseteq Q$ with ℓ_1 unpopular colours and ℓ_2 popular ones (the second sum). Then, we need to select specific colours (the term $\binom{q_1}{\ell_1}\binom{q_2}{\ell_2}$). Since all edges coloured with I are contained in T, we need to select which of the tk edges of G_{k-out} generated by vertices of T are coloured with I (the term $\binom{tk}{k\ell + \ell_2}$). The selected edges need to stay within T (the term $\left(\frac{t}{n}\right)^{k\ell + \ell_2}$) and be coloured with I (the last term).

Clearly, $\binom{tk}{k\ell + \ell_2} \le (tk)^{k\ell + \ell_2}/(k\ell + \ell_2)!$. Moreover, since $\ell \le n/20$ and $\ell_2 \le q_2 = k$,

$$
\begin{aligned}
(kn)(kn - 1) \cdots (kn - (k\ell + \ell_2 - 1)) &\ge (kn)^{k\ell + \ell_2} \left(1 - \frac{k\ell + \ell_2}{kn} \right)^{k\ell + \ell_2} \\
&\ge (kn)^{k\ell + \ell_2} \left(\frac{19}{20} + o(1) \right)^{k\ell + \ell_2} \\
&= (kn)^{k\ell + \ell_2} \left(\sqrt{\frac{20}{19}} + o(1) \right)^{-2(k\ell + \ell_2)} \\
&\ge (kn)^{k\ell + \ell_2} \, 1.03^{-2(k\ell + \ell_2)}.
\end{aligned}
$$

We get that

$$
\begin{aligned}
\mathbb{P}(\mathcal{B}_\ell) &\le \sum_{t=1}^{3(\ell-1)/2} \binom{n}{t} \sum_{\ell_1 + \ell_2 = \ell} \binom{q_1}{\ell_1} \binom{q_2}{\ell_2} (tk)^{k\ell + \ell_2} \left(\frac{t}{n} \right)^{k\ell + \ell_2} \frac{1.03^{2(k\ell + \ell_2)}}{(kn)^{k\ell + \ell_2}} \\
&= \sum_{t=1}^{3(\ell-1)/2} \binom{n}{t} \sum_{\ell_1 + \ell_2 = \ell} \binom{q_1}{\ell_1} \binom{q_2}{\ell_2} \left(\frac{1.03 t}{n} \right)^{2(k\ell + \ell_2)}.
\end{aligned}
$$

Since $\binom{q_1}{\ell_1}\binom{q_2}{\ell_2} \le \binom{q}{\ell} \le \binom{n}{\ell} \le (ne/\ell)^\ell$, $\binom{n}{t} \le (ne/t)^t$, and $2(k\ell + \ell_2) \ge 2k\ell = 4\ell$, we get

$$\mathbb{P}(\mathcal{B}_\ell) \le (\ell + 1) \sum_{t=1}^{3(\ell-1)/2} \left(\frac{ne}{t} \right)^t \left(\frac{ne}{\ell} \right)^\ell \left(\frac{1.03 t}{n} \right)^{4\ell}.$$

Note that since the ratio between the $(t+1)$-st and the t-th term is

$$\frac{ne}{t+1}\left(\frac{t+1}{t}\right)^{4\ell-t} \geq \frac{ne}{t+1} \geq 2,$$

the above sum is of the order of its last term. We get that

$$\begin{aligned}
\mathbb{P}(\mathcal{B}_\ell) &= O\left(\ell\left(\frac{ne}{3\ell/2}\right)^{3\ell/2}\left(\frac{ne}{\ell}\right)^{\ell}\left(\frac{1.03 \cdot 3\ell/2}{n}\right)^{4\ell}\right) \\
&= O\left(\ell\left(\frac{2}{3}\cdot e^{5/3}\cdot 1.545^{8/3}\,\frac{\ell}{n}\right)^{3\ell/2}\right) \\
&= O\left(\ell\left(\frac{12\,\ell}{n}\right)^{3\ell/2}\right).
\end{aligned}$$

Clearly, $\sum_{\ell=4}^{n/20}\mathbb{P}(\mathcal{B}_\ell) = O(n^{-6}) = o(1)$, since the sum is dominated by its first term.

3.2 $n/20 < \ell \leq n - \frac{5001n\log\log n}{\log n}$

We first bound the number of pairwise vertex disjoint cycles in G_{k-out}, including loops (cycles of length 1) and parallel edges (cycles of length 2). In our application, we concentrate on $k = 2$ but the following bound holds in general.

Lemma 3. *Fix any integer $k \geq 2$. W.h.p. no family of pairwise vertex disjoint cycles in G_{k-out} consists of more than $3n\log(2k)/\log n$ cycles.*

Proof. Let $\omega = \frac{\log n}{2\log(2k)}$. Let Z denote the number of cycles of length at most ω in G_{k-out} (possibly overlapping). There are $\binom{n}{s}\frac{s!}{2s} = \binom{n}{s}\frac{(s-1)!}{2}$ cycles of length s that one can form out of n labelled vertices. For a given pair of vertices, the probability that there is an edge between them is clearly at most $\frac{2k}{n}$. Hence,

$$\mathbb{E}(Z) \leq \sum_{s=1}^{\omega}\binom{n}{s}\frac{(s-1)!}{2}\left(\frac{2k}{n}\right)^s \leq \sum_{s=1}^{\omega}\frac{(2k)^s}{2s}.$$

Since the ratio between the two consecutive terms $((s+1)$st and sth$)$ is $(2k)s/(s+1) \geq k \geq 2$,

$$\mathbb{E}(Z) \leq \frac{(2k)^\omega}{\omega} = o\left(\exp\left(\frac{\log n}{2\log(2k)}\log(2k)\right)\right) = o(\sqrt{n}).$$

The Markov inequality implies that $Z \leq n\log(2k)/\log n$ w.h.p.

On the other hand, there are trivially at most $n/\omega = 2n\log(2k)/\log n$ pairwise vertex disjoint cycles of length greater than ω, and the lemma follows. $\qquad\square$

For the next property we need, we assume that $k = 2$. As in Section 2.2, let Γ_2 be the bipartite (multi)graph with vertex sets $[n]$ and $Q' = Q \cup \{q\}$, where q is a "dummy" vertex associated with special copies of popular colours. There is an edge vc in Γ_2 if one of v's *choices* has non-special copy of colour c; as before, an edge vq occurs in Γ_2 if one of v's choices is one of the two special copies of popular colours. Γ_2 is a random 2-regular bipartite graph. Any 2-regular graph is a collection of cycles. We will use a well-known and easy to prove fact that w.h.p. there are not too many cycles in Γ_2. For completeness, we provide an elementary proof.

Lemma 4. Γ_2 *contains at most* $(\log n)^2$ *cycles w.h.p.*

Proof. Let Z denote the number of cycles in Γ_2. We will use the configuration model of Bollobás [4] to estimate $\mathbb{E}(Z)$; see also Chapter 11 of Frieze and Karoński [10]. Let us select any vertex v and any of the two points in it. We expose the other endpoint of that edge associated with vertex y, and move on to the other point associated with y. We continue the process until the first cycle is discovered. Then, we select any other point that is not matched yet and continue from there. The probability of closing a cycle at ith step of this process is precisely $1/(2n - 2i + 1)$. Indeed, $2(i - 1)$ points are matched before step i and one point is considered at ith step, so there are $2n - 2i + 1$ points available to be matched with the considered point to form an edge; only one of these points (namely, the one associated with vertex v we start with) closes the cycle. We get that

$$\mathbb{E}(Z) = \sum_{i=1}^{n} \frac{1}{2n - 2i + 1} = \sum_{j=1}^{2n} \frac{1}{j} - \sum_{j=1}^{n} \frac{1}{2j} = \ln(2n) - \frac{1}{2}\ln(n) + O(1) = \frac{1}{2}\ln n + O(1).$$

Showing concentration of Z around its expectation is easy but, since we aim for a slightly weaker result, we may trivially use the Markov inequality to get the desired bound that holds w.h.p. $\qquad\square$

We continue concentrating on a minimal ℓ (in the range for ℓ considered in this section) for which $\mathcal{A}_\ell$ occurs and the corresponding set $I \subseteq Q$. We let $S = C_I$ and we expose $G_S = ([n], S)$, the sub-graph induced by S. In particular, $\kappa(S) \geq n - \ell + 1$ and we may assume that G_S is bridgeless by Claim 1. Let X_0 denote the number of isolated vertices of G_S. Since G_S is bridgeless, each non-trivial component has a cycle. Hence, the number of non-trivial components, η, is bounded by the number of pairwise disjoint cycles. By Lemma 3, we may assume that $\eta \leq \frac{3n \log(2k)}{\log n} \leq \frac{5n}{\log n}$. It follows that

$$n - \ell + 1 \leq \kappa(S) = X_0 + \eta \text{ where } \eta \leq \frac{5n}{\log n}. \tag{5}$$

Almost all sets of colours I yield a graph G_S with many isolated vertices ensuring that (5) does not hold. Only a small fraction of possible sets of colours need special attention. We will use Γ_2 to estimate how many different configurations we need to take care of.

Let us concentrate on the set of edges incident with colours I in Γ_2, ignoring the two edges incident with a "dummy" vertex q. (Note that removing at most two edges from G_S may only increase the number of isolated vertices, so $X_0 \leq X_0'$, where X_0' is the number of isolated

9

vertices in $G_{S'}$; S' is the set obtained from S after removing the edges incident with special copies of popular colours (there are at most two such edges).) By Lemma 4, this set of edges in Γ_2 induces a graph Γ consisting of $r \leq (\log n)^2$ cycles and s paths. The $2s$ endpoints of paths in Γ must be in $[n]$ and so each path corresponds to two distinct non-isolated vertices of $G_{S'}$. Let x be the number of vertices in $[n]$ in Γ that are of degree 2 (vertices on cycles or internal vertices on paths); they also correspond to non-isolated vertices of $G_{S'}$. The remaining vertices in Γ are isolated and the corresponding vertices in $G_{S'}$ *might* be isolated. Note that we did not expose edges yet only assigned colours. It is possible that after exposing edges such vertices will be chosen by other ones and become non-isolated. In any case, $X_0' \leq n - 2s - x$. By comparing the number of edges in Γ incident to $[n]$ with the number of edges incident to $I \subseteq Q$, we get $2 \cdot x + 1 \cdot (2s) = 2\ell$ so $x = \ell - s$. Combining the two observations together, we get $X_0' \leq n - 2s - x = n - \ell - s$.

If $s \geq \frac{5n}{\log n}$, then

$$\kappa(G_S) = X_0 + \eta \leq X_0' + \eta \leq n - \ell - s + \eta \leq n - \ell - \frac{5n}{\log n} + \eta \leq n - \ell,$$

contradicting (5).

Let us assume then that $s < \frac{5n}{\log n}$. Let $V_0 \subseteq [n]$ denote the set of vertices in Γ that are covered by paths and cycles, and let $V_1 = [n] \setminus V_0$. As argued above, $|V_1| = n - \ell - s$ and so $|V_1| \geq \frac{5000n \log\log n}{\log n}$ since $\ell \leq n - \frac{5001n \log\log n}{\log n}$ and $s < \frac{5n}{\log n} < \frac{n \log\log n}{\log n}$. Vertices in V_0 correspond to non-isolated vertices in $G_{S'}$. Vertices in $G_{S'}$ corresponding to vertices in V_1 that are isolated in Γ do not generate any random edges but it does not mean that they are isolated in $G_{S'}$. Let $V_2 \subseteq V_1$ denote the set of vertices that are incident with an I-coloured edge in Γ and are *chosen by some vertex in V_0*. It is important to notice that we have not conditioned on the other endpoints of the choices in Γ (endpoints of random edges generated by vertices in $G_{S'}$ corresponding to vertices in V_0 in Γ). We may expose these $2\ell \geq n/10$ edges now, one at a time, each time updating set V_2. Provided that $|V_2| \leq |V_1|/50$, we add a new vertex to V_2 with probability at least $(49|V_1|/50)/n \geq |V_1|/(2n)$. Let $X \in \mathrm{Bin}\left(n/10, |V_1|/(2n)\right)$ with $\mathbb{E}(X) = |V_1|/20 \geq \frac{250n \log\log n}{\log n}$. The established coupling implies that

$$\mathbb{P}\left(|V_2| \leq \frac{|V_1|}{50}\right) \leq \mathbb{P}\left(X \leq \frac{|V_1|}{50}\right) \leq \mathbb{P}\left(X \leq \frac{\mathbb{E}(X)}{2}\right),$$

and so the Chernoff's bounds imply that

$$\mathbb{P}\left(|V_2| \leq \frac{100n \log\log n}{\log n}\right) \leq \exp\left(-\frac{\mathbb{E}(X)}{12}\right) \leq \exp\left(-\frac{20n \log\log n}{\log n}\right).$$

Now, we need to estimate the number of configurations we need to investigate. After orienting (arbitrarily) cycles in Γ_2, there are at most $\binom{n}{s}$ choices for the beginnings and at most $\binom{n}{s}$ choices for the endings of paths. Such choices yield paths in Γ. By Lemma 4, there are at most $2^{(\log n)^2}$ choices to determine which cycles from Γ_2 should stay in Γ. Hence, the number of configurations (different bipartite graphs Γ) we need to deal with is at most

$$\sum_{s < 5n/\log n} \binom{n}{s}^2 2^{(\log n)^2} \leq 2 \cdot \binom{n}{5n/\log n}^2 2^{(\log n)^2} \leq 2 \cdot \left(\frac{en}{5n/\log n}\right)^{10n/\log n} 2^{(\log n)^2}$$

10

$$\leq \quad \exp\left(\frac{10n\log\log n}{\log n} + O((\log n)^2)\right) \leq \exp\left(\frac{15n\log\log n}{\log n}\right).$$

Comparing it with an upper bound for the failure probability for each configuration, we get that if $s < \frac{5n}{\log n}$ and $|V_1| \geq \frac{5000n\log\log n}{\log n}$, then w.h.p. $|V_2| > \frac{100n\log\log n}{\log n}$. Since $X_0' \leq n - \ell - s - |V_2| \leq n - \ell - |V_2|$, we get that

$$\kappa(G_S) \leq X_0' + \eta \leq n - \ell - \frac{100n\log\log n}{\log n} + \eta < n - \ell,$$

contradicting (5).

3.3 $\quad n - \frac{5001n\log\log n}{\log n} < \ell \leq n - 1$

The two special copies of the two popular colours can give us some unnecessary technical problems in the following calculations. So let us delete the two edges, e_1, e_2, associated with special copies so that each colour is used exactly twice. The graph G^* obtained this way has $2n-2$ edges. Deleting edges can only increase the number of connected components so it suffices to prove that this quantity is small enough to satisfy (2) after the deletion. It is straightforward to show that G^* remains connected w.h.p. but we will not need this fact in our argument.

For the range of ℓ considered in this section, it is easier to concentrate on the largest set of colours $I \subseteq Q$ and the associated set of edges $S = C_I \cup \{e_1, e_2\}$ for which G_S has too many components i.e. $\kappa(G_S) \geq n - \ell + 1$, in violation of (2). Note that for every colour c not in I, the two edges of colour c in G^* join distinct components of G_S; otherwise, $I \cup \{c\}$ also yields a graph with too many components. This means that there are no edges of colour belonging to $Q \setminus I$ in G^* joining vertices in the same component of G_S. Put another way, let $\mathcal{C}_m$ be the event in $G_{2,q}$ that we can find m colours and the associated m unique pairs of edges M such that

(i) $M \cap \{e_1, e_2\} = \emptyset$,

(ii) each pair in M has the same colour (that is distinct from colours of other pairs in M),

(iii) $G_{\bar{M}} = G^* - M = G_{2-out} - (M \cup \{e_1, e_2\})$ has at least $m + 2$ components, and

(iv) no edge of M joins two vertices of the same component of G^*.

We have to show that $\mathcal{C}_m$ is unlikely, for $1 \leq m \leq m_0 := \frac{5001n\log\log n}{\log n}$.

Suppose that $V_1, V_2, \ldots, V_p$ are the components of $G_{\bar{M}}$. Let $n_i = |V_i|$. We will use δ_v for the number of choices of v in G_{2-out} outside its component in $G_{\bar{M}}$, and let $\Delta_i = \sum_{v \in V_i} \delta_v$. Note that the number of edges in component V_i is $2n_i - \Delta_i$. Hence,

$$\Delta_i \leq n_i + 1, \tag{6}$$

as otherwise there are not enough edges inside V_i to get connectivity. It follows that

$$\mathbb{P}(\mathcal{C}_m) \leq \binom{n-1}{m} \sum_{\substack{p=m+2}}^{2m+3} \sum_{\substack{n_1+\cdots+n_p=n \\ n_1 \geq n_2 \geq \cdots \geq n_p \geq 1}} \frac{1}{\Psi(n_1, \ldots, n_p)} \binom{n}{n_1, n_2, \ldots, n_p}$$

11

$$\times \sum_{\substack{\Delta_1+\cdots+\Delta_p=2m \\ \Delta_1,\ldots,\Delta_p\geq 0}} \prod_{i=1}^{p} \left(\frac{n_i}{n}\right)^{2n_i-\Delta_i} \left(1-\frac{n_i}{n}\right)^{\Delta_i} \binom{2n_i}{\Delta_i} \frac{1}{\binom{2n}{2m}},$$

where

$$\Psi(n_1,\ldots,n_p) = \prod_{i=1}^{n} \ell_i! \quad \text{with} \quad \ell_i = |\{j \in [p] : n_j = i\}|.$$

Indeed, we first need to choose the colours M which can be done in $\binom{n-1}{m}$ ways. There are at least $m+2$ components in $G_{\bar{M}}$ but, since we removed exactly $2m+2$ edges from G_{2-out} to get $G_{\bar{M}}$, the number of them is at most $2m+3$. We then need to choose the component sizes and the components in $\sum_{\substack{n_1+\cdots+n_p=n \\ n_1\geq n_2\geq\cdots\geq n_p\geq 1}} \binom{n}{n_1,n_2,\ldots,n_p}/\Psi(n_1,\ldots,n_p)$ ways. Function $\Psi(n_1,\ldots,n_p)$ removes an implicit ordering of the components in the multinomial coefficient. We then consider all possibilities for the number of M coloured edges (the $2m$ edges present in G^*) leaving each component in $\sum_{\substack{\Delta_1+\cdots+\Delta_p=2m \\ \Delta_1,\ldots,\Delta_p\geq 0}}$ ways. The factor $\binom{2n_i}{\Delta_i}$ accounts for choosing which of the $2n_i$ edges generated by vertices in V_i have colour in M and are not one of the two special edges, e_1, e_2. The factor n_i/n (respectively, $1 - n_i/n$) is the probability that the edge choice of a vertex in V_i is in V_i (respectively, not in V_i). Finally, we need to make sure that the $2m$ edges we identified received precisely the $2m$ non-special copies of the m colours we selected. This happens with probability $1/\binom{2n}{2m}$ as any set of $2m$ colours from the set of $2n$ available colours (including repetitions and including special copies) is assigned to these edges with uniform probability; only one of them has colours that are exactly the ones we selected at the very beginning (note that special colours were excluded then).

Continuing,

$$\mathbb{P}(\mathcal{C}_m) \leq \frac{\binom{n}{m}}{\binom{2n}{2m}} \sum_{p=m+2}^{2m+3} \sum_{\substack{n_1+\cdots+n_p=n \\ n_1\geq n_2\geq\cdots\geq n_p\geq 1}} \frac{1}{\Psi(n_1,\ldots,n_p)} \binom{n}{n_1,n_2,\ldots,n_p}$$

$$\times \sum_{\Delta_1+\cdots+\Delta_p=2m} \prod_{i=1}^{p} \left(\frac{n_i}{n}\right)^{2n_i-\Delta_i} \left(1-\frac{n_i}{n}\right)^{\Delta_i} \binom{2n_i}{\Delta_i}$$

$$\leq \frac{\binom{n}{m}}{\binom{2n}{2m}} \sum_{p=m+2}^{2m+3} \sum_{\substack{n_1+\cdots+n_p=n \\ n_1\geq n_2\geq\cdots\geq n_p\geq 1}} \frac{1}{\Psi(n_1,\ldots,n_p)} \binom{n}{n_1,n_2,\ldots,n_p}$$

$$\times \sum_{\Delta_1+\cdots+\Delta_p=2m} \prod_{i=1}^{p} \left(\frac{n_i}{n}\right)^{2n_i-\Delta_i} \binom{2n_i}{\Delta_i}$$

$$\leq \frac{\binom{n}{m}}{\binom{2n}{2m}} \frac{n!}{n^{2n-2m}} \sum_{\substack{p\geq m+2 \\ n_1+\cdots+n_p=n \\ n_1\geq n_2\geq\cdots\geq n_p\geq 1 \\ \Delta_1+\cdots+\Delta_p=2m}} \frac{1}{\Psi(n_1,\ldots,n_p)} \prod_{i=1}^{p} \frac{n_i^{2n_i-\Delta_i}}{n_i!} \binom{2n_i}{\Delta_i}. \tag{7}$$

Let us prove the following simple structural property of G_{2-out}. It will imply that the largest component (of size n_1) has size asymptotic to n.

12

Lemma 5. *For $S \subseteq [n]$, let $e^+(S)$ denote the number of choices by vertices in S that are not in S, and let $e(S, [n] \setminus S) = e^+(S) + e^+([n] \setminus S)$ denote the number of edges in G_{2-out} that are between S and its complement. Then, w.h.p. the following property holds for any m such that $1 \leq m \leq m_0 := \frac{5001 n \log \log n}{\log n}$:*

$$\text{for all } S \subseteq [n], 9m \leq |S| \leq n/2, \text{ we have } e(S, [n] \setminus S) > 2m + 2.$$

Proof. We will independently deal with small and large sets by proving the following statement. For any m such that $1 \leq m \leq m_0 := \frac{5001 n \log \log n}{\log n} = o(n)$, the following two properties hold with probability $1 - O(n^{-1})$:

(a)
$$\text{for all } S \subseteq [n], 9m \leq |S| = s \leq n/200, \text{ we have } e^+(S) \geq s/2 > 2m + 2. \tag{8}$$

(b)
$$\text{for all } S \subseteq [n], n/200 \leq |S| = s \leq n/2, \text{ we have } e(S, [n] \setminus S) > 2m + 2. \tag{9}$$

Note that, since $16e^2 < 200$,

$$\mathbb{P}(\neg(8)) \leq \sum_{s=9m}^{n/200} \binom{n}{s} \binom{2s}{3s/2} \left(\frac{s}{n}\right)^{3s/2} \leq \sum_{s=9m}^{n/200} \left(\frac{en}{s}\right)^s 2^{2s} \left(\frac{s}{n}\right)^{3s/2}$$

$$\leq \sum_{s=9m}^{n/200} \left(\frac{s}{n} \cdot 16e^2\right)^{s/2} = O(n^{-1}),$$

so property (a) holds.

To see that property (b) holds too, note that

$$\mathbb{P}(\neg(9)) \leq \sum_{s=n/200}^{n/2} \sum_{i=0}^{2m+2} \sum_{j=0}^{i} \binom{n}{s} \binom{2s}{j} \left(\frac{s}{n}\right)^{2s-j} \left(1 - \frac{s}{n}\right)^j \binom{2(n-s)}{i-j} \left(\frac{s}{n}\right)^{i-j} \left(1 - \frac{s}{n}\right)^{2(n-s)-i+j}$$

$$= \sum_{s=n/200}^{n/2} \sum_{i=0}^{2m+2} \sum_{j=0}^{i} \binom{n}{s} \binom{2s}{j} \binom{2(n-s)}{i-j} \left(\frac{s}{n}\right)^{2s+i-2j} \left(1 - \frac{s}{n}\right)^{2(n-s)-i+2j}$$

$$= O(m^2) \sum_{s=\sigma n=n/200}^{n/2} \left(\frac{1}{\sigma^\sigma (1-\sigma)^{1-\sigma}}\right)^n \binom{2n}{2m+2}^2 \sigma^{2\sigma n - (2m+2)} (1-\sigma)^{2(1-\sigma)n - (2m+2)}$$

$$= O(m^2) \sum_{s=\sigma n=n/200}^{n/2} (\sigma^\sigma (1-\sigma)^{1-\sigma})^n \binom{2n}{2m+2}^2 (\sigma(1-\sigma))^{-(2m+2)}$$

$$= O(m^2 n) \, c^n \exp\left(O\left(\frac{n(\log \log n)^2}{\log n}\right)\right) = c^n e^{o(n)} = O(n^{-1}),$$

where $c = (1/200)^{1/200} (199/200)^{199/200} < 0.97$. (In the above computation, i corresponds to $e(S, [n] \setminus S)$ and j corresponds to $e^+(S)$.) $\quad\square$

13

The lemma implies that we may assume that

$$n_1 \geq n - 9m \text{ in (7);} \tag{10}$$

otherwise, the number of edges joining distinct components in $G_{\bar{M}}$ would be greater than $2m+2$. Hence, we may rewrite (7) as follows:

$$\mathbb{P}(\mathcal{C}_m) \leq \frac{\binom{n}{m}}{\binom{2n}{2m}} \frac{n!}{n^{2n-2m}} \sum_{p=m+2}^{2m+3} \sum_{s=p-1}^{9m} \sum_{\substack{n_1+\cdots+n_p=n \\ n-s=n_1 \geq n_2 \geq \cdots \geq n_p \geq 1 \\ \Delta_1+\cdots+\Delta_p=2m}} \frac{1}{\Psi(n_1,\ldots,n_p)} \prod_{i=1}^{p} \frac{n_i^{2n_i-\Delta_i}}{n_i!} \binom{2n_i}{\Delta_i}.$$

Define

$$f(a,x) = a^{2a-x} \binom{2a}{x}.$$

Note that if $x \geq 1$, then

$$\frac{f(a,x)}{f(a,x-1)} = \frac{2a-x+1}{ax} \leq \frac{2}{x},$$

and so $f(a,x) \leq \frac{2^x}{x!} f(a,0)$. Using this observation, we get that

$$\sum_{\Delta_1+\cdots+\Delta_p=2m} \prod_{i=1}^{p} n_i^{2n_i-\Delta_i} \binom{2n_i}{\Delta_i}$$

$$\leq \sum_{\Delta_1,\ldots,\Delta_p \leq 2m} \prod_{i=1}^{p} n_i^{2n_i-\Delta_i} \binom{2n_i}{\Delta_i}$$

$$= \sum_{\Delta_1,\ldots,\Delta_{p-1} \leq 2m} \left(\prod_{i=1}^{p-1} n_i^{2n_i-\Delta_i} \binom{2n_i}{\Delta_i} \right) \sum_{\Delta_p \leq 2m} n_p^{2n_p-\Delta_p} \binom{2n_p}{\Delta_p}$$

$$\leq \sum_{\Delta_1,\ldots,\Delta_{p-1} \leq 2m} \left(\prod_{i=1}^{p-1} n_i^{2n_i-\Delta_i} \binom{2n_i}{\Delta_i} \right) n_p^{2n_p-0} \binom{2n_i}{0} \left(1 + \frac{2}{1} + \frac{2}{1} \cdot \frac{2}{2} + \frac{2}{1} \cdot \frac{2}{2} \cdot \frac{2}{3} + \cdots \right)$$

$$\leq \sum_{\Delta_1,\ldots,\Delta_{p-1} \leq 2m} \left(\prod_{i=1}^{p-1} n_i^{2n_i-\Delta_i} \binom{2n_i}{\Delta_i} \right) n_p^{2n_p} \sum_{k \geq 0} \frac{2^k}{k!}$$

$$= e^2 n_p^{2n_p} \sum_{\Delta_1,\ldots,\Delta_{p-1} \leq 2m} \prod_{i=1}^{p-1} n_i^{2n_i-\Delta_i} \binom{2n_i}{\Delta_i}$$

$$\leq \cdots \leq (e^2)^p \prod_{i=1}^{p} n_i^{2n_i}.$$

Unfortunately, the constant

$$e^2 = \sum_{k \geq 0} \frac{2^k}{k!}$$

associated with the sum over all possible values of Δ_i is too large for the final argument to follow. Fortunately, any constant smaller than e^2 would work. We may squeeze a bit more by using the following observation. Since $\sum_{i=2}^{p} n_i = s \leq 9m$ (see (10)) and $p-1 \geq m+1$, there are at most $p/2$ values of n_i, $i \geq 2$, that are at least 18 ($n_1 \sim n$ certainly is at least

14

18); the remaining ones are at most 17. It is important to notice that the sequence of n_i's is non-increasing so we conclude that at least the last $p/2 - 1$ values of n_i are at most 17. As a result, since $\Delta_i \leq n_i + 1$ (see (6)), the corresponding values of Δ_i's are at most 18 (we will refer to them as small). The contribution from small Δ_i's is A, where

$$A = \sum_{k \leq 18} \frac{2^k}{k!} \leq e^2 - 10^{-12}.$$

We get that

$$\sum_{\Delta_1 + \cdots + \Delta_p = 2m} \prod_{i=1}^p n_i^{2n_i - \Delta_i} \binom{2n_i}{\Delta_i} \leq (e^2)^{p/2+1} A^{p/2-1} \prod_{i=1}^p n_i^{2n_i}$$

$$= \frac{e^2}{A} (e^2 A)^{p/2} \prod_{i=1}^p n_i^{2n_i} \leq 2B^p \prod_{i=1}^p n_i^{2n_i},$$

where

$$B = e\sqrt{A} \leq e^2 - 10^{-12}.$$

It follows that

$$\mathbb{P}(\mathcal{C}_m) \leq 2 \frac{\binom{n}{m}}{\binom{2n}{2m}} \frac{n!}{n^{2n-2m}} \sum_{p=m+2}^{2m+3} B^p \sum_{s=p-1}^{9m} \sum_{\substack{n_1 + \cdots + n_p = n \\ n-s=n_1 \geq n_2 \geq \cdots \geq n_p \geq 1}} \frac{1}{\Psi(n_1, \ldots, n_p)} \prod_{i=1}^p \frac{n_i^{2n_i}}{n_i!}.$$

Now, for a given sequence $n_1, \ldots, n_p$ and an integer i, $2 \leq i \leq p$, let us define

$$g(a, b) = \frac{a^{2a} b^{2b}}{a! b! \Psi(a, n_2, \ldots, n_{i-1}, b, n_{i+1}, \ldots, n_p)}$$

and suppose that $n \sim a \gg 9m \geq b > 1$. Then, since $(1 + 1/x)^x$ is an increasing function of x (tending to e but we do not need this fact) and

$$\frac{\Psi(a, n_2, \ldots, n_{i-1}, b, n_{i+1}, \ldots, n_p)}{\Psi(a+1, n_2, \ldots, n_{i-1}, b-1, n_{i+1}, \ldots, n_p)} \leq 2,$$

we get that

$$\frac{g(a, b)}{g(a+1, b-1)} \leq \frac{a^{2a}}{(a+1)^{2a+2}} \cdot \frac{b^{2b}}{(b-1)^{2b-2}} \cdot \frac{a+1}{b} \cdot 2$$

$$= \frac{\left(1 + \frac{1}{b-1}\right)^{2(b-1)}}{\left(1 + \frac{1}{a}\right)^{2a}} \cdot \frac{b}{a+1} \cdot 2 \leq \frac{2b}{a} \leq \frac{20m}{n}.$$

It implies that the terms corresponding to sequences of n_i's with larger values of n_1 (smaller values of s in our bound) are much larger. On the other hand, there are more sequences with smaller values of n_1 (larger values of s) to consider. However, since $n_2 + \cdots + n_p = s$ and $n_i \geq 1$, there are only $\binom{s-1}{p-2}$ choices for the sequence of n_i's to consider. Combining the two

observations together, we get that the term corresponding to $s = p - 1$ (and the associated unique sequence of n_i's) is a dominating term:

$$\mathbb{P}(\mathcal{C}_m) \le 2 \frac{\binom{n}{m}}{\binom{2n}{2m}} \frac{n!}{n^{2n-2m}} \sum_{p=m+2}^{2m+3} B^p \frac{(n-p+1)^{2(n-p+1)}}{(n-p+1)!} \frac{1}{\Psi(n-p+1,1,\ldots,1)}$$
$$\cdot \left(1 + \sum_{s=p}^{9m} \binom{s-1}{p-2} \left(\frac{20m}{n} \right)^{s-(p-1)} \right)$$
$$\le 3 \frac{\binom{n}{m}}{\binom{2n}{2m}} \frac{n!}{n^{2n-2m}} \sum_{p=m+2}^{2m+3} B^p \frac{(n-p+1)^{2(n-p+1)}}{(n-p+1)!(p-1)!}.$$

Note that the ratio between the $(p+1)$st term and the pth one is

$$B \cdot \frac{(n-p)^{2(n-p)}}{(n-p+1)^{2(n-p+1)}} \cdot \frac{(n-p+1)!}{(n-p)!} \cdot \frac{p-1}{p}$$
$$= B \cdot \left(1 - \frac{1}{n-p+1} \right)^{2(n-p+1)} \cdot \frac{n-p+1}{(n-p)^2} \cdot \frac{p-1}{p} = \Theta(1/n).$$

Hence,

$$\mathbb{P}(\mathcal{C}_m) \le 4 \frac{\binom{n}{m}}{\binom{2n}{2m}} \frac{n! B^{m+2}}{n^{2n-2m}} \frac{(n-m-1)^{2(n-m-1)}}{(n-m-1)!(m+1)!}. \tag{11}$$

If m is a constant, then $\mathbb{P}(\mathcal{C}_m) = O(1/n)$. In order to investigate larger values of m, note that the ratio between the $(m+1)$st term and the mth one is

$$\frac{\binom{n}{m+1}}{\binom{n}{m}} \cdot \frac{\binom{2n}{2m}}{\binom{2n}{2m+2}} \cdot Bn^2 \cdot \frac{(n-m-2)^{2(n-m-2)}}{(n-m-1)^{2(n-m-1)}} \cdot \frac{n-m-1}{m+2}$$
$$= \frac{n-m}{m+1} \cdot \frac{(2m+1)(2m+2)}{(2n-2m)(2n-2m-1)} \cdot Bn^2 \cdot \left(1 - \frac{1}{n-m-1} \right)^{2(n-m-1)}$$
$$\cdot \frac{n-m-1}{(n-m-2)^2} \cdot \frac{1}{m+2} \to (B)(e^{-2}) < 1 - 10^{-13} \qquad \text{as } m, n \to \infty.$$

Hence, if m is sufficiently large (say, $m \ge m'$) the ratio is at most, say, $1 - 10^{-14}$. Combining the two observations together, we get that

$$\sum_{1 \le m \le m_0} \mathbb{P}(\mathcal{C}_m) = \sum_{1 \le m < m'} \mathbb{P}(\mathcal{C}_m) + \sum_{m' \le m \le m_0} \mathbb{P}(\mathcal{C}_m) = O\Big(\mathbb{P}(\mathcal{C}_1) + \mathbb{P}(\mathcal{C}_{m'}) \Big) = O\left(\frac{1}{n} \right) = o(1).$$

That finishes the proof of Theorem 1.

4 Matchings and Hamilton Cycles

In this paper, we dealt with rainbow spanning trees proving the strongest possible result, both in terms of q, the number of colours, and k, the degree of the associated random graph. We leave investigating other rainbow structures for future research.

Recall that it was shown by Frieze [11] that G_{2-out} has a perfect matching w.h.p., and by Bohman and Frieze [3] that G_{3-out} is Hamiltonian w.h.p. Both results are sharp. Hence, based on our observation in Section 2.1, it is natural to investigate the following questions.

- What is the smallest value of q such that $G_{2,q}$ has a Rainbow Perfect Matching (**RPM**) w.h.p.? (Trivially, $q \geq n/2$ and $q \leq 2n$ as $G_{2,2n}$ is rainbow.)

- What is the smallest value of q such that $G_{3,q}$ has a Rainbow Hamilton Cycle (**RHC**) w.h.p.? (Trivially, $q \geq n$ and $q \leq 3n$ as $G_{3,3n}$ is rainbow.)

References

[1] D. Bal, P. Bennett, A. Frieze, and P. Prałat, *Power of k choices and rainbow spanning trees in random graphs*, Electronic Journal of Combinatorics **22(1)** (2015), #P1.29.

[2] D. Bal, P. Bennett, X. Pérez-Giménez, and P. Prałat, *Rainbow perfect matchings and Hamilton cycles in the random geometric graph*, Random Structures and Algorithms **51(4)** (2017), 587–606.

[3] T. Bohman and A.M. Frieze, *Hamilton cycles in 3-out*, Random Structures and Algorithms **35** (2009), 393–417.

[4] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled graphs, *European Journal on Combinatorics* 1 (1980) 311–316.

[5] B. Bollobás, *Random graphs*, Academic press, 1985.

[6] C. Cooper and A.M. Frieze, *Multi-coloured Hamilton cycles in random edge-coloured graphs*, Combinatorics, Probability and Computing **11** (2002), 129–134.

[7] J. Edmonds, *Submodular functions, matroids and certain polyhedra*, in Combinatorial Structures and their Applications, R.Guy et al, eds., Gordon and Breach, 1970, 69–87.

[8] T.I. Fenner and A.M. Frieze, *On the connectivity of random m-orientable graphs and digraphs*, Combinatorica **2** (1982), 347–359.

[9] T.I. Fenner and A.M. Frieze, *On the existence of polychromatic sets of edges in graphs and digraphs*, Progress in Graph Theory, Edited by J.A. Bondy and U.S.R. Murty, Academic Press, (1984) 219-232.

[10] A. Ferber and M. Krivelevich, *Rainbow Hamilton cycles in random graphs and hypergraphs*, Recent trends in combinatorics, IMA Volumes in Mathematics and its applications, A Beveridge et al, Eds., Springer 2016, 167–189.

[11] A.M. Frieze, *Maximum matchings in a class of random graphs*, Journal of Combinatorial Theory B **40** (1986), 196–212.

[12] A.M. Frieze and M. Karoński, Introduction to Random Graphs, Cambridge University Press, 2015.

[13] A.M. Frieze and P. Loh, *Rainbow Hamilton cycles in random graphs*, Random Structures and Algorithms **44** (2014), 328–354.

[14] A. Frieze and B.D. Mckay, *Multicoloured trees in random graphs*, Random Structures and Algorithms **5** (1994), 45–56.

[15] S. Janson and N. Wormald. *Rainbow Hamilton cycles in random regular graphs*, Random Structures Algorithms, **30(1-2)** (2007), 35–49.

[16] J. Oxley, Matroid Theory, Oxford: Oxford University Press, 1992.

[17] K. Suzuki, *A necessary and sufficient condition for the existence of a heterochromatic spanning tree in a graph*, Graphs and Combinatorics 22 (2006) 261-269.

[18] N.C. Wormald, Models of random regular graphs, *In J.D. Lamb and D.A. Preece, editors, Surveys in Combinatorics, 1999, volume 267 of London Mathematical Society Lecture Note Series, Cambridge University Press* (1999) 239-298.