RAINBOW ARBORESCENCE IN RANDOM DIGRAPHS

DEEPAK BAL, PATRICK BENNETT, COLIN COOPER, ALAN FRIEZE, AND PAWEŁ PRAŁAT

Abstract. We consider the Erdős-Rényi random directed graph process, which is a stochastic process that starts with n vertices and no edges, and at each step adds one new directed edge chosen uniformly at random from the set of missing edges. Let $\mathcal{D}(n, m)$ be a graph with m edges obtained after m steps of this process. Each edge e_i $(i = 1, 2, \ldots, m)$ of $\mathcal{D}(n, m)$ independently chooses a colour, taken uniformly at random from a given set of $n(1 + O(\log \log n / \log n)) = n(1 + o(1))$ colours. We stop the process prematurely at time M when the following two events hold: $\mathcal{D}(n, M)$ has at most one vertex that has in-degree zero and there are at least $n - 1$ distinct colours introduced ($M = n(n - 1)$ if at the time when all edges are present there are still less than $n - 1$ colours introduced; however, this does not happen asymptotically almost surely). The question addressed in this paper is whether $\mathcal{D}(n, M)$ has a rainbow arborescence (that is, a directed, rooted tree on n vertices in which all edges point away from the root and all the edges are different colours). Clearly, both properties are necessary for the desired tree to exist and we show that, asymptotically almost surely, the answer to this question is "yes".

1. Introduction and the main result

A set of edges S in an edge-coloured graph is said to be **rainbow coloured** if every edge in S has a different colour. There has recently been a deal of research on the existence of rainbow coloured objects in random graphs. The existence of rainbow coloured Hamilton cycles is studied in Cooper and Frieze [8] and in Frieze and Loh [10]. In the latter paper, it is shown that if $G_{n,m}$, a random graph on n vertices and $m = (\frac{1}{2} + o(1))n \log n$ edges, is randomly coloured with $n + o(n)$ colours, then a.a.s. there is a rainbow Hamilton cycle.[1] In a related paper, Bal and Frieze [3] showed that if $m = Kn \log n$ and if $G_{n,m}$ is randomly coloured with exactly n colours then a.a.s. there is a rainbow Hamilton cycle. Janson and Wormald [14] showed that if a random $2r$ regular graph $\mathcal{G}_{n,2r}$ on vertex set $[n]$ is randomly coloured so that each colour is used exactly r times then a.a.s. $\mathcal{G}_{n,2r}$ contains a rainbow Hamilton cycle. More recently, Ferber, Kronenberg, Mousset and Shikhelman [9] showed that if $np \gg \log n$ and the binomial random graph $G_{n,p}$ is randomly edge-coloured using $O(n)$ colours then a.a.s. $G_{n,p}$ contains $\Omega(np)$ edge disjoint rainbow Hamilton cycles.

Hamilton cycles are just one example of a well-studied spanning subgraph. Frieze and McKay [11] considered the following process. Let $e_1, e_2, \ldots, e_N$, $N = \binom{n}{2}$ be a random ordering of the edges of the complete graph K_n. Let $G_m = ([n], E_m = \{e_1, e_2, \ldots, e_m\})$. Suppose

---

[1]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 a probability space holds **asymptotically almost surely** (or **a.a.s.**) if the probability that it holds tends to 1 as n goes to infinity.

that the edges of K_n are randomly coloured from a set of $cn, c \geq 1$ colours. Then let

$$m_1 = \min\{m : G_m \text{ is connected}\}.$$
$$m_2 = \min\{m : E_m \text{ contains edges of } n - 1 \text{ different colours}\}.$$
$$m_3 = \min\{m : G_m \text{ contains a rainbow spanning tree}\}.$$

It is clear that $m_3 \geq \max\{m_1, m_2\}$ and [11] shows that a.a.s. $m_3 = \max\{m_1, m_2\}$. Bal, Bennett, Frieze and Pralat [2] proved a similar tight result for the case where each edge receives a choice of two random colours.

In this paper we turn our attention to a directed analogue of the result of [11]. We consider a **random digraph process**, which is a stochastic process that starts with n vertices and no edges, and at each step adds one new (directed) edge chosen uniformly at random from the set of missing edges. Formally, let $N = n(n-1)$ and let $e_1, e_2, \ldots, e_N$ be a random permutation of the edges of the complete digraph $\vec{K}_n$. The graph process consists of the sequence of random digraphs $(\mathcal{D}(n,m))_{m=0}^{N}$, where $\mathcal{D}(n,m) = (V, E_m)$, $V = [n] := \{1, 2, \ldots, n\}$, and $E_m = \{e_1, e_2, \ldots, e_m\}$. It is clear that $\mathcal{D}(n,m)$ is a digraph taken uniformly at random from the set of all digraphs on n vertices and m edges. (See, for example, [4, 13] for more details.)

Let

$$\varepsilon = \varepsilon(n) = \frac{\log \log n}{\log n}.$$

Suppose that each edge e_i $(i = 1, 2, \ldots, m)$ of $\mathcal{D}(n,m)$ is independently assigned a colour, uniformly at random, from a given set W of $(1 + 50\varepsilon)n = (1 + o(1))n$ colours.

We often write $\mathcal{D}(n,m)$ when we mean a graph drawn from the distribution $\mathcal{D}(n,m)$.

An **arborescence** is a spanning digraph in which, for a vertex u called the **root** and any other vertex v, there is exactly one directed path from u to v. Equivalently, an arborescence is a directed, rooted tree in which all edges point away from the root. Every arborescence is a directed acyclic graph (DAG), but not every DAG is an arborescence. A set of edges S is said to be **rainbow** if each edge of S is in a different colour. An arborescence is said to be rainbow if its edge set is. We are concerned with the following four events:

$$
\begin{aligned}
\mathcal{C}_m &= \{\mathcal{D}(n,m) \text{ contains edges in at least } n - 1 \text{ colours}\}, \\
\mathcal{Z}_m &= \{\text{At most one vertex of } \mathcal{D}(n,m) \text{ has in-degree zero}\}, \\
\mathcal{A}_m &= \{\mathcal{D}(n,m) \text{ has an arborescence}\}, \\
\mathcal{R}_m &= \{\mathcal{D}(n,m) \text{ has a rainbow arborescence}\}.
\end{aligned}
$$

Let $\mathcal{E}_m$ stand for one of the above four sequences of events and let

$$m_\mathcal{E} = \min\{m \in \mathbb{N} : \mathcal{E}_m \text{ occurs}\},$$

provided that such an m exists. (Note that $m_\mathcal{Z}$ and $m_\mathcal{A}$ are always defined but the other two might not be.) It is obvious that $m_\mathcal{A} \geq m_\mathcal{Z}$. Moreover, if $m_\mathcal{R}$ is defined, then so is $m_\mathcal{C}$ and clearly

$$m_\mathcal{R} \geq \max\{m_\mathcal{A}, m_\mathcal{C}\} \geq m_\mathcal{Z}.$$

It is known (and easy to show using, say, the Brun's sieve—see, for example, Section 8.3 in [1]) that the sharp threshold for property $\mathcal{Z}$ is $n \log n$; in fact, the random variable

counting the number of vertices of in-degree zero tends to the Poisson random variable with expectation e^{-c}, provided that $m = n(\log n + c), c \in \mathbb{R}$, and so

$$\mathbb{P}(\mathcal{Z}_m) = \left(1 + e^{-c} + o(1)\right) e^{-e^{-c}}. \tag{1}$$

Moreover, given other "hitting time" results in random graph processes (see, for example [5, 6] or Chapters 7 and 8 in [4]) it is natural to expect that a.a.s. $m_{\mathcal{A}} = m_{\mathcal{Z}}$. To the best of our knowledge, this result is not published anywhere, but is implied by our result. Moreover, it follows from the coupon collector problem that, say, a.a.s. $m_{\mathcal{C}} < \frac{n}{2} \log n$ and so the desired condition for the number of distinct colours present is satisfied much earlier. Indeed, the expected number of colours not present at time $\frac{n}{2} \log n$ is equal to

$$(1 + 50\varepsilon)n \left(1 - \frac{1}{(1 + 50\varepsilon)n}\right)^{\frac{n}{2} \log n} < 2n \exp\left(-\frac{\log n}{2(1 + 50\varepsilon)}\right) = n^{1/2 + o(1)}.$$

Hence, a.a.s. at most $n^{2/3} \le 50\varepsilon n$ colours are missing at that point of the process.

In this paper, we show the following result.

Theorem 1.1. *We have that under the previous assumptions a.a.s.*

$$m_{\mathcal{R}} = m_{\mathcal{A}} = m_{\mathcal{Z}}.$$

The following corollary follows immediately from (1).

Corollary 1.2. *Let $m = n(\log n + c)$ for some $c \in \mathbb{R}$. Then,*

$$\mathbb{P}(\mathcal{R}_m) = \left(1 + e^{-c} + o(1)\right) e^{-e^{-c}}.$$

Our results refer to the random graph process. However, it will be sometimes easier to work with the $D(n, p)$ model instead of $\mathcal{D}(n, m)$. The **random digraph** $D(n, p)$ consists of the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the set of all digraphs with vertex set $\{1, 2, \ldots, n\}$, $\mathcal{F}$ is the family of all subsets of Ω, and for every $G \in \Omega$,

$$\mathbb{P}(G) = p^{|E(G)|}(1 - p)^{n(n-1) - |E(G)|}.$$

This space may be viewed as the set of outcomes of $n(n-1)$ independent coin flips, one for each ordered pair (u, v) of vertices, where the probability of success (that is, adding directed edge (u, v)) is p. Note that $p = p(n)$ may (and usually does) tend to zero as n tends to infinity. We often write $D(n, p)$ when we mean a graph drawn from the distribution $D(n, p)$.

Lemma 1.3 below provides us with a tool to translate results from $D(n, p)$ to $\mathcal{D}(n, m)$—see the first Proposition in [15].

Lemma 1.3. *Let P be an arbitrary property and set $p = p(n) = m/n(n-1)$. If $m = m(n) \to \infty$ is any function such that $m(1 - p) \to \infty$, then,*

$$\mathbb{P}(\mathcal{D}(n, m) \in P) \le 5\sqrt{m} \cdot \mathbb{P}(D(n, p) \in P).$$

We will also use the following version of the Chernoff bound:

Lemma 1.4 (Chernoff Bound). *If X is a binomial random variable with expectation μ, and $0 < \delta < 1$, then*

$$\Pr[X < (1 - \delta)\mu] \leq \exp\left(-\frac{\delta^2\mu}{2}\right)$$

and if $\delta > 0$,

$$\Pr[X > (1 + \delta)\mu] \leq \exp\left(-\frac{\delta^2\mu}{2 + \delta}\right).$$

For a subset of colours $C \subseteq W$, we use the notation $\deg_C^-(v)$ ($\deg_C^+(v)$, respectively) to refer to the in-degree (out-degree, respectively) of v in the graph induced by edges with colours from C.

2. Proof of Theorem 1.1

2.1. **Overview.** The proof is technical and involves revealing some aspects of the random object one by one. We give this overview to summarize the order in which randomness is revealed in the proof.

The model $\mathcal{D}(n, m)$ may be thought of as a collection of m triples, (u, v, c) where (u, v) is the edge and c is the colour that it is assigned. In the random digraph process, when an edge arrives, we do not have to reveal the entire triple. We may, for example, only reveal the second coordinate (the head of the edge) and whether the third coordinate (colour) lies in some fixed subset of colours. This is what we mean by revealing the randomness in a certain order. Before the process begins we arbitrarily select $5\varepsilon n$ colours to be "special". The remaining colours are "regular". Now, we consider the random graph process, and for each edge we reveal whether its colour is special or regular. If the edge is special, then we do not reveal anything else at this time; on the other hand, if the edge is regular, then we reveal its head vertex (but not its tail and not its colour).

After m_- many steps of the random digraph process, we identify some "dangerous" vertices that have in-degree at most 2 (in regular colours). Now, we go back to edges with special colours and reveal whether their heads are dangerous. If so, then we reveal this head. We then continue the process until time $m_{\mathcal{Z}}$ revealing the heads of regular coloured edges and revealing the heads of special edges if the head is a dangerous vertex. By doing this, we are able to determine exactly when the event $\mathcal{Z}_m$ holds for the first time (*i.e.* $m_{\mathcal{Z}}$) since the unique vertex of in-degree zero at time $m_{\mathcal{Z}}$ must be among the dangerous vertices. Actually, this vertex, which we call u, obviously must be among the vertices of in-degree zero, but we define dangerous in this way for another reason. Also note that we do not need to know the special heads to non-dangerous vertices to determine $m_{\mathcal{Z}}$. Our goal here is to show that at time $m_{\mathcal{Z}}$, each vertex (except for u, which will become the root of the final arborescence) can select a unique colour from its in-edges. The $n - 1$ colours used for this will now be called "mapping colours". Note that the mapping colours are either regular or they are special and belong to an edge directed to a dangerous vertex. Thus we will be left with at least $(1 + 45\varepsilon)n - (n - 1) = 45\varepsilon n + 1$ many regular colours which are not mapping colours.

We will then show that a.a.s. there is an arborescence of size at least $n^{2/3}$, using only regular non-mapping colours, and rooted at the vertex u. At that point we reveal the tails of all the edges of mapping colours. These edges, together with the size $n^{2/3}$ arborescence, will a.a.s. give us one large arborescence with almost all of the vertices in the graph, together

with a few small arborescences. To show that we can connect them all together, we finally reveal the heads and tails of the remaining unrevealed special edges.

2.2. **Proof.** Let $\omega = \omega(n)$ be any function tending to infinity (but sufficiently slowly) together with n. For definiteness, let $\omega = \omega(n) = \log \log n$. Let

$$m_- := \lfloor n(\log n - \omega) \rfloor \quad \text{and} \quad m_+ := \lceil n(\log n + \omega) \rceil .$$

(Here we make sure m_- and m_+ are both integers but later on expressions such as $n(\log n + \omega)$ that clearly have to be an integer, we round up or down but do not specify which: the choice of which does not affect the argument.) As we already mentioned, $m_{\mathcal{R}} \geq m_{\mathcal{A}} \geq m_{\mathcal{Z}}$ and it is known that a.a.s.

$$m_- \leq m_{\mathcal{Z}} \leq m_+. \tag{2}$$

Our goal is to show that at time $m_{\mathcal{Z}}$, a.a.s. there is a rainbow arborescence in $\mathcal{D}(n, m_{\mathcal{Z}})$; that is, that a.a.s. $m_{\mathcal{R}} = m_{\mathcal{Z}}$.

Let us start with showing that at time $m_{\mathcal{Z}}$, a.a.s. all but the only vertex of in-degree zero can select a unique colour assigned to an edge from some of their in-neighbours. However, as we mentioned earlier, it is important that we apply the so-called **multi-round exposure** technique. We arbitrarily choose $5\varepsilon n$ colours to be **special** and let the other $(1 + 45\varepsilon)n$ colours be **regular**. Until time m_- we expose only the heads of regular-coloured edges. We also reveal the in-degrees in special colours of the dangerous vertices (formally defined in the proof of Lemma 2.1 below), but nothing else. As a consequence of this, if we condition on the event that vertex v has precisely k in-neighbours (and perhaps also on the event that some of the edges from them to v are coloured in the desired way), in-neighbours can be selected uniformly at random from all $\binom{n-1}{k}$ possible k-tuples.

Lemma 2.1. *A.a.s. the following property holds in $\mathcal{D}(n, m_{\mathcal{Z}})$. There exists a set $T \subset W$ of size $n - 1$, a vertex $u \in V$, and a perfect matching $f : V \setminus \{u\} \to T$ so that every vertex $v \in V \setminus \{u\}$ has an edge directed to v in colour $f(v)$.*

Note that u must be the only vertex of in-degree zero in $\mathcal{D}(n, m_{\mathcal{Z}})$. This vertex will be used as a root of the arborescence. The $n - 1$ colours from the set T will now be called **mapping colours**.

Proof. We think of the assignment problem we deal with as a bipartite graph: on one side we have the set of n vertices V, on the other side we have the set of $(1 + 50\varepsilon)n$ colours W. A vertex $v \in V$ is adjacent to colour $c \in W$ if at least one edge to v has colour c. We use Hall's necessary and sufficient condition to show that the desired matching exists.

There is no matching saturating V if and only if for some $k \geq 1$ there exists a k-**witness**, that is, a pair (S, T) of sets $S \subseteq V$, $T \subseteq W$ such that $|S| = k$, $|T| = k - 1$, and $N(S) \subseteq T$. We say that vertex v is **dangerous** if its in-degree (in regular colours) is at most 2 at time m_-. The probability that a given vertex has $\ell \leq 2$ regular-coloured edges incident to it is

equal to

$$\binom{m_-}{\ell}\left(\frac{n - O(\ell)}{n(n-1) - O(m_-)} \cdot \frac{1 + 45\varepsilon}{1 + 50\varepsilon}\right)^{\ell}\left(1 - \frac{n - O(\ell)}{n(n-1) - O(m_-)} \cdot \frac{1 + 45\varepsilon}{1 + 50\varepsilon}\right)^{m_- - \ell}$$

$$= (1 + o(1))\frac{(n\log n)^{\ell}}{\ell!} \cdot \frac{1}{n^{\ell}} \cdot \exp\left(-\frac{m_-}{n}\left(1 - 5\varepsilon + O(\varepsilon^2)\right)\right)$$

$$= (1 + o(1))\frac{(\log n)^{\ell}}{\ell!} \cdot \exp\left(-\log n + \omega + 5\varepsilon\log n\right)$$

$$= (1 + o(1))\frac{e^{\omega + 5\varepsilon\log n}(\log n)^{\ell}}{\ell!n}.$$

Hence, the expected number of dangerous vertices is $O(e^{\omega + 5\varepsilon\log n}\log^2 n) = O(\log^8 n)$ and so a.a.s. there are at most, say, $O(\log^9 n)$ dangerous vertices. We now reveal whether the heads of special-coloured edges are dangerous and if so we reveal the head. Thus at time m_-, we know the in-degree in regular-coloured edges of every vertex and we know the in-degree in special-coloured edges for dangerous vertices. For the rest of the process, until time $m_{\mathcal{Z}}$, we continue to reveal the heads of regular-coloured edges. Additionally, when a special-coloured edge arrives, we reveal whether its head is dangerous, and if so we reveal it completely. Now, it follows immediately from the definition of $m_{\mathcal{Z}}$ that there is precisely one 1-witness at that point of the process: $S = \{u\}$, where u is the only vertex of in-degree zero.

Now, our goal is to show that a.a.s. there is a matching saturating $V \setminus \{u\}$; that is, a.a.s. there is no k-witness (S, T) with $S \subseteq V \setminus \{u\}$ and $2 \leq k \leq n - 1$. First, let us concentrate on the case $k = 2$. Note that at time $m > m_-$, a dangerous vertex may or may not have at least 3 in-neighbours but a non-dangerous vertex will definitely have at least 3. Moreover, note that vertex u is dangerous but every other dangerous vertex has in-degree at least 1 at time $m_{\mathcal{Z}}$. Let us focus on the first incoming edges assigned to these dangerous vertices (regardless of whether they are regular-coloured or special), conditioning on the event that there are $O(\log^9 n)$ such vertices. By "first incoming edges" we are referring to the order given by the random digraph process. The probability that two of these dangerous vertices receive the same regular colour on their first incoming edge is at most

$$\binom{O(\log^9 n)}{2} \cdot (1 + 45\varepsilon)\, n \cdot \left(\frac{1}{(1 + 45\varepsilon)n}\right)^2 = O\left(\frac{\log^{18} n}{n}\right) = o(1).$$

Similarly, the probability that two of these dangerous vertices receive the same special colour on their first incoming edge is at most

$$\binom{O(\log^9 n)}{2} \cdot 5\varepsilon n \cdot \left(\frac{1}{5\varepsilon n}\right)^2 = O\left(\frac{\log^{18} n}{\varepsilon n}\right) = o(1).$$

So we may assume that at time $m_{\mathcal{Z}}$, every pair of dangerous vertices (excluding u) sees at least 2 different colours. On the other hand, for each non-dangerous vertex we focus on the first 3 regular-coloured incoming edges assigned to them. The probability that all 3 of these edges receive the same colour is equal to

$$(1 + 45\varepsilon)n\left(\frac{1}{(1 + 45\varepsilon)n}\right)^3 = O\left(\frac{1}{n^2}\right).$$

Hence, a.a.s. all non-dangerous vertices have at least least 2 distinct colours assigned at time $m_{\mathcal{Z}}$. Conditioning on these two events, we get that there is no 2-witness at time $m_{\mathcal{Z}}$. Indeed,

every vertex in a set $S \subseteq V$ consisting of two dangerous vertices yields a unique colour; on the other hand, S containing at least one non-dangerous vertex yields at least 2 colours.

Now, we move to the case $k \geq 3$. This time, it will be convenient to focus on minimal configurations. A k-witness is **minimal** if there does not exists $S' \subset S$ and $T' \subset T$ such that (S', T') is a k'-witness, where $k' < k$. It is straightforward to see that if (S, T) is a minimal k-witness, then every c in T has degree at least 2 in the graph induced by $S \cup T$. Hence there are at least $2(k-1)$ edges between S and T. The goal is to show that there is no minimal k-witness in $\mathcal{D}(n, m_{\mathcal{Z}})$, even when only regular colours are taken into consideration. However, having no minimal k-witness at time m_1 does not imply that there is no k-witness at time $m_2 > m_1$. Hence, we need to estimate the probability that the desired property holds for a given m such that $m_- \leq m \leq m_+$ and then take a union bound over all $O(n\omega) = O(n \log \log n)$ possible values of m. In fact, we will show that with probability $1 - o((n \log n)^{-1}(n \log n)^{-1/2})$, $D(n, p)$ has no k-witness for $k \geq 3$, for

$$p_- \leq p \leq p_+$$

where

$$p_{\pm} = \frac{m_{\pm}}{n(n-1)} = \frac{\log n \pm \omega + o(1)}{n}.$$

As already mentioned, the claim will hold by Lemma 1.3 and the union bound.

Let

$$p' := p \cdot \frac{1 + 45\varepsilon}{1 + 50\varepsilon} = \frac{\log n - O(\log \log n)}{n}$$

be the probability that an edge is both present and of a regular colour. Fix $S \subseteq V, T \subseteq W$, $|S| = k$, $|T| = k - 1$. The probability that the number of in-neighbours (to S, in regular colours) is equal to t is

$$\binom{k(n-1)}{t} p'^t (1 - p')^{k(n-1)-t}.$$

So the probability that (S, T) is a k-witness is at most

$$p_{S,T} = \sum_{t \geq 2(k-1)} \binom{k(n-1)}{t} p'^t (1 - p')^{k(n-1)-t} \left(\frac{k-1}{(1+45\varepsilon)n} \right)^t.$$

Thus, the expected number of minimal witnesses for any $k \geq 3$ is at most

$$\sum_{k \geq 3} \binom{n}{k} \binom{(1+45\varepsilon)n}{k-1} \sum_{t \geq 2(k-1)} \binom{k(n-1)}{t} p'^t (1 - p')^{k(n-1)-t} \left(\frac{k-1}{(1+45\varepsilon)n} \right)^t. \qquad (3)$$

To estimate the inner sum, first note that the ratio of consecutive terms is equal to

$$\frac{\binom{k(n-1)}{t+1} p'^{t+1} (1-p')^{k(n-1)-t-1} \left(\frac{k-1}{(1+45\varepsilon)n} \right)^{t+1}}{\binom{k(n-1)}{t} p'^t (1-p')^{k(n-1)-t} \left(\frac{k-1}{(1+45\varepsilon)n} \right)^t} = \frac{k(n-1)-t}{t+1} \cdot \frac{p'}{1-p'} \cdot \frac{k-1}{(1+45\varepsilon)n}$$

$$\leq \frac{k(n-1) - 2(k-1)}{2(k-1)+1} \cdot \frac{p'}{1-p'} \cdot \frac{k-1}{(1+45\varepsilon)n}$$

$$\leq \frac{k \log n}{1.5n}$$

for n sufficiently large. So, when $k < \frac{n}{\log n}$, the inner sum is of the order of its first term. Now, for $k \geq \frac{n}{\log n}$, note that each term is small:

$$\binom{n}{k}\binom{(1+45\varepsilon)n}{k-1}\binom{k(n-1)}{t}p'^t(1-p')^{k(n-1)-t}\left(\frac{k-1}{(1+45\varepsilon)n}\right)^t$$

$$\leq \left(\frac{ne}{k}\right)^k\left(\frac{(1+45\varepsilon)ne}{k-1}\right)^{k-1}\left(\frac{k(n-1)e}{t}\right)^t\left(\frac{p'}{1-p'}\right)^t\exp\{-k(n-1)p'\}\left(\frac{k-1}{(1+45\varepsilon)n}\right)^t$$

$$\leq \exp\left\{-k(n-1)p'+2k\log\left(\frac{2ne}{k}\right)\right\}\left(\frac{k^2p'e}{t(1-p')(1+45\varepsilon)}\right)^t$$

$$\leq \exp\left\{-k(n-1)p'+\frac{k^2p'}{(1-p')(1+45\varepsilon)}+3k\log\log n\right\},$$

for n large enough, since $\left(\frac{c}{t}\right)^t \leq \exp\left\{\frac{c}{e}\right\}$ which holds for all positive numbers c,t. Indeed, this inequality follows by noting that $\log(x) \leq x/e$ for all $x > 0$ and letting $x = c/t$. Hence, each term is at most

$$\exp\left\{-k(n-1)p'+knp'(1+O(p'))(1-45\varepsilon+O\left(\varepsilon^2\right))+3k\log\log n\right\}$$
$$= \exp\left\{-(45+o(1))\left(\varepsilon knp'\right)+3k\log\log n\right\}$$
$$\leq \exp\left\{-41k\log\log n\right\} = \exp\left\{-\Omega\left(\frac{n\log\log n}{\log n}\right)\right\}.$$

Therefore, the contribution from these terms is $O(n^2\exp(-\Omega(n\log\log n/\log n))) = o(n^{-2})$ and (3) can be upper bounded by

$$o(n^{-2})+\sum_{k=3}^{\frac{n}{\log n}}\sum_{t\geq 2(k-1)}\binom{n}{k}\binom{(1+45\varepsilon)n}{k-1}\binom{k(n-1)}{t}p'^t(1-p')^{k(n-1)-t}\left(\frac{k-1}{(1+45\varepsilon)n}\right)^t$$

$$= o(n^{-2})+O\left(\sum_{k=3}^{\frac{n}{\log n}}\binom{n}{k}\binom{(1+45\varepsilon)n}{k-1}\binom{k(n-1)}{2(k-1)}p'^{2(k-1)}(1-p')^{k(n-1)-2(k-1)}\left(\frac{k}{(1+45\varepsilon)n}\right)^{2(k-1)}\right)$$

$$= o(n^{-2})+O\left(\sum_{k=3}^{\frac{n}{\log n}}\left(\frac{ne}{k}\right)^k\left(\frac{(1+45\varepsilon)ne}{k-1}\right)^{k-1}\left(\frac{k(n-1)e}{2(k-1)}\cdot\frac{p'}{1-p'}\cdot\frac{k-1}{(1+45\varepsilon)n}\right)^{2(k-1)}(1-p')^{k(n-1)}\right)$$

$$= o(n^{-2})+O\left(\sum_{k=3}^{\frac{n}{\log n}}\exp\left(-(k-1)\log n+O(k\omega+k\log\log n)\right)\right)$$

$$= o(n^{-2})+O\left(\sum_{k=3}^{\frac{n}{\log n}}\exp\left(-(1+o(1))(k-1)\log n\right)\right)$$

$$= o\left(n^{-2}\right)+n^{-2+o(1)} = o((n\log n)^{-1}(n\log n)^{-1/2})$$

as desired, since the sum on the second to last line is dominated by its first term. The proof of the lemma is finished, as explained earlier.   □

Now we will show that some of the regular colours not in T can be used to start building a rainbow arborescence of size at least $n^{2/3}$ rooted at the special vertex u. After that, we will show how to convert this arborescence into a spanning one. Let us note that Lemma 2.1 proves only the existence of a set of $n - 1$ colours and a corresponding matching. Unfortunately, we do not control which colours are used and which are still available for us to construct the arborescence around u. Therefore, we can only use properties that hold for *all* sets of colours of the desired size.

Let us start with the following useful observation that holds a.a.s. in $\mathcal{D}(n, m_-)$, and is deterministically preserved in $\mathcal{D}(n, m)$, provided that $m > m_-$.

Lemma 2.2. *A.a.s. the following property holds in $\mathcal{D}(n, m_-)$. Every set of colours $C \subseteq W$ with $|C| = 45\varepsilon n$ satisfies*

$$\left|\left\{v \in V \,:\, \deg_C^-(v) \le 43\varepsilon \log n\right\}\right| \le \frac{n}{\varepsilon \log n} = \frac{n}{\log \log n}.$$

Proof. For convenience, we will work with $D(n, p_-)$ with

$$p_- = \frac{m_-}{n(n-1)} = \frac{\log n - \omega}{n - 1} = \frac{\log n - \omega - o(1)}{n}.$$

Let C be any set of colours of size $45\varepsilon n$. Then

$$\mathbb{E}\left[\deg_C^-(v)\right] = \frac{45\varepsilon}{1 + 50\varepsilon} \cdot p_- \cdot (n - 1) > 44\varepsilon \log n.$$

It follows from the Chernoff Bound that

$$\Pr\left[\deg_C^-(v) \le 43\varepsilon \cdot \log n\right] \le \exp\left(-\frac{1}{2} \cdot \left(\frac{1}{44}\right)^2 \cdot \varepsilon \log n\right) \le \exp\left(-\frac{\varepsilon \log n}{5000}\right).$$

So the probability that there exist $\frac{n}{\varepsilon \log n}$ many vertices of small in-degree is at most

$$\binom{n}{\frac{n}{\varepsilon \log n}} \cdot \exp\left(-\frac{\varepsilon \log n}{5000} \cdot \frac{n}{\varepsilon \log n}\right) \le \exp\left(\frac{n}{\varepsilon \log n} \cdot \log(e\varepsilon \log n) - \frac{n}{5000}\right) \le \exp\left(-\Omega(n)\right).$$

Since there are

$$\binom{(1 + 50\varepsilon)n}{45\varepsilon n} \le \exp\left(45\varepsilon n \log\left(\frac{e \cdot (1 + 50\varepsilon)}{45\varepsilon}\right)\right) = \exp\left(O\left(\frac{\log^2 \log n}{\log n} n\right)\right) = \exp(o(n))$$

sets of colours to consider, the desired property holds with probability $1 - o(1/n)$ in $D(n, p_-)$ by the union bound, and so it holds a.a.s. in $\mathcal{D}(n, m_-)$ by Lemma 1.3. $\qquad\square$

We will need the following three claims. The first claim is an easy observation. We say a vertex and an edge are incident if the vertex is either the head or tail of the edge.

Claim 2.3. *In $\mathcal{D}(n, m)$ with $m \le m_+$, a.a.s.,*
  (i) *no colour appears more than $10 \log n$ times,*
  (ii) *no vertex is incident to more than $10 \log n$ edges,*
  (iii) *no vertex is incident to more than 10 edges of the same colour.*

Proof. It is enough to prove these properties for $D(n, p_+)$ and apply Lemma 1.3. Each follows easily from Chernoff and union bounds. $\qquad\square$

While the above lemma holds for all colours, we will only use it for regular colours and for special colours incident to dangerous vertices.

As we will not have failure probabilities small enough to union bound over all choices of m with $m_- \leq m \leq m_+$, we would like to prove the existence of the desired structure at time m_-. One issue is that at time m_-, we do not know which vertex is going to be the root. However, we do know (as it is shown in the proof of Lemma 2.1) that a.a.s. there are $O(\log^9 n)$ many dangerous vertices; that is, vertices of in-degree at most 2 in regular colours. One of these vertices must be the root u at time $m_{\mathcal{Z}}$.

Let C' be an arbitrarily chosen set of $45\varepsilon n$ regular non-mapping colors. (Recall that there are at least $45\varepsilon n+1$ many of them to choose from. By Lemma 2.2, all but at most $n/\log\log n$ many vertices v have $\deg_{C'}^-(v) \geq 43\varepsilon \log n$. We will call such vertices **good**.

Claim 2.4. *In $\mathcal{D}(n, m_-)$, a.a.s. every dangerous vertex w has $\deg_{C'}^+(w) \geq 10\varepsilon \log n = 10 \log\log n$.*

Proof. For any good vertex v and any vertex w, we have that

$$\mathbb{P}\left[(w, v) \in E_{C'}\right] = \frac{\deg_{C'}^-(v)}{n-1} \geq \frac{42\varepsilon \log n}{n},$$

where $E_{C'}$ represents the edges coloured with C'. Thus we have that for every vertex w, $\deg_{C'}^+(w)$ stochastically dominates

$$\mathrm{Bin}\left[n\left(1 - \frac{1}{\log\log n}\right), \frac{42\varepsilon \log n}{n}\right],$$

with the expected value of $(42 + o(1))\varepsilon \log n$. So by the Chernoff Bound,

$$\mathbb{P}\left[\deg_{C'}^+(w) < 10\varepsilon \log n\right] \leq \exp\left(-\frac{1}{2} \cdot \left(\frac{42 - 10}{42}\right)^2 \cdot (42 + o(1))\varepsilon \log n\right) = o(\log^{-9} n).$$

Thus, a.a.s. each dangerous vertex has at least $10\varepsilon \log n$ many out neighbours since we again condition on the event that there are only $O(\log^9 n)$ of them. □

We will show that, if S is not too small nor too large, there are many edges outgoing from S with different heads and different colours. This will be enough to grow the tree to the desired size.

Claim 2.5. *Let S be a set of vertices such that $|S| = \Omega(\varepsilon \log n) = \Omega(\log\log n)$, $|S| \leq n^{2/3}$, and let T be a set of at least $n/2$ good vertices which is disjoint from S. Then with probability at least $1 - \exp\left(-\Omega(\varepsilon^2 \log^2 n)\right) = 1 - \exp\left(-\Omega(\log^2 \log n)\right)$ the following holds: there is a set of $3|S|\varepsilon \log n = 3|S|\log\log n$ edges going from S to T whose head vertices are all distinct and whose colours are all distinct and in C'.*

Proof. We use a two-round exposure. First we will expose whether some edges are present and in the colour set C', without exposing exactly which colour they are. In the second round we take the edges we found in the first round and expose their colours, conditioning on the colours being in C' of course. Let v be any good vertex and set $d = \deg_{C'}^-(v)$. Suppose

for now that S is any set of vertices not including v and let $|S| = s$. Then the probability that there is an edge from a vertex of S to v with a colour from C' is

$$1 - \frac{\binom{n-1-s}{d}}{\binom{n-1}{d}} = 1 - \frac{(n-1-s)_d}{(n-1)_d}$$

$$= 1 - \prod_{i=1}^{d}\left(1 - \frac{s}{n-i}\right)$$

$$\geq 1 - \left(1 - \frac{s}{n}\right)^d$$

$$\geq 1 - \left(1 - \frac{s}{n}\right)^{43\varepsilon \log n}.$$

Now suppose S satisfies $s = \Omega(\varepsilon \log n)$ and $s \leq n^{2/3}$. Then $|N_{C'}^{+}(S) \cap T|$ (where $N_{C'}^{+}(S)$ represents the out-neighbours of vertices in S coloured with colours from C') stochastically dominates

$$\mathrm{Bin}\left[\frac{n}{2}, 1 - \left(1 - \frac{s}{n}\right)^{43\varepsilon \log n}\right].$$

Since $s \leq n^{2/3}$, we have

$$1 - \left(1 - \frac{s}{n}\right)^{43\varepsilon \log n} \sim \frac{43\varepsilon s \log n}{n}.$$

So by the Chernoff Bound, the probability that $|N_{C'}^{+}(S) \cap T|$ less than $10s\varepsilon \log n$ is at most $\exp\left(-\Theta\left(s\varepsilon \log n\right)\right) \leq \exp\left(-\Omega(\varepsilon^2 \log^2 n)\right)$.

Now we do the second round of exposure for the proof of the claim. Conditioning on having $10s\varepsilon \log n$ edges going from S to T with distinct heads and colours in C', we reveal which colours the edges are to see if they are distinct. The probability that 4 of these edges get the same colour is at most

$$\binom{10s\varepsilon \log n}{4} \cdot 45\varepsilon n \cdot \left(\frac{1}{45\varepsilon n}\right)^4 \leq n^{(2/3)\cdot 4 + 1 - 4 + o(1)} = n^{-1/3 + o(1)} \leq \exp\left(-\Omega(\varepsilon^2 \log^2 n)\right),$$

where we have used the fact that $s \leq n^{2/3}$. We conclude that there are at least $3s\varepsilon \log n$ distinct colours on these edges and the proof of the claim is finished. $\qquad \square$

With these three claims, we can begin to grow our arborescence.

Lemma 2.6. *A.a.s. $\mathcal{D}(n, m_{\mathcal{Z}})$ has a rainbow arborescence with $n^{2/3}$ many vertices, rooted at vertex u which uses only colours from C'.*

Proof. As mentioned, we will actually show that this statement holds at time m_- for all dangerous vertices, of which u is one. For each dangerous vertex w we can do the following. Let B initially be the set of vertices which are not good, (that is, those which have $\deg_{C'}^{-}(v) < 43\varepsilon \log n$) so $|B| \leq \frac{n}{\varepsilon \log n}$ by Lemma 2.2. As we proceed, we will want to ignore certain vertices and we will move them into B. Note that some of the dangerous vertices might be in B (but that does not affect our argument). Let T be the remaining good vertices that are not dangerous. By Claims 2.3(iii) and 2.4, w has at least $\varepsilon \log n$ many out-neighbours with distinct colours from C'. Let C'' be a set of precisely $\varepsilon \log n$ of these colours and let S be the corresponding out-neighbours. Move from T to B any vertex which is incident to an edge with a colour from C''. Then by Claim 2.3(i), we have removed at most $O(\log^2 n)$ vertices

from T. We may now repeatedly apply Claim 2.5 to find edges from S to T with new colours. After each application of Claim 2.5, we replace S by the set of heads of the newly found edges and we move from T to B all vertices incident to an edge with the same colour as any of the new edges. We stop when we have first accumulated at least $n^{2/3}$ vertices. At most $O(n^{2/3} \log n) = o(n)$ many vertices are removed from T and thus we always have $|T| \geq n/2$. Each application of Claim 2.5 multiplies the number of vertices in our arborescence by a factor of $\Omega(\varepsilon \log n) = \Omega(\log \log n)$, so we apply it $O(\log n / \log \log \log n) = O(\log n)$ many times (to each of $O(\log^9 n)$ dangerous vertices). We succeed a.a.s. by the union bound since Claim 2.5 holds with probability $1 - \exp\left(-\Omega(\log^2 \log n)\right)$ and

$$O(\log^{10} n \exp\left(-\Omega(\log^2 \log n)\right)) = \exp\left(-\Omega(\log^2 \log n)\right) = o(1).$$

The proof of the lemma is finished.   □

Now we expose the $n - 1$ edges we reserved at the beginning, associated with mapping colours. In the digraph induced by these $n - 1$ edges, every vertex has in-degree 1 except for the root, which has in-degree 0. Thus if we randomly put an extra edge pointing to the root, we have a "loop-less random mapping" (which is actually independent from the arborescence we have built so far).

We call the reader's attention to a few small but important points. First, a **random mapping** is a function chosen uniformly at random from all functions $f : [n] \to [n]$. A random mapping can be identified with a digraph where each vertex chooses a random *out-neighbour*. We have the reverse situation. Each vertex is choosing a random *in-neighbour*. Second, a pure random mapping allows vertices to map to themselves, *i.e.* allows self-loops. We have no loops in our situation. However, in a random mapping, the number of loops is asymptotically distributed as Poisson with mean 1. Thus with probability bounded away from 0, a random mapping has no loops and any properties which hold a.a.s. for random mappings also hold a.a.s. for loop-less random mappings. (See Section 14.5 in [4] for basic facts about random mappings)

We now consider the so called "inverse epidemic process" (See [12] or Section 14.5 in [4]). The process is as follows: suppose that m members of a population are initially infected with a disease. The disease spreads to other vertices along the edges of a random mapping in the reverse orientation. Note that for us, since our mapping edges are already reversed, this means that the disease is spreading along our digraph edges in the forward direction.

We would like to apply a result of Burtin [7]. The result implies that if $x = n^{2/3}$ vertices are initially infected, and η is the random variable for the number of eventually infected vertices, then $\left(\frac{x}{n}\right)^2 (n - \eta)$ is a.a.s. less than any function going to infinity with n, say $n^{1/6}$. So all but at most, say $n^{5/6}$ vertices total are eventually infected. We let our initially infected set, A, be the vertices in the arborescence we found of size $n^{2/3}$. The eventually infected vertices are in the same component as the root vertex. Denote the eventually infected vertices as M_I and the uninfected vertices as M_U. First, we would like to show that a.a.s. M_U contains no dangerous vertices. Since our random mapping is independent of A, we have that any vertex in $V \setminus A$ is equally likely to be in M_U and this probability is $|M_U|/|V \setminus A| = O\left(n^{5/6}/(n - n^{2/3})\right) = O(n^{-1/6})$. Thus the expected number of dangerous vertices in M_U is $O(\log^9 n \cdot n^{-1/6}) = o(1)$ and so a.a.s. M_U contains no dangerous vertices.

Now, it is well known and easy to see that a.a.s. there are $O(\log n)$ many cycles in a random mapping. Here we consider double edges to be cycles and note that each component

of the digraph induced by the mapping edges is unicyclic. We arbitrarily remove one edge from each cycle to get $O(\log n)$, many arborescence components, each with its own "root". Our goal is to show that we can connect the large component M_I to the small arborescence components, whose vertices comprise M_U.

Note that since there are $O(\log^9 n)$ dangerous vertices, each incident to at most $10 \log n$ colours, there are at least $5\varepsilon n - O(\log^{10} n)$ special, non-mapping colours. We will use these colours to connect the big component to the small ones. Recall that so far we have exposed all regular edges, as well as the in-degrees in special edges of the dangerous vertices (and these are the only special in-degrees which have been revealed). Suppose A_1 is a small arborescence component (contained in M_U) and suppose its root is u_1. To connect the large component to A_1, we search for an edge whose tail is anywhere in M_I, whose head is u_1, and whose colour has not yet been used. Since no vertex of M_U is dangerous, u_1's special in-degree is unrevealed. If we find a special edge from the large component to u_1, we add it to the arborescence (and discard that colour from future use). We will attempt to do this for $O(\log n)$ many small components. Hence, at most $O(\log n)$ colours are discarded this way, and since each colour appears on at most $10 \log n$ edges, at most $O(\log^2 n)$ edges are discarded. A.a.s. there are at least $(5 - o(1))\varepsilon n \log n$ many edges in the set of eligible colours when we start exposing edges, so there are at least $4\varepsilon n \log n$ edges to use each time. Also, by Claim 2.3(ii) we may assume that no vertex (and so none of the roots) has in-degree more than $10 \log n$. So when we look for our random special edge going to a particular root there are at most $10 \log n$ possibilities forbidden by our conditioning. Thus if there are x many small components that remain to be connected to the giant, then the probability that any of them fail to connect is at most

$$x \left(1 - \frac{n - n^{5/6} - 10 \log n}{n(n-1)}\right)^{4\varepsilon n \log n} \leq x \left(1 - \frac{1}{2n}\right)^{4\varepsilon n \log n} \leq x \exp\{-2\varepsilon \log n\} = \frac{x}{\log^2 n} = o(1),$$

since $x = O(\log n)$. This finishes the proof of Theorem 1.1.

3. Conclusion and open questions

Our result, Theorem 1.1, is tight in one sense in that the hitting time is correct in terms of the structure of the digraph. It would be of interest to see precisely how many colours we need to choose from to get this result. Can we find the precise hitting time if we only have $n - 1$ colours? Suppose we want a rainbow copy of some specific spanning tree, for example, a directed Hamilton path. How many random edges and colours are needed? How many edge disjoint rainbow spanning trees (respectively, arborescences) can we pack into $G_{n,m}$ (respectively, $\mathcal{D}_{n,m}$)?

References

[1] N. Alon and J.H. Spencer, *The Probabilistic Method*, Wiley, 1992 (Third Edition, 2008).
[2] D. Bal, P. Bennett, A.M. Frieze and P. Prałat, *Power of k choices and rainbow spanning trees in random graphs*, Electronic Journal of Combinatorics **22(1)** (2015), #P1.29.
[3] D. Bal and A.M. Frieze, *Rainbow Matchings and Hamilton Cycles in Random Graphs*, preprint.
[4] B. Bollobás, *Random Graphs*, Cambridge University Press, Cambridge, 2001.
[5] B. Bollobás, *The evolution of sparse graphs.* Graph theory and Combinatorics (Cambridge, 1983), 35-57, Academic Press, London, 1984.
[6] B. Bollobás, A. Thomason, *Random graphs of small order.* Random graphs '83 (Poznań, 1983), 47-97, North-Holland Math. Stud., 118, North-Holland, Amsterdam, 1985.

[7] Y. Burtin, *On a simple formula for random mappings and its applications*, Journal of Applied Probability **17** (1980), 403-414

[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. Ferber, G. Kronenberg, Mousset and Shikhelman, *Packing a randomly edge-colored random graph with rainbow k-outs*, preprint.

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

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

[12] I. B. Gertsbakh, *Epidemic process on a random graph: some preliminary results.* J. Appl. Probability **14** (1977), 427-438.

[13] S. Janson, T. Łuczak and A. Rucinski, *Random Graphs*, Wiley-Intersci. Ser. Discrete Math. Optim., 2000.

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

[15] B. Pittel, *On the probable behaviour of some algorithms for finding the stability number of a graph*, Math. Proc. Camb. Philos. Soc. **92** (1980) 511–526.

DEPARTMENT OF MATHEMATICAL SCIENCES, MONTCLAIR STATE UNIVERSITY, MONTCLAIR, NJ, 07043 U.S.A.
 E-mail address: `bald@mail.montclair.edu`

DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF TORONTO, TORONTO, ON, CANADA, M5S 3G4
 E-mail address: `patrickb@cs.toronto.edu`

DEPARTMENT OF INFORMATICS, KINGS COLLEGE, UNIVERSITY OF LONDON, LONDON WC2R 2LS, UK
 E-mail address: `colin.cooper@kcl.ac.uk`

DEPARTMENT OF MATHEMATICAL SCIENCES, CARNEGIE MELLON UNIVERSITY, 5000 FORBES AV., 15213, PITTSBURGH, PA, U.S.A
 E-mail address: `alan@random.math.cmu.edu`

DEPARTMENT OF MATHEMATICS, RYERSON UNIVERSITY, TORONTO, ON, CANADA, M5B 2K3
 E-mail address: `pralat@ryerson.ca`