Rainbow matchings and rainbow connectedness

Alexey Pokrovskiy*
Department of Mathematics
ETH Zürich
Zürich, Switzerland
alexey.pokrovskiy.maths@gmail.com

Submitted: May 9, 2015; Accepted: Jan 20, 2017; Published: Feb 3, 2017
Mathematics Subject Classifications: 05C70, 05C40, 05C20, 05B15, 05D15

Abstract

Aharoni and Berger conjectured that every collection of n matchings of size $n+1$ in a bipartite graph contains a rainbow matching of size n. This conjecture is related to several old conjectures of Ryser, Brualdi, and Stein about transversals in Latin squares. There have been many recent partial results about the Aharoni-Berger Conjecture. The conjecture is known to hold when the matchings are much larger than $n+1$. The best bound is currently due to Aharoni, Kotlar, and Ziv who proved the conjecture when the matchings are of size at least $3n/2+1$. When the matchings are all edge-disjoint and perfect, the best result follows from a theorem of Häggkvist and Johansson which implies the conjecture when the matchings have size at least $n+o(n)$.

In this paper we show that the conjecture is true when the matchings have size $n+o(n)$ and are all edge-disjoint (but not necessarily perfect). We also give an alternative argument to prove the conjecture when the matchings have size at least $\phi n+o(n)$ where $\phi \approx 1.618$ is the Golden Ratio.

Our proofs involve studying connectedness in coloured, directed graphs. The notion of connectedness that we introduce is new, and perhaps of independent interest.

1 Introduction

A Latin square of order n is an $n \times n$ array filled with n different symbols, where no symbol appears in the same row or column more than once. Latin squares arise in different branches of mathematics such as algebra (where Latin squares are exactly the multiplication tables of quasigroups) and experimental design (where they give rise to designs

*Supported by SNSF grant 200021-149111 and the Methods for Discrete Structures, Berlin graduate school (GRK 1408).

called Latin square designs). They also occur in recreational mathematics—for example completed Sudoku puzzles are Latin squares.

In this paper we will look for *transversals* in Latin squares. A transversal in a Latin square of order n is a set of n entries such that no two entries are in the same row, same column, or have the same symbol. One reason transversals in Latin squares are interesting is that a Latin square has an orthogonal mate if, and only if, it has a decomposition into disjoint transversals. See [15] for a survey about transversals in Latin squares. It is easy to see that not every Latin square has a transversal (for example the unique 2×2 Latin square has no transversal). However, it is possible that every Latin square contains a large *partial transversal*. Here, a partial transversal of size m means a set of m entries such that no two entries are in the same row, same column, or have the same symbol.

There are several closely related, old, and difficult conjectures which say that Latin squares should have large partial transversals. The first of these is a conjecture of Ryser that every Latin square of odd order contains a transversal [13]. Brualdi and Stein conjectured that every Latin square contains a partial transversal of size $n-1$.

Conjecture 1.1 (Brualdi and Stein, [7, 14])**.** *Every Latin square contains a partial transversal of size $n-1$.*

There have been many partial results about this conjecture. It is known that every Latin square has a partial transversal of size $n-o(n)$—Woolbright [16] and independently Brower, de Vries and Wieringa [6] proved that every Latin square contains a partial transversal of size $n-\sqrt{n}$. This has been improved by Hatami and Schor [11] to $n-O(\log^2 n)$. Häggkvist and Johansson proved a related result about Latin rectangles. For $m \leqslant n$ a $m \times n$ Latin rectangle is an $m \times n$ array of n symbols where no symbol appears in the same row or column more than once. A transversal in a Latin rectangle is a set of m entries no two of which are in the same row, column, or have the same symbol. Häggkvist and Johansson proved the following.

Theorem 1.2 (Häggkvist and Johansson, [10])**.** *For every ϵ, there is an $m_0 = m_0(\epsilon)$ such that the following holds. For every $n \geqslant (1+\epsilon)m \geqslant m_0$, every $m \times n$ Latin rectangle can be decomposed into disjoint transversals.*

This theorem is proved by a probabilistic argument, using a "random greedy process" to construct the transversals. The above theorem gives yet another proof that every sufficiently large $n \times n$ Latin square has a partial transversal of size $n-o(n)$—indeed if we remove ϵn rows of a Latin square we obtain a Latin rectangle to which Theorem 1.2 can be applied.

In this paper we will look at a strengthening of Conjecture 1.1. The strengthening we'll look at is a conjecture due to Aharoni and Berger which takes place in a more general setting than Latin squares—namely edge coloured bipartite graphs. To see how the two settings are related, notice that there is a one-to-one correspondence between $n \times n$ Latin squares and proper edge colourings of $K_{n,n}$ with n colours—indeed to a Latin square S we associate the colouring of $K_{n,n}$ with vertex set $\{x_1, \ldots, x_n, y_1, \ldots, y_n\}$ where the edge between x_i and y_j receives colour $S_{i,j}$. Notice that this colouring is proper i.e. adjacent

edges receive different colours. Recall that a *matching* in a graph is a set of disjoint edges. We call a matching *rainbow* if all of its edges have different colours. It is easy to see that partial transversals in the Latin square S correspond to rainbow matchings in the corresponding coloured $K_{n,n}$. Thus Conjecture 1.1 is equivalent to the statement that "in any proper n-colouring of $K_{n,n}$, there is a rainbow matching of size $n-1$."

One could ask whether a large rainbow matching exists in other settings. Recall that a *simple graph* is a graph which contains at most one edge between any pair of vertices. A *multigraph* is a graph which may contain multiple edges between vertices. Aharoni and Berger posed the following conjecture, which generalises Conjecture 1.1.

Conjecture 1.3 (Aharoni and Berger, [1])**.** *Let G be a multigraph, properly edge coloured by n colours, with at least $n+1$ edges of each colour. Then G contains a rainbow matching with n edges.*

This conjecture was first posed in a different form in [1] as a conjecture about matchings in tripartite hypergraphs (Conjecture 2.4 in [1]). It was first stated as a conjecture about rainbow matchings in [2].

The above conjecture has attracted a lot of attention recently, and there are many partial results. Just like in Conjecture 1.1, one natural way of attacking Conjecture 1.3 is to prove approximate versions of it. As observed by Barat, Gyárfás, and Sárközy [4], the arguments that Woolbright, Brower, de Vries, and Wieringa used to find partial transversals of size size $n-\sqrt{n}$ in Latin squares actually generalise to bipartite graphs to give the following.

Theorem 1.4 (Woolbright, [16]; Brower, de Vries, and Wieringa, [6]; Barat, Gyárfás, and Sárközy, [4])**.** *Let G be a bipartite multigraph, properly edge coloured by n colours, with at least n edges of each colour. Then G contains a rainbow matching with $n-\sqrt{n}$ edges.*

Barat, Gyárfás, and Sárközy actually proved something a bit more general in [4]—for every k, they gave an upper bound on the number of colours needed to find a rainbow matching of size $n-k$.

Another approximate version of Conjecture 1.3 comes from Theorem 1.2. It is easy to see that Theorem 1.2 is equivalent to the following "let G be a bipartite graph consisting of n edge-disjoint perfect matchings, each with at least $n+o(n)$ edges. Then G can be decomposed into disjoint rainbow matchings of size n" (to see that this is equivalent to Theorem 1.2, associate an m-edge coloured bipartite graph with any $m \times n$ Latin rectangle by placing a colour k edge between i and j whenever (k,i) has symbol j in the rectangle).

The main result of this paper is an approximate version of Conjecture 1.3 in the case when the matchings in G are disjoint, but not necessarily perfect.

Theorem 1.5. *For all $\epsilon > 0$, there exists an $N = N(\epsilon) = 10^{20}\epsilon^{-16\epsilon^{-1}}$ such that the following holds. Let G be a bipartite simple graph, properly edge coloured by $n \geqslant N$ colours, with at least $(1+\epsilon)n$ edges of each colour. Then G contains a rainbow matching with n edges.*

Unlike the proof of Theorem 1.2 which can be used to give a randomised process to find a rainbow matching, the proof of Theorem 1.5 is algorithmic. In fact, it can be shown that the matching guaranteed by Theorem 1.5 can be found in polynomial time.

Another very natural approach to Conjecture 1.3 is to to prove it when the matchings have size much larger than $n+1$. When the matchings have size $2n$, the result becomes trivial.

Lemma 1.6. *Let G be a multigraph, properly edge coloured with n colours, each with at least $2n$ edges of each colour. Then G contains a rainbow matching with n edges.*

This lemma is proved by greedily choosing disjoint edges of different colours. We can always choose n edges this way, since each colour class has $2n$ edges (one of which must be disjoint from previously chosen edges).

There have been several improvements to the $2n$ bound in Lemma 1.6. Aharoni, Charbit, and Howard [2] proved that matchings of size $\lfloor 7n/4 \rfloor$ are sufficient to guarantee a rainbow matching of size n. Kotlar and Ziv [12] improved this to $\lfloor 5n/3 \rfloor$. Clemens and Ehrenmüller [8] further improved the constant to $3n/2 + o(n)$. Finally, Aharoni, Kotlar, and Ziv improved the $o(n)$ term to give the following.

Theorem 1.7 (Aharoni, Kotlar, and Ziv, [3])**.** *Let G be a multigraph, properly coloured with n colours, with at least $\lceil 3n/2 \rceil + 1$ edges of each colour. Then G contains a rainbow matching with n edges.*

Though we won't improve on this theorem, we give an alternative proof which gives a weaker bound of $\phi n + o(n)$ where $\phi \approx 1.618$ is the Golden Ratio.

Theorem 1.8. *Let G be a bipartite graph, properly coloured by n colours, with at least $\phi n + 20n/\log n$ edges of each colour. Then G contains a rainbow matching with n edges.*

Theorems 1.5 and 1.8 are proved by studying paths in auxiliary directed graphs. This approach is new and the results we prove about directed graphs may be of independent interest. In particular, we introduce a new notion of connectivity, which we call "rainbow k-connectivity". In the next section we give an informal sketch of the proof of Theorem 1.5. In Section 2 we prove the results about a number of lemmas about directed graphs which we will need. In Section 3 we prove Theorem 1.5. In Section 4 we prove Theorem 1.8. In Section 5 we make some concluding remarks about the techniques used in this paper. For all standard notation we follow [5].

Sketch of proofs

In this section we informally present the main ideas in our proof of Theorem 1.5. Let G be a bipartite simple graph with bipartition classes X and Y properly coloured by n colours and with at least $(1+\epsilon)n$ edges of each colour. Let M be a maximum rainbow matching in G. Let $X_0 = X \setminus V(M)$ and $Y_0 = Y \setminus V(M)$. If G doesn't satisfy the conclusion of Theorem 1.5, then there is some colour, c^*, which is not present in M. Previous approaches to the Aharoni-Berger Conjecture [2, 3, 8, 12] revolved around trying

to perform local manipulations on M until it can be extended to a larger matching. These local manipulations have a similar flavour to "alternating paths" which are often used to study matchings in uncoloured graphs. The following definition is a special case of the kinds of manipulations we will use in the full proof of Theorem 1.5.

Definition 1.9. *An X_0-switching σ is a sequence of distinct edges $\sigma = (e_0, m_1, e_1, \ldots, m_{\ell-1}, e_{\ell-1}, m_\ell)$ such that $m_i \in M$, e_i goes between X_0 and $m_{i+1} \cap Y$, and e_i and m_i have the same colour.*

If e_0 has colour c_0 and m_ℓ has colour c_ℓ, then we say that σ is an X_0-switching from c_0 to c_ℓ. If σ is an X_0-switching from the colour c^* which is not present in M to some other colour c_ℓ, then it is an easy exercise to show that the following is another rainbow matching in G.

$$M_\sigma = M + e_0 - m_1 + e_1 - m_2 + e_3 - \cdots - m_{\ell-1} + e_{\ell-1} - m_\ell.$$

In Section 3 we prove Lemma 3.3 which is a generalization of the statement of M_σ being a rainbow matching. Furthermore M_σ is a rainbow matching of the same size as M and missing c_ℓ, the colour of m_ℓ. Thus X_0-switchings can be used to go between maximum rainbow matchings with different missing colours. At a very high level, our proof can be summarized as "we look for switchings in a graph until we find one which allows us to find a matching bigger than M."

The key idea of this paper is that switchings can be studied using auxiliary directed graphs. The following is a special case of the directed graphs which we will use to study switchings.

Definition 1.10. *The directed graph D_{X_0} is defined as follows.*

- *The vertex set of D_{X_0} is the set of colours of edges in G.*
- *For two colours u and $v \in V(D_{X_0})$, there is a directed edge from u to v in D_{X_0} whenever there is an $x \in X_0$ such that there is a colour u edge from x to the vertex $m_v \cap Y$ in G, where m_v is the colour v edge of M. In this case uv is coloured by "x".*

What does the above definition have to do with X_0-switchings? It is an easy exercise to show that X_0-switchings are in one-to-one corrospondance with rainbow paths in the directed graph D_{X_0}. Specifically, given an X_0-switching $\sigma = (e_0, m_1, e_1, m_2, e_2, \ldots, m_{\ell-1}, e_{\ell-1}, m_\ell)$ from c to c', then the set of edges of D_{X_0} corresponding to $e_0, e_1, \ldots, e_{\ell-1}$ form a rainbow path from c to c'. This correspondence is made precise in Lemma 3.6 in Section 3. Thus studying X_0-switchings is exactly the same as studying rainbow paths in D_{X_0}. Since rainbow paths are more familiar objects that X_0-switchings, this opens up more powerful techniques to study them. For two vertices $u, v \in D_{X_0}$, we use $d_R(u, v)$ to denote the length of the shortest rainbow path from c^* to v

The first observation one makes about D_{X_0} is that any vertex $v \in D_{X_0}$ must have $d^+(v) \geqslant \epsilon n - d_R(c^*, v)$. This is a special case of Lemma 3.7 which we prove in Section 3.

The idea behind it's proof is to consider an X_0-switching σ corresponding to a minimal length path from c^* to v. If $d^+(v) \geqslant \epsilon n - d_R(c^*, v)$ didn't hold, then it is possible to show that a colour v edge between X_0 and Y_0 can be added to M_σ in order to get a larger matching, contradicting the maximality of the original matching M.

Thus we have that vertices close to c^* in D_{X_0} have linear degree. A large part of the proof of Theorem 1.5 involves showing that such graphs have a large highly connected subgraph. But how do we define connectivity of coloured graphs? The following definition is new, and perhaps of independent interest.

Definition 1.11. *An edge coloured directed graph G is said to be rainbow k-edge-connected if for any set S of at most $k-1$ colours and any pair of vertices u and v, there is a rainbow u to v path whose edges have no colours from S.*

The above definition differs from usual notions of connectivity, since generally the avoided set S is a set of *edges* rather than colours.

The key intermediate result we prove is that every properly edge coloured directed graph D has a rainbow k-edge-connected subset C of size roughly $\delta^+(D)$. Lemma 2.10 will make this precise. Since every vertex in D_{X_0} has $d^+(v) \geqslant \epsilon n - d_R(c^*, v)$, this can be used to get a k-edge-connected subset C of size roughly $\epsilon n - o(n)$ where the "$o(n)$" term depends on k. How does such a highly connected subset help in the original problem of finding a matching in G? Recall that the vertices of D_{X_0} correspond to colours in G. In particular all vertices $v \in V(D_{X_0})$ other that c^* have a coresponding edge $m_v \in M$ coloured by v. Given a k-connected subset C, let $X_1 = \{m_v \cap X : v \in C\} \cup X_0$.

Now X_1 is a subset of X which has some kind of "flexibility" property which comes from the rainbow k-connectedness of C. In particular, given any small set S of vertices in X_1 it is possible to find an X_0-switching σ corresponding to some rainbow path in C, such that M_σ is disjoint from S. The precise notion of flexibility which we obtain is given in Definition 3.2. Though far from obvious, it turns out that all the arguments in this section can be repeated with X_0 replaced by X_1. This is made precise in Lemma 3.8. Repeating all the arguments with X_1 instead of X_0 gives a new set X_2 containing X_1, which still has a degree of "flexibility". This process cannot go on forever since at each iteration X_i is roughly ϵn bigger than X_{i-1}. The only way the process can terminate is by finding a matching larger than M in the step during which we established that vertices in D_{X_0} have large out-degree. Thus, if the various parameters are chosen suitably, then we obtain that the original matching M must have used every colour i.e. Theorem 1.5.

2 Paths in directed and coloured graphs

In this section we prove results about paths in various types of directed graphs. All graphs in this section have no multiple edges, although we allow the same edge to appear twice in opposite directions. In directed graphs, "path" will always mean a sequence of vertices $x_1, x_2, \ldots, x_k$ such that $x_i x_{i+1}$ is a directed edge for $i = 1, \ldots, k-1$. The vertices x_1 and x_k are called the *endpoints* of the path, and $x_2, \ldots, x_{k-1}$ are called the *internal vertices* of the path. We say that a path P *internally avoids* a set of vertices S if $\{x_2, \ldots, x_{k-1}\} \cap S = \varnothing$.

We will use additive notation for concatenating paths—for two paths $P = p_1 \dots p_i$ and $Q = q_1 \dots q_j$, $P + Q$ denotes the path with vertex sequence $p_1 \dots p_i q_1 \dots q_j$. Let $N_G^+(v)$ denote the out-neighbourhood of a vertex v in a graph G i.e. the set of vertices x for which vx is an edge in G. When the identity of G is clear from the context, we will abbreviate this to $N^+(v)$. Let $d_G(u, v)$ denote the length of the shortest part from u to v in a graph G.

We will look at coloured graphs. An edge colouring of a graph is an assignment of colours to the edges of a graph. A total colouring is an assignment of colours to both the vertices and edges of a graph. For any coloured graph we denote by $c(v)$ and $c(uv)$ the colour assigned to a vertex or edge respectively.

An edge colouring is *out-proper* if for any vertex v, the outgoing edges from v all have different colours. Similarly an edge colouring is *in-proper* if for any vertex v, the ingoing edges from v all have different colours. We say that an edge colouring is *proper* if it is both in and out-proper (notice that by this definition it is possible to have two edges with the same colour at a vertex v—as long as one of the edges is oriented away from v and one is oriented towards v). A total colouring is proper if the underlying edge colouring and vertex colourings are proper and the colour of any vertex is different from the colour of any edge containing it. A totally coloured graph is *rainbow* if all its vertices and edges have different colours. For two vertices u and v in a coloured graph, $d_R(u, v)$ denotes the length of the shortest rainbow path from u to v. We say that a graph is *vertex-rainbow* if all its vertices have different colours.

This section will mostly be about finding highly connected subsets in directed graphs. The following is the notion of connectivity that we will use.

Definition 2.1. *Let A be a set of vertices in a digraph D. We say that A is (k, d)-connected in D if, for any set of vertices $S \subseteq V(D)$ with $|S| \leqslant k - 1$ and any vertices $x, y \in A \setminus S$, there is an x to y path of length $\leqslant d$ in D avoiding S.*

Notice that a directed graph D is strongly k-connected if, and only if, $V(D)$ is (k, ∞)-connected in D. Also notice that it is possible for a subset $A \subseteq V(D)$ to be highly connected without the induced subgraph $D[A]$ being highly connected—indeed if D is a bipartite graph with classes X and Y where all edges between X and Y are present in both directions, then X is a $(|Y|, 2)$-connected subset of D, although the induced subgraph on X has no edges.

We will also need a generalization this notion of connectivity to coloured graphs

Definition 2.2. *Let A be a set of vertices in a coloured digraph D. We say that A is (k, d)-rainbow connected in D if, for any set of at most $k - 1$ colours S and any vertices $x, y \in A$, there is a rainbow x to y path of length at most d in D internally avoiding colours in S.*

Notice that in the above definition, we did not specify whether the colouring was a edge colouring, vertex colouring, or total colouring. The definition makes sense in all three cases. For edge colourings a path P "internally avoiding colours in S" means P not having edges having colours in S. For vertex colourings a path P "internally avoiding

colours in S" means P not having vertices having colours in S (except possibly for the vertices x and y). For total colourings a path P "internally avoiding colours in S" means P having no edges or vertices with colours in S (except possibly for the vertices x and y).

Comparing the above definition to "rainbow k-edge-connectedness" defined in the previous section we see that an edge coloured graph is rainbow k-connected exactly when it is (k, ∞)-rainbow connected.

We'll need the following lemma which could be seen as a weak analogue of Menger's Theorem. It will allow us to find rainbow paths through prescribed vertices in a highly connected set.

Lemma 2.3. *Let D be a totally coloured digraph and A a $(3kd, d)$-rainbow connected subset of D. Let S be a set of colours with $|S| \leqslant k$ and $a_1, \ldots, a_k$ be vertices in A such that no a_i has a colour from S and $a_1, \ldots, a_k$ all have different colours.*

Then there is a rainbow path P from a_1 to a_k of length at most kd which passes through each of $a_1, \ldots, a_k$ and avoids S

Proof. Using the definition of $(3kd, d)$-rainbow connected, there is a rainbow path P_1 from a_1 to a_2 of length $\leqslant d$ avoiding colours in S. Similarly for $i \leqslant k$, since the total number of colours in S together with $P_1 + \cdots + P_{i-1}$ is $k + 2d(i-1) + 1 \leqslant 3kd$ there is a rainbow path P_i from a_i to a_{i+1} of length $\leqslant d$ internally avoiding the k colours in S and the $2d(i-1)+1$ colours in $P_1 + \cdots + P_{i-1}$. Joining the paths $P_1, \ldots, P_{k-1}$ gives the required path. □

To every coloured directed graph we associate an uncoloured directed graph where two vertices are joined whenever they have a lot of short paths between them.

Definition 2.4. *Let D be a totally coloured digraph and $m \in \mathbb{N}$. We define an uncoloured directed graph D_m as follows. The vertex set of D_m is $V(D)$, and xy is an edge of D_m whenever there are m internally vertex disjoint paths $P_1, \ldots, P_m$, each of length 2 and going from x to y such that $P_1 \cup \cdots \cup P_m$ is rainbow.*

It turns out that for properly coloured directed graphs D, the uncoloured graph D_m has almost the same minimum degree as D. The following lemma will allow us to study short rainbow paths in coloured graphs by first proving a result about short paths in uncoloured graphs.

Lemma 2.5. *For all $\epsilon > 0$ and $m \in \mathbb{N}$, there is an $N = N(\epsilon, m) = (5m+4)/\epsilon^2$ such that the following holds. Let D be a properly totally coloured vertex-rainbow directed graph on at least N vertices. Then we have*

$$\delta^+(D_m) \geqslant \delta^+(D) - \epsilon|D|.$$

Proof. Let v be an arbitrary vertex in D_m. It is sufficient to show that $|N^+_{D_m}(v)| \geqslant |\delta^+(D)| - \epsilon|D|$.

For $w \in V(D)$, we define

$$r_v(w) = \#\text{rainbow paths of length 2 from } v \text{ to } w.$$

Let $W = \{w : r_v(w) \geqslant 5m\}$. The following claim shows that W is large.

Claim 2.6. $|W| \geqslant \delta^+(D) - \epsilon|D|$.

Proof. For any $u \in N_D^+(v)$ we let

$$N'(u) = N_D^+(u) \setminus \{x \in N_D^+(u) : ux \text{ or } x \text{ has the same colour as } v \text{ or } vu\}.$$

Since D is properly coloured, and all the vertices in D have different colours, we have that $|\{x \in N^+(u) : ux \text{ or } x \text{ has the same colour as } v \text{ or } vu\}| \leqslant 4$. This implies that $|N'(u)| \geqslant \delta^+(D) - 4$.

Notice that for a vertex x, we have $x \in N'(u)$ if, and only if, the path vux is rainbow. Indeed vu has a different colour from v and u since the colouring is proper. Similarly ux has a different colour from u and x. Finally ux and x have different colours from v and vu by the definition of $N'(u)$.

Therefore there are $\sum_{u \in N_D^+(v)} |N'(u)|$ rainbow paths of length 2 starting at v i.e. we have $\sum_{x \in V(D)} r_v(x) = \sum_{u \in N_D^+(v)} |N'(u)|$. For any $x \in D$, we certainly have $r_v(x) \leqslant |N^+(v)|$. If $x \notin W$ then we have $r_v(x) < 5m$. Combining these we obtain

$$(|D| - |W|)5m + |W||N_D^+(v)| \geqslant \sum_{x \in V(D)} r_v(x) = \sum_{u \in N^+(v)} |N'(u)| \geqslant |N_D^+(v)|(\delta^+(D) - 4).$$

The last inequality follows from $|N'(u)| \geqslant \delta^+(D) - 4$ for all $u \in N_D^+(v)$. Rearranging we obtain

$$|W| \geqslant \frac{|N_D^+(v)|(\delta^+(D) - 4) - 5m|D|}{|N_D^+(v)| - 5m} \geqslant \delta^+(D) - \frac{5m|D|}{|N_D^+(v)|} - 4 \geqslant \delta^+(D) - (5m+4)\frac{|D|}{\delta^+(D)}.$$

If $(5m+4)/\delta^+(D) \leqslant \epsilon$, then this implies the claim. Otherwise we have $\delta^+(D) < (5m+4)/\epsilon$ which, since $|D| \geqslant N_0 = (5m+4)/\epsilon^2$, implies that $\delta^+(D) \leqslant \epsilon|D|$ which also implies the claim. □

The following claim shows that W is contained in $N_{D_m}^+(v)$.

Claim 2.7. *If $w \in W$, then we have $vw \in E(D_m)$.*

Proof. From the definition of W, we have $5m$ distinct rainbow paths $P_1, \ldots, P_{5m}$ from v to w of length 2. Consider an auxiliary graph G with $V(G) = \{P_1, \ldots, P_{5m}\}$ and $P_iP_j \in E(G)$ whenever $P_i \cup P_j$ is rainbow.

We claim that $\delta(G) \geqslant 5m - 4$. Indeed if for $i \neq j$ we have $P_i = vxw$ and $P_j = vyw$, then, using the fact that the colouring on D is proper and the vertex-rainbow, it is easy to see that the only way $P_i \cup P_j$ could not be rainbow is if one of the following holds:

$$\begin{array}{ll} c(vx) = c(yw) & \qquad c(vx) = c(y) \\ c(vy) = c(xw) & \qquad c(vy) = c(x). \end{array}$$

Thus if $P_i = vxw$ had five non-neighbours $vy_1w, \ldots, vy_5w$ in G, then by the Pigeonhole Principle for two distinct j and k we would have one of $c(y_jw) = c(y_kw)$, $c(y_j) = c(y_k)$,

or $c(vy_j) = c(vy_k)$. But none of these can occur for distinct paths vy_jw and vy_kw since the colouring on D is proper and the vertex-rainbow. Therefore $\delta(G) \geqslant 5m - 4$ holds.

It is easy to see that G has a clique of size at least $|V(G)|/5 = m$ (for example by Turán's Theorem, or Brooks' Theorem.) The union of the paths in this clique is rainbow, showing that $vw \in E(D_m)$. □

Claim 2.7 shows that $W \subseteq N^+_{D_m}(v)$, and so Claim 2.6 implies that $|N^+_{D_m}(v)| \geqslant \delta^+(D) - \epsilon|D|$. Since v was arbitrary, this implies the lemma. □

The following lemma shows that every directed graph with high minimum degree contains a large, highly connected subset.

Lemma 2.8. *For all $\epsilon > 0$ and $k \in \mathbb{N}$, there is a $d = d(\epsilon) = 40\epsilon^{-2}$ and $N = N(\epsilon, k) = 32k\epsilon^{-2}$ such that the following holds. Let D be a directed graph of order at least N. Then there is a (k, d)-connected subset $A \subseteq V(D)$ satisfying*

$$|A| \geqslant \delta^+(D) - \epsilon|D|.$$

Proof. We start with the following claim.

Claim 2.9. *There is a set $\tilde{A} \subseteq V(D)$ satisfying the following*

- *For all $B \subseteq \tilde{A}$ with $|B| > \epsilon|D|/4$ there is a vertex $v \in \tilde{A} \setminus B$ such that $|N^+(v) \cap B| \geqslant \epsilon^2|D|/16$.*
- *$\delta^+(D[\tilde{A}]) \geqslant \delta^+(D) - \epsilon|D|/4$.*

Proof. Let $A_0 = V(D)$. We define $A_1, A_2, \ldots, A_M$ recursively as follows.

- If A_i contains a set B_i such that $|B_i| > \epsilon|D|/4$ and for all $v \in A_i \setminus B_i$ we have $|N^+(v) \cap B_i| < \epsilon^2|D|/16$, then we let $A_{i+1} = A_i \setminus B_i$.
- Otherwise we stop with $M = i$.

We will show that that $\tilde{A} = A_M$ satisfies the conditions of the claim. Notice that by the construction of A_M, it certainly satisfies the first condition. Thus we just need to show that $\delta^+(D[A_M]) \geqslant \delta^+(D) - \epsilon|D|/4$.

From the definition of A_{i+1} we have that $\delta^+(D[A_{i+1}]) \geqslant \delta^+(D[A_i]) - \epsilon^2|D|/16$ which implies $\delta^+(D[A_M]) \geqslant \delta^+(D) - M\epsilon^2|D|/16$. Therefore it is sufficient to show that we stop with $M \leqslant 4\epsilon^{-1}$. This follows from the fact that the sets $B_0, \ldots, B_{M-1}$ are all disjoint subsets of $V(D)$ with $|B_i| > \epsilon|D|/4$. □

Let $\tilde{A}$ be the set given by the above claim. Let $A = \{v \in \tilde{A} : |N^-(v) \cap \tilde{A}| \geqslant \frac{\epsilon}{2}|D|\}$. We claim that A satisfies the conditions of the lemma.

To show that $|A| \geqslant \delta^+(D) - \epsilon|D|$, notice that we have

$$\frac{\epsilon}{2}|D|(|\tilde{A}| - |A|) + |A||\tilde{A}| \geqslant \sum_{v \in \tilde{A}} |N^-(v) \cap \tilde{A}| = \sum_{v \in \tilde{A}} |N^+(v) \cap \tilde{A}| \geqslant |\tilde{A}|(\delta^+(D) - \epsilon|D|/4).$$

The first inequality come from bounding $|N^-(v) \cap \tilde{A}|$ by $\frac{\epsilon}{2}|D|$ for $v \notin A$ and by $|\tilde{A}|$ for $v \in A$. The second inequality follows from the second property of $\tilde{A}$ in Claim 2.9. Rearranging we obtain

$$|A| \geqslant \frac{|\tilde{A}|}{|\tilde{A}| - \epsilon|D|/2}(\delta^+(D) - 3\epsilon|D|/4) \geqslant \delta^+(D) - \epsilon|D|.$$

Now, we show that A is (k,d)-connected in D. As in Definition 2.1, let S be a subset of $V(D)$ with $|S| \leqslant k-1$ and let x, y be two vertices in $A \setminus S$. We will find a path of length $\leqslant d$ from x to y in $\tilde{A} \setminus S$. Notice that since $|D| \geqslant 32k\epsilon^{-2}$, we have $|S| \leqslant \epsilon^2|D|/32$.

Let $N^t(x) = \{u \in \tilde{A} \setminus S : d_{D[\tilde{A}\setminus S]}(x,u) \leqslant t\}$. We claim that for all $x \in \tilde{A}$ and $t \geqslant 0$ we have

$$|N^{t+1}(x)| \geqslant \min(|\tilde{A}| - \epsilon|D|/4, |N^t(x)| + \epsilon^2|D|/32).$$

For $t = 0$ this holds since we have $|N^1| = |\tilde{A}| \geqslant \epsilon|D|/4$. Indeed if $|N^t(x)| < |\tilde{A}| - \epsilon|D|/4$ holds for some t and x, then letting $B = \tilde{A} \setminus N^t(x)$ we can apply the first property of $\tilde{A}$ from Claim 2.9 in order to find a vertex $u \in N^t(x)$ such that $|N^+(u) \cap (\tilde{A} \setminus N^t(x))| \geqslant \epsilon^2|D|/16$. Using $|S| \leqslant \epsilon^2|D|/32$ we get $|(N^+(u) \setminus S) \cap (\tilde{A} \setminus N^t(x))| \geqslant |N^+(u) \cap (\tilde{A} \setminus N^t(x))| - |S| \geqslant \epsilon^2|D|/32$. Since $(N^+(u) \cap \tilde{A} \setminus S) \cup N^t(x) \subseteq N^{t+1}(x)$, we obtain $|N^{t+1}(x)| \geqslant |N^t(x)| + \epsilon^2|D|/32$.

Thus we obtain that $|N^t(x)| \geqslant \min(|\tilde{A}| - \epsilon|D|/4, t\epsilon^2|D|/32)$. Since $(d-1)\epsilon^2/32 > 1$, we have that $|N^{d-1}(x)| \geqslant |\tilde{A}| - \epsilon|D|/4$. Recall that from the definition of A, we also have also have $|N^-(y) \cap \tilde{A}| \geqslant \epsilon|D|/2$. Together these imply that $N^-(y) \cap N^{d-1}(x) \neq \varnothing$ and hence there is a $x - y$ path of length $\leqslant d$ in $\tilde{A} \setminus S$. □

The following is a generalization of the previous lemma to coloured graphs. This is the main intermediate lemma we need in the proof of Theorem 1.5.

Lemma 2.10. *For all $\epsilon > 0$ and $k \in \mathbb{N}$, there is an $d = d(\epsilon) = 1280\epsilon^{-2}$ and $N = N(\epsilon, k) = 1800k\epsilon^{-4}$ such that the following holds.*

Let D be a properly totally coloured vertex-rainbow directed graph on at least N vertices. Then there is a (k,d)-rainbow connected subset $A \subseteq V(D)$ satisfying

$$|A| \geqslant \delta^+(D) - \epsilon|D|.$$

Proof. Set $m = 9d + 3k$, and consider the directed graph D_m as in Definition 2.4. Using $|D| \geqslant 1800k\epsilon^{-4}$, we can apply Lemma 2.5 with the constant $\epsilon/4$ we have that $\delta^+(D_m) \geqslant \delta^+(D) - \epsilon|D|/4$.

Apply Lemma 2.8 to D_m with the constants $\epsilon/4$, and k. This gives us a $(k, d/2)$-connected set A in D_m with $|A| \geqslant \delta^+(D_m) - \epsilon|D_m|/4 \geqslant \delta^+(D) - \epsilon|D|/2$. We claim that A is (k,d)-rainbow connected in D. As in Definition 2.2, let S be a set of k colours and $x, y \in A$. Let S_V be the vertices of D with colours from S. Since D is vertex-rainbow, we have $|S_V| \leqslant k$. Since A is $(k, d/2)$-connected in D_m, there is a $x - y$ path P in $(D_m \setminus S_V) + x + y$ of length $\leqslant d/2$.

Using the property of D_m, for each edge $uv \in P$, there are at least m choices for a triple of three distinct colours (c_1, c_2, c_3) and a vertex $y(uv)$ such that there is a path

$uy(uv)v$ with $c(uy(uv)) = c_1$, $c(y(uv)) = c_2$, and $c(y(uv)v) = c_3$. Since $m \geqslant 9d + 3k \geqslant 6|E(P)| + 3|V(P)| + 3|S|$, we can choose such a triple for every edge $uv \in P$ such that for two distinct edges in P, the triples assigned to them are disjoint, and also distinct from the colours in S and colours of vertices of P.

Let the vertex sequence of P be $u, x_1, x_2, \ldots, x_p, v$. The following sequence of vertices is a rainbow path from u to v of length $2|P| \leqslant d$ internally avoiding colours in S

$$P' = u, y(ux_1), x_1, y(x_1x_2), x_2, y(x_2x_3), x_3, \ldots, x_{p-1}, y(x_{p-1}x_p), x_p, y(x_pv), v.$$

To show that P' is a rainbow path we must show that all its vertices and edges have different colours. The vertices all have different colours since the vertices in D all had different colours. The edges of P' all have different colours from each other and the vertices of P' by our choice of the vertices $y(x_ix_{i+1})$ and the triples of colours associated with them. □

We'll need the following simple lemma which says that for any vertex v there is a set of vertices N^{t_0} close to v with few edges going outside N^{t_0}.

Lemma 2.11. *Suppose we have $\epsilon > 0$ and D a totally coloured directed graph. Let v be a vertex in D and for $t \in \mathbb{N}$, let $N^t(v) = \{x : d_R(v, x) \leqslant t\}$. There is a $t_0 \leqslant \epsilon^{-1}$ such that we have*

$$|N^{t_0+1}(v)| \leqslant |N^{t_0}(v)| + \epsilon|D|.$$

Proof. Notice that if $|N^{t+1}(v)| > |N^t(v)| + \epsilon|D|$ held for all $t \leqslant \epsilon^{-1}$, then we would have $|N^t(v)| > \epsilon t|D|$ for all $t \leqslant \epsilon^{-1}$. When $t = \epsilon^{-1}$ this gives $|N^{\epsilon^{-1}}(v)| > |D|$, which is a contradiction. □

A corollary of the above lemma is that for any vertex v in a properly coloured directed graph, there is a subgraph of D close to v which has reasonably large minimum out-degree.

Lemma 2.12. *Suppose we have $\epsilon > 0$ and D a properly totally coloured vertex-rainbow directed graph on $\geqslant 2\epsilon^{-2}$ vertices. Let v be a vertex in D and $\delta^+ = \min_{x:d_R(v,x)\leqslant\epsilon^{-1}} d^+(x)$. Then there is a set N such that $d_R(v, N) \leqslant \epsilon^{-1}$ and we have*

$$\delta^+(D[N]) \geqslant \delta^+ - 2\epsilon|D|.$$

Proof. Apply Lemma 2.11 to D in order to obtain a number $t_0 \leqslant \epsilon^{-1}$ with $|N^{t_0+1}(v)| \leqslant |N^{t_0}(v)| + \epsilon|D|$. We claim that the set $N = N^{t_0}(v)$ satisfies the conditions of the lemma.

Suppose, for the sake of contradiction that there is a vertex $x \in N^{t_0}(v)$ with $|N^+(x) \cap N^{t_0}(v)| < \delta^+ - 2\epsilon|D|$. Since $\delta^+ \leqslant |N^+(x)|$, we have $|N^+(x) \setminus N^{t_0}(v)| > 2\epsilon|D|$. Let P be a length $\leqslant t_0$ path from v to x. Notice that since the colouring on D is proper and all vertices in D have different colours, the path $P + y$ is rainbow for all except at most $2|P|$ of the vertices $y \in N^+(x)$. Therefore we have $|N^+(x) \setminus N^{t_0+1}(v)| \leqslant 2|P| \leqslant 2\epsilon^{-1}$. Combining this with $|D| \geqslant 2\epsilon^{-2}$, this implies

$$\begin{aligned}|N^{t_0+1}(v)| &\geqslant |N^{t_0}(v)| + |N^+(x) \setminus N^{t_0}(v)| - |N^+(x) \setminus N^{t_0+1}(v)| \\ &> |N^{t_0}(v)| + 2\epsilon|D| - 2\epsilon^{-1} \\ &\geqslant |N^{t_0}(v)| + \epsilon|D|.\end{aligned}$$

This contradicts the choice of t_0 in Lemma 2.11. □

3 Proof of Theorem 1.5

The goal of this section is to prove an approximate version of Conjecture 1.3 in the case when all the matchings in G are disjoint. The proof will involve considering auxiliary directed graphs to which Lemmas 2.10 and 2.12 will be applied.

We begin this section by proving a series of lemmas (Lemmas 3.3 – 3.8) about bipartite graphs consisting of a union of n_0 matchings. The set-up for these lemmas will always be the same, and so we state it in the next paragraph to avoid rewriting it in the statement of every lemma.

We will always have bipartite graph called "G" with bipartition classes X and Y consisting of $n+1$ edge-disjoint matchings $M_1, \ldots, M_{n+1}$. These matchings will be referred to as colours, and the colour of an edge e means the matching e belongs to. There will always be a rainbow matching called M of size n in G. We set $X_0 = X \setminus V(M)$ and $Y_0 = Y \setminus V(M)$. The colour missing from M will denoted by c^*.

Notice that for any edge e, there is a special colour (the colour c_e of the edge e) as well as a special vertex in X (i.e. $e \cap X$) and in Y (i.e. $e \cap Y$). In what follows we will often want to refer to the edge e, the colour c_e, and the vertices $e \cap X$ and $e \cap Y$ interchangeably. To this end we make a number of useful definitions:

- For an edge e, we let $(e)_C$ be the colour of e, $(e)_X = e \cap X$, and $(e)_Y = e \cap Y$.
- For a vertex $x \in X$, we let $(x)_M$ be the edge of M passing through x (if it exists), $(x)_C$ the colour of $(x)_M$, and $(x)_Y$ the vertex $(x)_M \cap Y$. If there is no edge of M passing through x, then $(x)_M$, $(x)_C$, and $(x)_Y$ are left undefined.
- For a vertex $y \in Y$, we let $(y)_M$ be the edge of M passing through y (if it exists), $(y)_C$ the colour of $(y)_M$, and $(y)_X$ the vertex $(y)_M \cap X$. If there is no edge of M passing through y, then $(y)_M$, $(y)_C$, and $(y)_X$ are left undefined.
- For a colour c, we let $(c)_M$ be the colour c edge of M (if it exists), $(c)_X = (c)_M \cap X$, and $(c)_Y = (c)_M \cap Y$. For the colour c^*, we leave $(c)_M$, $(c)_X$, and $(c)_Y$ undefined.

For a set S of colours, edges of M, or vertices, we let $(S)_M = \{(s)_M : s \in S\}$, $(S)_X = \{(s)_X : s \in S\}$, $(S)_Y = \{(s)_Y : s \in S\}$, and $(S)_C = \{(s)_C : s \in S\}$. Here S is allowed to contain colours/edges/vertices for which $(*)_M/(*)_X/(*)_Y/(*)_C$ are undefined—in this case $(S)_M$ is just the set of $(s)_M$ for $s \in S$ where $(s)_M$ is defined (and similarly for $(S)_X/(S)_Y/(S)_C$. It is useful to observe that from the above definitions we get identities such as $(((S)_X)_C)_M = S$ for a set S of edges of M.

We will now introduce two important and slightly complicated definitions. Both Definition 3.1 and 3.2 will apply in the setting of a bipartite graph G with bipartition $X \cup Y$ consisting of $n+1$ edge-disjoint matchings, and a rainbow matching M of size n missing colour c^*. The first definition is that of a *switching*—informally this should be thought of as a sequence of edges of $G \setminus M$ which might be exchanged with a sequence of edges of M in order to produce a new rainbow matching of size n. See Figure 1 for an illustration of a switching.

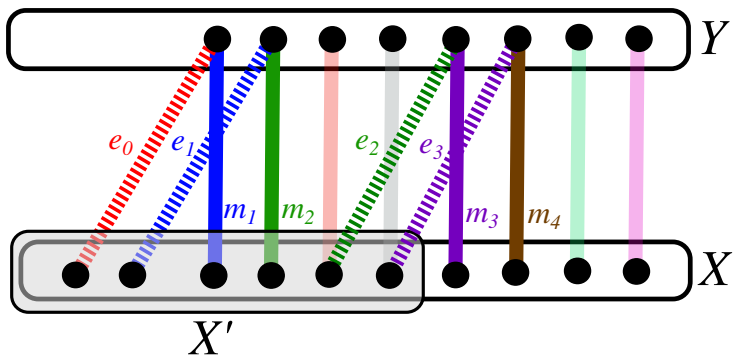

Figure 1: An X'-switching of length 4. The solid lines represent edges of M and the dashed lines represent edges not in M.

Definition 3.1. *Let $X' \subseteq X$. A sequence of edges, $\sigma = (e_0, m_1, e_1, m_2, e_2, \ldots, e_{\ell-1}, m_\ell)$, is an X'-switching if the following hold.*

(i) For all i, m_i is an edge of M and e_i is not an edge of M.

(ii) For all i, m_i and e_i have the same colour, c_i.

(iii) For all i, $e_{i-1} \cap m_i = (m_i)_Y$.

(iv) For all $i \neq j$, we have $e_i \cap e_j = e_{i-1} \cap m_j = \varnothing$ and also $c_i \neq c_j$.

(v) For all i, $(e_i)_X \in X'$.

If σ is a switching defined as above, then we say that σ is a length ℓ switching from c_0 to c_ℓ. Let $e(\sigma) = \{e_0, \ldots, e_{\ell-1}\}$ and $m(\sigma) = \{m_1, \ldots, m_\ell\}$. For a switching σ we define $(\sigma)_X = (e(\sigma))_X \cup (m(\sigma))_X$.

The next definition is that of a *free* subset of X—informally a subset $X' \subset X$ is free if there are matchings M' which "look like" M, but avoid small subsets of X'.

Definition 3.2. *Let $X', T \subseteq X$, $k \in \mathbb{N}$, and c be a colour. We say that X' is (k, T, c)-free if $T \cap X' = \varnothing$, $c \notin (X' \cup T)_C$, and the following holds:*

Let A be any set of at most k edges in $M \setminus ((T)_M \cup (c)_M)$, $B \subseteq X'$ any set of at most k vertices such that $(A)_X \cap B = \varnothing$. Then there is a rainbow matching M' of size n satisfying the following:

- *M' agrees with M on A.*
- *$(M')_X \cap B = \varnothing$.*
- *M' misses the colour c.*

In other words, M' replaces all edges touching B by edges not touching B, whilst maintaining the edges of A and avoiding the colour c. It is worth noticing that X_0 is $(n, \varnothing, c^*)$-free (always taking the matching M' to be M in the definition).

The following lemma is crucial—it combines the preceding two definitions together and says that if we have an X'-switching σ for a free set X', then there is a new rainbow matching of size n which avoids $(m(\sigma))_X$.

Lemma 3.3. *Suppose that X' is $(2k, T, c)$-free and $\sigma = (e_0, m_1, e_1, \ldots, e_{\ell-1}, m_\ell)$ is an X'-switching from c to $(m_\ell)_C$ of length $\ell \leqslant k$. Let A be any set of at most k edges in $M - (c)_M$ and let B be any subset of X' of order at most k. Suppose that the following disjointness conditions hold.*

$$(\sigma)_X \cap T = \varnothing \qquad (\sigma)_X \cap (A)_X = \varnothing \qquad (\sigma)_X \cap B = \varnothing$$
$$T \cap (A)_X = \varnothing \qquad (A)_X \cap B = \varnothing.$$

Then there is a rainbow matching $\tilde{M}$ of size n in G which misses colour $(m_\ell)_C$, agrees with M on A, and has $(\tilde{M})_X \cap (m(\sigma))_X = (\tilde{M})_X \cap B = \varnothing$.

Proof. We let $A' = m(\sigma) \cup A$ and $B' = (e(\sigma))_X \cup B$. Notice that we have $|A'|, |B'| \leqslant 2k$. Also from the definition of "switching", we have that for any i and j, the edges e_i and m_j never intersect in X which together with $(A)_X \cap (\sigma)_X = \varnothing$, $B \cap (\sigma)_X = \varnothing$, and $(A)_X \cap B = \varnothing$ implies that $(A')_X \cap B' = \varnothing$. Also, $(\sigma)_X \cap T = \varnothing$ and $(A)_X \cap T = \varnothing$ imply that $A' \cap (T)_M = \varnothing$. Since σ is a switching starting at c we have $A' \subseteq M \setminus ((T)_M \cup (c)_M)$. Finally, σ being an X'-switching and $B \subseteq X'$ imply that $B' \subseteq X'$.

Therefore we can invoke the definition of X' being $(2k, T, c)$-free in order to obtain a rainbow matching M' of size n avoiding B', agreeing with M on A', and missing colour $c = (e_0)_C$. We let

$$\tilde{M} = (M' \setminus m(\sigma)) \cup e(\sigma) = M' + e_0 - m_1 + e_1 - m_2 + e_2 - \cdots + e_{\ell-1} - m_\ell.$$

We claim that $\tilde{M}$ is a matching which satisfies all the conditions of the lemma.

Recall that $B' \supseteq (e(\sigma))_X$, $A' \supseteq m(\sigma)$, and $(A')_X \cap B' = \varnothing$. Since M' agreed with M on A' and was disjoint from B', we get $m(\sigma) \subseteq M'$ and $e(\sigma) \cap (M' \setminus m(\sigma)) = \varnothing$. This implies that $\tilde{M}$ is a set of n edges and also that $(\tilde{M})_X = \big((M')_X \setminus (m(\sigma))_X\big) \cup (e(\sigma))_X$ is a set of n vertices. Finally notice that since $(e_i)_Y = (m_{i+1})_Y$ we have $(\tilde{M})_Y = (M')_Y$. Thus $\tilde{M}$ is a set of n edges with n vertices in each of X and Y i.e. a matching. The matching $\tilde{M}$ is clearly rainbow, missing the colour $(m_\ell)_C$ since m_i and e_i always have the same colour.

To see that $\tilde{M}$ agrees with M on edges in A, notice that M' agreed with M on these edges since we had $A \subseteq A'$. Since $(\sigma)_X \cap (A)_X = \varnothing$ implies that σ contains no edges of A, we obtain that $\tilde{M}$ agrees with M on A.

To see that $\tilde{M} \cap (m(\sigma))_X = \varnothing$, recall that $(\tilde{M})_X = \big((M')_X \setminus (m(\sigma))_X\big) \cup (e(\sigma))_X$ and $(m(\sigma))_X \cap (e(\sigma))_X = \varnothing$. Finally, $\tilde{M} \cap B = \varnothing$ follows from $M' \cap B = \varnothing$, $(\tilde{M})_X = \big((M')_X \setminus (m(\sigma))_X\big) \cup (e(\sigma))_X$, and $B \cap (\sigma)_X = \varnothing$. $\square$

We study X'-switchings by looking at an auxiliary directed graph. For any $X' \subseteq X$, we will define a directed, totally labelled graph $D_{X'}$. We call $D_{X'}$ a "labelled" graph rather than a "coloured" graph just to avoid confusion with the coloured graph G. Of

course the concepts of "coloured" and "labelled" graphs are equivalent, and we will freely apply results from Section 2 to labelled graphs. The vertices and edges of $D_{X'}$ will be labelled by elements of the set $X \cup \{*\}$.

Definition 3.4. *Let X' be a subset of X. The directed graph $D_{X'}$ is defined as follows:*

- *The vertex set of $D_{X'}$ is the set of colours of edges in G. For any colour $v \in V(D_{X'})$ present in M, v is labelled by "$(v)_X$". The colour c^* is labelled by "$*$".*
- *For two colours u and $v \in V(D_{X'})$, there is a directed edge from u to v in $D_{X'}$ whenever there is an $x \in X'$ such that there is a colour u edge from x to the vertex $(v)_Y$ in G. In this case uv is labelled by "x".*

Notice that in the second part of this definition the labelling is well-defined since there cannot be colour u edges from two distinct vertices x and x' to $(v)_Y$ (since the colour u edges form a matching in G).

Recall that a total labelling is proper if outgoing edges at a vertex always have different labels, ingoing edges at a vertex always have different labels, adjacent vertices have different labels, and an edge always has different labels from its endpoints. Using the fact that the matchings in G are disjoint we can show that $D_{X'}$ is always properly labelled.

Lemma 3.5. *For any $X' \subseteq X$ the total labelling on $D_{X'}$ is always proper. In addition $D_{X'}$ is vertex-rainbow.*

Proof. Suppose that uv and $u'v'$ are two distinct edges of $D_{X'}$ with the same label $x \in X'$. By definition of $D_{X'}$ they correspond to two edges $x(v)_Y$ and $x(v')_Y$ of G having colours u and u' respectively. This implies that u and u' are different since otherwise we would have two edges of the same colour leaving x in G (which cannot happen since colour classes in G are matchings). We also get that v and v' are distinct since otherwise we would have edges of colours both u and u' between x and $(v)_Y$ in G (contradicting the matchings forming G being disjoint).

Let uv be an edge of $D_{X'}$ labelled by x and $x(v)_Y$ the corresponding colour u edge of G. Then u cannot be labelled by "x" (since that would imply that the colour u edge at x would end at $(u)_Y$ rather than $(v)_Y$), and v cannot be labelled by "x" (since then there would be edges from x to $(v)_Y$ in G of both colours u and v).

The fact that D'_X is vertex-rainbow holds because M being a matching implies that $(c)_X$ is distinct for any colour c. □

Recall that a path in a totally labelled graph is defined to be rainbow whenever all its vertices and edges have different colours. The reason we defined the directed graph $D_{X'}$ is that rainbow paths in $D_{X'}$ correspond exactly to X'-switchings in G. Let $P = v_0, \ldots, v_\ell$ be a path in $D_{X'}$ for some X'. For each $i = 0, \ldots, \ell - 1$ let e_i be the colour v_i edge of G corresponding to the edge $v_i v_{i+1}$ in $D_{X'}$. We define σ_P to be the sequence of edges $(e_0, (v_1)_M, e_1, (v_2)_M, e_2, \ldots, (v_{\ell-1})_M, e_{\ell-1}, (v_\ell)_M)$. Notice that $(e(\sigma_P))_X$ is the set of labels of edges in P, and $(m(\sigma_P))_X$ is the set of labels of vertices in $P - v_0$.

The following lemma shows that if P is rainbow then σ_P is a switching.

Lemma 3.6. *Let $P = v_0, \ldots, v_\ell$ be a rainbow path in $D_{X'}$ for some $X' \subseteq X$. Then σ_P is an X'-switching from v_0 to v_ℓ of length ℓ.*

Proof. As in the definition of σ_P, let e_i be the colour v_i edge of G corresponding to the edge $v_i v_{i+1}$ in $D_{X'}$.

We need to check all the parts of the definition of "X'-switching". For part (i), notice that $(v_1)_M, \ldots, (v_\ell)_M$ are edges of M by definition of $(.)_M$, whereas e_i cannot be the colour v_i matching edge $(v_i)_M$ since $(e_i)_Y = (v_{i+1})_M \cap Y$ which is distinct from $(v_i)_M \cap Y$. Parts (ii), (iii), and (v) follow immediately from the definition of e_i and the graph $D_{X'}$.

Part (iv) follows from the fact that P is a rainbow path. Indeed to see that for $i \neq j$ we have $e_i \cap e_j = \varnothing$, notice that $e_i \cap e_j \cap X = \varnothing$ since $v_i v_{i+1}$ and $v_j v_{j+1}$ have different labels in $D_{X'}$, and that $e_i \cap e_j \cap Y = \varnothing$ since $(e_i)_Y \in (v_{i+1})_M$, $(e_j)_Y \in (v_{j+1})_M$, and $(v_{i+1})_M \cap (v_{j+1})_M = \varnothing$. Similarly for $i \neq j$, $e_{i-1} \cap (v_j)_M \cap X = \varnothing$ since $v_{i-1} v_i$ and v_j have different labels in $D_{X'}$, and $e_{i-1} \cap (v_j)_M \cap Y = \varnothing$ since $(e_{i-1})_Y \in (v_i)_M$ and $(v_i)_M \neq (v_j)_M$. Finally, $c_i \neq c_j$ since $v_0, \ldots, v_\ell$ are distinct. □

Although it will not be used in our proof, it is worth noticing that the converse of Lemma 3.6 holds i.e. to every X'-switching σ there corresponds a unique rainbow path P in $D_{X'}$ such that $\sigma = \sigma_P$.

So far all our lemmas were true regardless whether the rainbow matching M was maximum or not. Subsequent lemmas will assume that M is maximum. The following lemma shows that for a free set X', vertices in $D_{X'}$ have large out-degree.

Lemma 3.7. *Suppose that G has at least $(1+\epsilon)n$ edges of each colour and no rainbow matching of size $n+1$. Let X', T, k and c be such that X' is $(2k, T, c)$-free. Let $D = D_{X'} \setminus (T)_C$, v a vertex of D, and P a rainbow path in D from c to v of length at most k. Then we have*

$$|N_D^+(v)| \geqslant (1+\epsilon)n + |X'| - |X| - 2|P| - |T|.$$

Proof. Notice that since P is contained in $D_{X'} \setminus (T)_C$ and since X' being $(2k, T, c)$-free implies $X' \cap T = \varnothing$, we can conclude that $(\sigma_P)_X \cap T = \varnothing$.

Therefore, Lemma 3.3 applied with $A = \varnothing$ implies that for any $B \subseteq X'$ with $|B| \leqslant k$ and $B \cap (\sigma_P)_X = \varnothing$, there is a rainbow matching M' of size n which is disjoint from B and misses colour v. Since there are no rainbow matchings of size $n+1$ in G this means that there are no colour v edges from $X' \setminus (\sigma_P)_X$ to Y_0. Indeed if such an edge xy existed, then we could apply Lemma 3.3 with $B = \{x\}$ in order to obtain a rainbow matching M' missing colour v and vertex x which can be extended to a rainbow $n+1$ matching by adding the edge xy.

We claim that there are at least $(1+\epsilon)n + |X'| - |X| - 2|P|$ colour v edges from $X' \setminus (\sigma_P)_X$. Indeed out of the $(1+\epsilon)n$ colour v edges in G at most $|X| - |X'|$ of them can avoid X', and at most $2|P|$ of them can pass through $(\sigma_P)_X$, leaving at least $(1+\epsilon)n - (|X| - |X'|) - 2|P|$ colour v edges to pass through $X' \setminus (\sigma_P)_X$. Since none of these edges can touch Y_0, each of them must give rise to an out-neighbour of v in $D_{X'}$. This shows that $|N_{D_{X'}}^+(v)| \geqslant (1+\epsilon)n + |X'| - |X| - 2|P|$ which implies the result. □

The following lemma is the essence of the proof of Theorem 1.5. It roughly says that given a free set X_1 containing X_0, there is another free set X_2 containing X_0 such that X_2 is much bigger than X_1, but has worse parameter k. The proof of this lemma combines everything in this section with Lemmas 2.3, 2.10 and 2.12 from Section 2.

Lemma 3.8. *Let $k_1, n \in \mathbb{N}$, and $\epsilon \in [0,1]$ be such that $n \geqslant 10^{20}\epsilon^{-8}k_1$ and $k_1 \geqslant 20\epsilon^{-1}$. Set $k_2 = 10^{-6}\epsilon^2 k_1$. Suppose that G has at least $(1+\epsilon)n$ edges of each colour and no rainbow matching of size $n+1$.*

- *Suppose that we have $X_1, T_1 \subseteq X$ and a colour c_1 such that X_1 is (k_1, T_1, c_1)-free and we also have $X_0 \subseteq X_1 \cup T_1$ and $|T_1| \leqslant k_1 - 30\epsilon^{-1}$.*
- *Then there are $X_2, T_2 \subseteq X$ and a colour c_2 such that X_2 is (k_2, T_2, c_2)-free and we also have $X_0 \subseteq X_2 \cup T_2$, $|T_2| \leqslant |T_1| + 30\epsilon^{-1}$ and*

$$|X_2| > |X_1| + \frac{\epsilon}{2}n.$$

Proof. Set $d = 10^5\epsilon^{-2}$. Let $D = D_{X_1} \setminus (T_1)_C$. Recall that Lemma 3.5 implies that D is properly labelled and vertex-rainbow.

Lemma 3.7, together with $n \geqslant 10^{20}\epsilon^{-8}k_1$, $k_1 \geqslant 20\epsilon^{-1}$, and $|T_1| \leqslant k_1$ imply that all vertices in D within rainbow distance $(10\epsilon)^{-1}$ of c_1 satisfy $d^+(v) \geqslant (1+\epsilon)n + |X_1| - |X| - 30\epsilon^{-1} \geqslant (1+0.9\epsilon)n + |X_1| - |X|$.

Lemma 2.12 applied with $\epsilon = 0.1\epsilon$ implies that there is a subgraph D' in D satisfying $\delta^+(D') \geqslant (1+0.7\epsilon)n + |X_1| - |X|$ and $d_R(c_1, v) \leqslant 10\epsilon^{-1}$ for all $v \in D'$. Therefore, using $n \geqslant 10^{20}\epsilon^{-8}k_1$, we can apply Lemma 2.10 to D' with $\epsilon = 0.1\epsilon$ and $k = 9k_2d$ in order to find a set W with $|W| \geqslant (1+0.6\epsilon)n + |X_1| - |X|$ which is $(9k_2d, d)$-rainbow connected in D'.

Since $W \subseteq D'$, there is a path, Q, of length $\leqslant 10\epsilon^{-1}$ from c_1 to some $q \in W$. Let c_2 be any vertex in W with $(c_2)_X \notin (\sigma_Q)_X$. Let $T_2 = T_1 \cup (\sigma_Q)_X \cup (c_1)_X$. Let $X_2 = ((W)_X \cup X_0) \setminus (T_2 \cup (c_2)_X)$. We claim that X_2, T_2, and c_2 satisfy the conclusion of the lemma.

First we show that X_2 is (k_2, T_2, c_2)-free. The facts that $T_2 \cap X_2 = \varnothing$ and $c_2 \notin X_2 \cup T_2$ follow from the construction of X_2, T_2, and c_2. Let A be any set of k_2 edges of $M \setminus ((T_2)_M \cup (c_2)_M)$, and $B \subseteq X_2$ any set of k_2 vertices such that $(A)_X \cap B = \varnothing$. Let $B_{X_0} = B \cap X_0$ and $B_W = B \cap (W)_X = B \setminus B_{X_0}$. By Lemma 2.3, applied with $k = 3k_2$, $d = d$, $A = W$, $\{q, a_1, \ldots, a_k, c_2\} = (B_W)_C$, and $S = (A)_X \cup (\sigma_Q)_X \cup B_{X_0}$, there is a rainbow path P in D' of length $\leqslant 3k_2d$ from q to c_2 which is disjoint from $V(Q-q)$ and $(A)_C$, passes through every colour of $(B_W)_C$, and whose edges and vertices don't have labels in $(A)_X \cup (\sigma_Q)_X \cup B_{X_0} \setminus (q)_X$. Notice that this means that $Q+P$ is a rainbow path from c_1 to c_2.

We apply Lemma 3.3 with $X' = X_1$, $T = T_1$, $c = c_1$, $\sigma = \sigma_{Q+P}$, $A = A$, $B = B_{X_0}$. For this application notice that σ_{Q+P} is an X_1-switching of length $\leqslant k_1/2$, which holds because of Lemma 3.6 and because $2|Q| + 2|P| \leqslant 20\epsilon^{-1} + 2k_2d \leqslant k_1/2$. We also need to check the various disjointness conditions—$(A)_X \cap T_1 = (A)_X \cap (\sigma_{Q+P})_X = (A)_X \cap B_{X_0} = \varnothing$

(which hold because $(A)_X$ was disjoint from T_2, P, and B), $(\sigma_{Q+P})_X \cap T_1 = \varnothing$ (which holds since vertices and edges in D have no labels from T_1), and $(\sigma_{Q+P})_X \cap B_{X_0} = \varnothing$ (which holds since B was disjoint from T_2 and P had no labels from B_{X_0}). Therefore Lemma 3.3 produces a rainbow matching M' of size n which agrees with M on A, avoids $(m(\sigma_{Q+P}))_X \cup B_{X_0}$, and misses colour c_2. Since P passes through every colour in $(B_W)_C$, we have $B_W \subseteq (m(\sigma_{Q+P}))_X$ and so M' avoids all of B. Since A and B were arbitrary, we have shown that X_2 is (k_2, T_2, c_2)-free.

The containment $X_0 \subseteq X_2 \cup T_2$ holds because $X_0 \subseteq X_1 \cup T_1 \subseteq X_2 \cup T_2$. Notice that $|T_2| \leqslant |T_1| + 30\epsilon^{-1}$ follows from $|Q| \leqslant 10\epsilon^{-1}$.

Finally, $|X_2| > |X_1| + \epsilon n/2$ holds because since $(W)_X$ was disjoint from X_0 we have

$$|X_2| \geqslant |X_0| + |W| \geqslant |X_0| + (1 + 0.6\epsilon)n + |X_1| - |X| = |X_1| + 0.6\epsilon n. \qquad \square$$

We are finally ready to prove Theorem 1.5. The proof consists of starting with X_0 and applying Lemma 3.8 repeatedly, at each step finding a free set X_i which is $\epsilon n/2$ bigger than X_{i-1}. This clearly cannot be performed more than $2\epsilon^{-1}$ times (since otherwise it would contradict $|X_i| \leqslant |X| = |X_0| + n$), and hence the "there is no rainbow matching in G of size $n+1$" clause of Lemma 3.8 could not be true.

Proof of Theorem 1.5. Let G be a bipartite graph which is the union of $n_0 \geqslant N$ disjoint matchings each of size at least $(1+\epsilon)n_0$. Let M be the largest rainbow matching in G and c^* the colour of any matching not used in M. Let n be the number of edges of M. Since M is maximum, Lemma 1.6 tells us that $n \geqslant N/2$. Let $X_0 = X \setminus M$ and $Y_0 = Y \setminus M$. Suppose for the sake of contradiction that $n < n_0$.

Let $T_0 = \varnothing$, $k_0 = (10^{-6}\epsilon^{-2})^{2\epsilon^{-1}}$, and $c_0 = c^*$. Notice that since X_0 is (n, T_0, c_0)-free and $n \geqslant N/2 \geqslant k_0$ we get that X_0 is (k_0, T_0, c_0)-free. For $i = 1, \ldots, 2\epsilon^{-1}$, we set $k_i = 10^{-6}\epsilon^2 k_{i-1}$.

For $i = 0, \ldots, 2\epsilon^{-1}$ we repeatedly apply Lemma 3.8 to X_i, k_i, T_i, c_i in order to obtain sets X_{i+1}, $T_{i+1} \subseteq X$ and a colour c_{i+1} such that X_{i+1} is $(k_{i+1}, T_{i+1}, c_{i+1})$-free, $X_0 \subseteq X_{i+1} \cup T_{i+1}$, $|T_{i+1}| \leqslant |T_i| + 30\epsilon^{-1}$, and $|X_{i+1}| > |X_i| + \epsilon n/2$. To see that we can repeatedly apply Lemma 3.8 this way we only need to observe that there are no rainbow $n+1$ matchings in G, and that for $i \leqslant 2\epsilon^{-1}$ we always have $n \geqslant 10^{20}\epsilon^{-8}k_i$, $k_i \geqslant 10\epsilon^{-1}$, and $|T_i| \leqslant 30\epsilon^{-1}i \leqslant k_i - 30\epsilon^{-1}$.

But now we obtain that $|X_{2\epsilon^{-1}}| > |X_0| + n = |X|$ which is a contradiction since X_i is a subset of X. $\square$

4 Golden Ratio Theorem

In this section we prove Theorem 1.8. The proof uses Theorem 1.4 as well as Lemma 2.11.

Proof of Theorem 1.8. The proof is by induction on n. The case "$n = 1$" is trivial since here G is simply a matching. Suppose that the theorem holds for all G which are unions of $< n$ matchings. Let G be a graph which is the union of n matchings each of size $\phi n + 20n/\log n$. Suppose that G has no rainbow matching of size n. Let M be a maximum

rainbow matching in G. By induction we can suppose that $|M| = n - 1$. Let c^* be the missing colour in M.

Let $X_0 = X \setminus V(M)$ and $Y_0 = Y \setminus V(M)$. Notice that for any colour c there are at least $(\phi - 1)n + 20n/\log n$ colour c edges from X_0 to Y and at least $(\phi - 1)n + 20n/\log n$ colour c edges from Y_0 to X. If $n < 10^6$, then this would give more than n colour c^* edges from X_0 to Y, one of which could be added to M to produce a larger matching. Therefore, we have that $n \geqslant 10^6$.

We define an edge-labelled directed graph D whose vertices are the colours in G, and whose edges are labelled by vertices from $X_0 \cup Y_0$. We set cd an edge in D with label $v \in X_0 \cup Y_0$ whenever there is a colour c edge from v to the colour d edge of M. Notice that D is out-proper—indeed if edges ux and $uy \in E(D)$ had the same label $v \in X_0 \cup Y_0$, then they would correspond to two colour u edges touching v in G (which cannot happen since the colour classes of G are matchings).

Recall that $d_R(x, y)$ denotes the length of the shortest rainbow x to y path in D.

We'll need the following two claims.

Claim 4.1. *For every $c \in V(D)$, there are at most $d_R(c^*, c)$ colour c edges between X_0 and Y_0.*

Proof. Let $P = c^* p_1 \dots p_k c$ be a rainbow path of length $d_R(c^*, c)$ from c^* to c in D. For each i, let m_i be the colour p_i edge of M, and let e_i be the colour p_i edge from the label of $p_i p_{i+1}$ to m_{i+1}. Similarly, let e_{c^*} be the colour c^* edge from the label of $c^* p_1$ to m_1, and let m_c be the colour c edge of M. If there are more than $d_R(c^*, c)$ colour c edges between X_0 and Y_0, then there has to be at least one such edge, e_c, which is disjoint from $e_{c^*}, e_1, \dots, e_k$. Let

$$M' = M + e_{c^*} - m_1 + e_1 - m_2 + e_2 \cdots - m_{k-1} + e_{k-1} - m_c + e_c.$$

The graph M' is clearly a rainbow graph with n edges. We claim that it is a matching. Distinct edges e_i and e_j satisfy $e_i \cap e_j = \varnothing$ since P is a rainbow path. The edge e_i intersects $V(M)$ only in one of the vertices of m_i, which are not present in M'. This means that M' is a rainbow matching of size n contradicting our assumption that M was maximum. □

Claim 4.2. *There is a set $A \subseteq V(D)$ containing c^* such that for all $a \in A$ we have $|N^+(a) \setminus A| \leqslant n/\log n$ and $d_R(c, a) \leqslant \log n$.*

Proof. This follows by applying Lemma 2.11 to D with $\epsilon = (\log n)^{-1}$. □

Let A be the set of colours given by the above claim. Let M' be the submatching of M consisting of the edges with colours not in A. Since $c^* \in A$, we have $|M'| + |A| = n$.

Let A_X be the subset of X spanned by edges of M with colours from A, and A_Y be the subset of Y spanned by edges of M with colours from A. Claim 4.1 shows that for any $a \in A$ there are at most $\log n$ colour a edges between X_0 and Y_0. Therefore there are at least $(\phi - 1)n + 20n/\log n - \log n$ colour a edges from X_0 to $Y \cap (M)_Y$. Using the property of A from Claim 4.2 we obtain that there are at least $(\phi - 1)n + 19n/\log n - \log n$ colour a

edges from X_0 to A_Y. Similarly, for any $a \in A$ we obtain at least $(\phi-1)n+19n/\log n-\log n$ colour a edges from Y_0 to A_X.

By applying Theorem 1.4 to the subgraph of G consisting of the colour A edges between X_0 and A_Y we can find a subset $A_0 \subseteq A$ and a rainbow matching M_0 between X_0 and A_Y using exactly the colours in A_0 such that we have

$$\begin{aligned}|A_0| &\geqslant (\phi-1)n + 19n/\log n - \log n - \sqrt{(\phi-1)n + 19n/\log n - \log n} \\ &\geqslant (\phi-1)n - 6\sqrt{n}\end{aligned}$$

Let $A_1 = A \backslash A_0$. We have $|A_1| \leqslant n-|A_0| \leqslant (2-\phi)n+6\sqrt{n}$. Recall that for each $a \in A_1$ there is a colour a matching between Y_0 and A_X of size at least $(\phi-1)n+19n/\log n-\log n$. Notice that the following holds

$$\begin{aligned}(\phi-1)n + \frac{19n}{\log n} - \log n &\geqslant \phi((2-\phi)n + 6\sqrt{n}) + \frac{20((2-\phi)n + 6\sqrt{n})}{\log((2-\phi)n + 6\sqrt{n})} \\ &\geqslant \phi|A_1| + \frac{20|A_1|}{\log|A_1|}.\end{aligned}$$

The first inequality follows from $\phi^2 - \phi - 1 = 0$ as well as some simple bounds on $\sqrt{n}$ and $\log n$ for $n \geqslant 10^6$. The second inequality holds since $x/\log x$ is increasing.

By induction there is a rainbow matching M_1 between Y_0 and A_X using exactly the colours in A_1. Now $M' \cup M_0 \cup M_1$ is a rainbow matching in G of size n. □

5 Concluding remarks

Here we make some concluding remarks about the techniques used in this paper.

Analogues of Menger's Theorem for rainbow k-edge-connectedness

One would like to have a version of Menger's Theorem for rainbow k-edge-connected graphs as defined in the introduction. In this section we explain why the most natural analogue fails to hold.

Consider the following two properties in an edge coloured directed graph D and a pair of vertices $u, v \in D$.

(i) For any set of $k-1$ colours S, there is a rainbow u to v path P avoiding colours in S.

(ii) There are k edge-disjoint u to v paths $P_1, \ldots, P_k$ such that $P_1 \cup \cdots \cup P_k$ is rainbow.

The most natural analogue of Menger's Theorem for rainbow k-edge-connected graphs would say that for any graph we have (i) $\iff$ (ii). One reason this would be a natural analogue of Menger's Theorem is that there is fractional analogue of the statement (i) $\iff$ (ii). We say that a rainbow path P contains a colour c if P has a colour c edge.

Proposition 5.1. *Let D be a edge coloured directed graph, u and v two vertices in D, and k a real number. The following are equivalent.*

(a) For any assignment of non-negative real numbers y_c to every colour c, satisfying $\sum_{c \text{ a colour}} y_c < k$, there is a rainbow u to v path P with $\sum_{c \text{ contained in } P} y_c < 1$.

(b) We can assign a non-negative real number x_P to every rainbow u to v path P, such that for any colour c we have $\sum_{P \text{ contains } c} x_P \leqslant 1$ and also $\sum_{P \text{ a rainbow } u \text{ to } v \text{ path}} x_P \geqslant k$.

Proof. Let k_a be the minimum of $\sum_{c \text{ a colour}} y_c$ over all choices of non-negative real numbers y_c satisfying $\sum_{c \text{ contained in } P} y_c \geqslant 1$ for all u to v paths P. Similarly, we let k_b be the maximum of $\sum_{P \text{ a rainbow } u \text{ to } v \text{ path}} x_P$ over all choices of non-negative real numbers x_P satisfying $\sum_{P \text{ contains } c} x_P \leqslant 1$ for all colours c.

It is easy to see that k_a and k_b are solutions of two linear programs which are dual to each other. Therefore, by the strong duality theorem (see [9]) we have that $k_a = k_b$ which implies the proposition. □

The reason we say that Proposition 5.1 is an analogue of the statement "(i) $\iff$ (ii)" is that if the real numbers y_c and x_P were all in $\{0, 1\}$ then (a) would be equivalent to (i) and (b) would be equivalent to (ii) (this is seen by letting $S = \{c : y_c = 1\}$ and $\{P_1, \ldots, P_k\} = \{P : x_P = 1\}$).

Unfortunately (i) does not imply (ii) in a very strong sense. In fact even if (ii) was replaced by the weaker statement "there are k edge-disjoint rainbow u to v paths", then (i) would still not imply (ii).

Proposition 5.2. *For any k there is a coloured directed graph D_k with two vertices u and v such that the following hold.*

(I) For any set of k colours S, there is a rainbow u to v path P avoiding colours in S.

(II) Any pair P_1, P_2 of rainbow u to v paths have a common edge.

Proof. We will construct a multigraph having the above property. It is easy to modify the construction to obtain a simple graph. Fix $m > 2k + 1$. The vertex set of D is $\{x_0, \ldots, x_m\}$ with $u = x_0$ and $v = x_m$. For each $i = 0, \ldots, m - 1$, D has $k + 1$ copies of the edge $x_i x_{i+1}$ appearing with colours i, $m + 1$, $m + 2$, $\ldots$, $m + k$. In other words G is the union of $k + 1$ copies of the path $x_0 x_1 \ldots x_m$ one of which is rainbow, and the rest monochromatic.

Notice that D satisfies (II). Indeed if P_1 and P_2 are u to v paths, then they must have vertex sequence $x_0 x_1 \ldots x_m$. Since there are only $m + k$ colours in D both P_1 and P_2 must have at least $m - k$ edges with colours from $\{0, \ldots, m - 1\}$. By the Pigeonhole Principle, since $2(m - k) > m$, there is some colour $i \in \{0, \ldots, m\}$ such that both P_1 and P_2 have a colour i edge. But the only colour i edge in D is $x_i x_{i+1}$ which must therefore be present in both P_1 and P_2. □

There is another, more subtle, reason why (i) does not imply (ii). Indeed if we had "(i) $\implies$ (ii)" then this would imply that every bipartite graph consisting of n matchings of size n contains a rainbow matching of size n.

Indeed given a bipartite graph G with bipartition $X \cup Y$ consisting of n matchings of size n construct an auxiliary graph G' by adding two vertices u and v to G with all edges from u to X and from Y to v present. These new edges all receive different colours which were not in G. It is easy to see for any set S of $n-1$ colours, there is a rainbow u to v path in G' i.e. (i) holds for this graph with $k = n$. In addition, for a set of paths $P_1, \ldots, P_t$ with $P_1 \cup \cdots \cup P_t$ rainbow, it is easy to see that $\{P_1 \cap E(G), \ldots, P_t \cap E(G)\}$ is a rainbow matching in G of size t.

Therefore if "(i) $\implies$ (ii)" was true then we would have a rainbow matching in G of size n. However, as noted in the introduction, there exist Latin squares without transversals, and hence bipartite graphs consisting of n matchings of size n containing no rainbow matching of size n.

The above discussion has hopefully convinced the reader that the natural analogue of Menger's Theorem for rainbow k-edge-connectedness is not true. Nevertheless, it would be interesting to see if any statements about connectedness carry over to rainbow k-edge-connected graphs.

Improving Theorem 1.5

One natural open problem is to improve the dependency of N on ϵ in Theorem 1.5. Throughout our proof we made no real attempt to do this. However there is one interesting modification which one can make in order to significantly improve the bound on N which we mention here.

Notice that the directed graphs $D_{X'}$ in Section 1.5 and the directed graph D in Section 1.8 had one big difference in their definition—to define the graphs $D_{X'}$ we only considered edges starting in X, whereas to define the graph D, we considered edges starting from both X_0 and Y_0. It is possible to modify the proof of Theorem 1.5 in order to deal with directed graphs closer to those we used in the proof of Theorem 1.8. There are many nontrivial modifications which need to be made for this to work. However, the end result seems to be that the analogue of Lemma 3.8 only needs to be iterated $O(\log \epsilon^{-1})$ many times (rather than $O(\epsilon^{-1})$ as in the proof of Theorem 1.5). This would lead to an improved bound on N in Theorem 1.5 $N = O\left(\epsilon^{C \log \epsilon}\right)$ for some constant C. In the grand scheme of things, this is still a very small improvement to bound in Theorem 1.5, and so we do not include any further details here. It is likely that completely new ideas would be needed for a major improvement in the bound in Theorem 1.5.

Another desirable improvement to Theorem 1.5 would be to remove the condition that G is simple i.e. prove an approximate version of the Aharoni-Berger Conjecture. The assumption that G is simple appears only once in the proof of Theorem 1.5—it appears in the proof of Lemma 3.5. If G was a multigraph, then the labeling of the corresponding directed graph $D_{X'}$ would not necessarily be proper. To see an example of this, let G be a graph with vertex set $\{x_1, y_1, \ldots, x_n, y_n\}$, such that the edge $x_i y_i$ appears n times, once with each of the colours $1, \ldots n$. For $X' = \{x_1, \ldots, x_n\}$ and M a rainbow matching

of size n, the corresponding directad graph $D_{X'}$ has vertex set $\{1, \ldots, n\}$ and ij an edge labeled by x_j for all $i \neq j$. Thile this directed graph $D_{X'}$ is out-proper, it is not in-proper. Moreover, it can be checked that this $D_{X'}$ doesn't have any rainbow k-edge-connected subsets A with $|A| \geqslant 2$ and $k \geqslant 1$. This is a serious barrier to our proof strategy since it stops most of the machinery from Section 2 from working. This barrier didn't occur in the proof of Theorem 1.8, since in that theorem we only used the fact that $D_{X'}$ contains a subgraph with high minimum degree, rather than a highly connected one.

Acknowledgement

The author would like to thank János Barát for introducing him to this problem as well as Ron Aharoni, Dennis Clemens, Julia Ehrenmüller, and Tibor Szabó for various discussions related to it. The author would also like to thank a careful referee for many suggestions for how to improve this paper. This research was supported by the Methods for Discrete Structures, Berlin graduate school (GRK 1408), and SNSF grant 200021-149111.

References

[1] R. Aharoni and E. Berger. Rainbow matchings in r-partite r-graphs. *Electron. J. Combin.*, 16 (2009), #R119.

[2] R. Aharoni, P. Charbit, and D. Howard. On a generalization of the Ryser-Brualdi-Stein conjecture. *J. Graph Theory*, 78:143–156, 2015.

[3] R. Aharoni, D. Kotlar, and R. Ziv. Representation of large matchings in bipartite graphs. arXiv:1601.00943, 2016.

[4] J. Barát, A. Gyárfás, and G. Sárközy. Rainbow matchings in bipartite multigraphs. *Period. Math. Hungar.*, pages 1–4, 2016.

[5] B. Bollobás. *Modern Graph Theory*. Springer, 1998.

[6] A. E. Brouwer, A. J. de Vries, and R. M. A. Wieringa. A lower bound for the length of partial transversals in a Latin square. *Nieuw Archief Voor Wiskunde*, 26:330–332, 1978.

[7] R. A. Brualdi and H. J. Ryser. *Combinatorial matrix theory*. Cambridge University Press, 1991.

[8] D. Clemens and J. Ehrenmüller. An improved bound on the sizes of matchings guaranteeing a rainbow matching. *Electron. J. Combin.*, 23(2) 2016, #P2.11.

[9] B. Gärtner and J. Matoušek. *Understanding and Using Linear Programming*. Springer, 2006.

[10] R. Häggkvist and A. Johansson. Orthogonal Latin rectangles. *Combin. Probab. Comput.*, 17:519–536, 2008.

[11] P. Hatami and P. W. Shor. A lower bound for the length of a partial transversal in a Latin square. *J. Combin. Theory Ser. A*, 115:1103–1113, 2008.

[12] D. Kotlar and R. Ziv. Large matchings in bipartite graphs have a rainbow matching. *European J. Combin.*, 38:97–101, 2014.

[13] H. Ryser. Neuere probleme der kombinatorik. *Vorträge über Kombinatorik, Oberwolfach*, pages 69–91, 1967.

[14] S. K. Stein. Transversals of Latin squares and their generalizations. *Pacific J. Math.*, 59:567–575, 1975.

[15] I. M. Wanless. Transversals in Latin squares: A survey. In R. Chapman, editor, *Surveys in Combinatorics 2011*. Cambridge University Press, 2011.

[16] D. E. Woolbright. An $n \times n$ latin square has a transversal with at least $n - \sqrt{n}$ distinct symbols. *J. Combin. Theory Ser. A*, 24:235–237, 1978.