

An approximate version of a conjecture of Aharoni and Berger

Alexey Pokrovskiy¹

Department of Mathematics, ETH Zurich, 8092 Zurich, Switzerland.

Abstract

Aharoni and Berger conjectured that in every proper edge-colouring of a bipartite multigraph by n colours with at least $n + 1$ edges of each colour there is a rainbow matching using every colour. This conjecture generalizes a longstanding problem of Brualdi and Stein about transversals in Latin squares. Here an approximate version of the Aharoni-Berger Conjecture is proved—it is shown that if there are at least $n + o(n)$ edges of each colour in a proper n -edge-colouring of a bipartite multigraph then there is a rainbow matching using every colour.

Keywords: Latin squares, rainbow matchings, connectedness

2000 MSC: 05B15, 05C40, 05C70, 05D15

1. Introduction

The research in this paper is motivated by some old problems about transversals in Latin squares. Recall that 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. 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. 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

Email address: dr.alexey.pokrovskiy@gmail.com ()

¹Research supported in part by SNSF grant SNSF grant 200021-175573 and the Methods for Discrete Structures, Berlin graduate school (GRK 1408).

entries such that no two entries are in the same row, same column, or have the same symbol. The study of transversals in Latin squares goes back to Euler who studied orthogonal Latin squares i.e. order n Latin squares which can be decomposed into n disjoint transversals. For a survey of transversals in Latin squares, see [1].

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 [2]. Brualdi conjectured that every Latin square contains a partial transversal of size $n - 1$ (see [3].) Stein independently made the stronger conjecture that every $n \times n$ array filled with n symbols, each appearing exactly n times contains a partial transversal of size $n - 1$ [4]. Because of the similarity of the above two conjectures, the following is often referred to as “the Brualdi-Stein Conjecture”.

Conjecture 1.1 (Brualdi and Stein, [3, 4]). *Every $n \times n$ Latin square has a partial transversal of size $n - 1$.*

In this paper we will study a generalization of the Brualdi-Stein Conjecture to the setting of rainbow matchings in properly coloured bipartite multigraphs. How are these related? 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 consider a Latin square S whose set of symbols is $\{1, \dots, n\}$ with the i, j symbol $S_{i,j}$. To S we associate an edge-colouring of $K_{n,n}$ with the colours $\{1, \dots, n\}$, by setting $V(K_{n,n}) = \{x_1, \dots, x_n, y_1, \dots, y_n\}$ and letting the edge between x_i and y_j receive 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 the Brualdi-Stein Conjecture is equivalent to the statement that “in any proper n -edge-colouring of $K_{n,n}$, there is a rainbow matching of size $n - 1$.” Once the conjecture is phrased in this form, one begins to wonder whether large rainbow matchings should exist in more general coloured graphs. Aharoni and Berger made the following generalization of the Brualdi-Stein Conjecture.

Conjecture 1.2 (Aharoni and Berger, [5]). *Let G be a properly edge-coloured bipartite multigraph with n colours having at least $n + 1$ edges of each colour. Then G has a rainbow matching using every colour.*

This conjecture attracted a lot of attention since it was made. A most natural way of attacking it is to consider graphs which have substantially more than $n + 1$ edges in each colour, and show that such graphs have a rainbow matching using every colour. For example an easy greedy argument shows that every properly edge-coloured bipartite multigraph with n colours and at least $2n$ edges of each colour has a rainbow matching of size n . Indeed, if the largest matching M in such a graph had size $\leq n - 1$, then one of the $2n$ edges of the unused colour would be disjoint from M , and we could get a larger matching by adding it. This simple bound has been successively improved by many authors. Aharoni, Charbit, and Howard [6] proved that matchings of size $\lceil 7n/4 \rceil$ are sufficient to guarantee a rainbow matching of size n . Kotlar and Ziv [7] improved this to $\lceil 5n/3 \rceil$. The author proved that $\phi n + o(n)$ is sufficient, where $\phi \approx 1.618$ is the Golden Ratio [8]. Clemens and Ehrenmüller [9] showed that $3n/2 + o(n)$ is sufficient. The best currently known bound is by Aharoni, Kotlar, and Ziv [10] who showed that having $3n/2 + 1$ edges of each colour in an n -edge-coloured bipartite multigraph guarantees a rainbow matching of size n .

Additionally, there are two results showing that just $(1 + o(1))n$ edges in each colour are enough if we place additional assumptions on G . A special case of a theorem of Haggkvist and Johansson [11] (proved by probabilistic methods) is that “every bipartite graph consisting of n edge-disjoint perfect matchings of size $n + o(n)$ edges has a rainbow matchings of size n ”. The author showed that the assumption that the matchings are perfect can be removed i.e. every bipartite graph consisting of n edge-disjoint matchings of size $n + o(n)$ edges has a rainbow matching of size n [8].

The goal of this paper is to improve on all previous asymptotic results by showing that $(1 + o(1))n$ edges are sufficient for all bipartite multigraphs.

Theorem 1.3. *For all $\epsilon > 0$, there exists an $N_0 = N_0(\epsilon)$ such that the following holds. Let G be a properly coloured bipartite multigraph with $n \geq N_0$ colours and at least $(1 + \epsilon)n$ edges of each colour. Then G contains a rainbow matching using every colour.*

The above theorem is the natural approximate version of Conjecture 1.2. Now the interesting direction for further research is to try and improve the second order term.

This theorem is proved by associating an auxiliary directed graph with G and studying certain kinds of paths in the directed graph. Such an approach was also taken in the author’s previous paper [8], and is substantially refined

here. In the next section we give an overview of the various components of the proof of Theorem 1.3.

2. Proof sketch

The proof of Theorem 1.3 is quite long and complicated. The basic idea is to associate an auxiliary directed graph to G and then study properties of this directed graph. The directed graph is studied by introducing five new concepts—“switching paths”, “amidstness”, “reaching”, “bypassing”, and “ λ -components”—and then proving many lemmas about these concepts. Since these concepts are quite foreign, we use this section to give a slow and detailed introduction to all of them. In particular we motivate some of these concepts by showing how they relate to the initial undirected graph in Theorem 1.3.

This section and the main proof of Theorem 1.3 (Sections 3 and 4) can be read completely independently of one another. All concepts that we introduce in this section, will be reintroduced during the main proof of Theorem 1.3 (usually more concisely.)

2.1. Associating a directed graph

Let G be a properly coloured bipartite multigraph as in Theorem 1.3, and let M be a rainbow matching of maximum size in G . Suppose for the sake of contradiction that M doesn't use every colour. Aside from [11], all approaches to Conjecture 1.2 have involved performing local manipulations on M to try and produce a larger rainbow matching. Here a “local manipulation” on M means choosing some edge $m \in M$ and $e \notin M$ such that $M - m + e$ is another rainbow matching of the same size as M . The basic idea of the proof is to perform a sequence of such local manipulations to obtain a new matching M' of the same size as M such that there is some edge which can be added to M' to give a larger rainbow matching. Since M was originally chosen to have maximum size, this gives a contradiction.

Thus the main aim throughout the proof is to find a suitable sequence of local manipulations. A key idea in [8] was that such sequences correspond to paths in a suitable auxiliary directed graph. The following is the directed graph which we will use.

Definition 2.1 (The directed graph $D_{G,M}$). *Let G have bipartition classes X and Y , C_G the set of colours in G , and C_M the set of colours on M . Let*

$X_0 = X \setminus V(M)$. For any colour $c \in C_M$, let m_c be the colour c edge of M . The digraph $D_{G,M}$ corresponding to G and M is defined as follows:

- The vertex set of $D_{G,M}$ is the set C_G of colours of edges in G .
- For two colours u and $v \in V(D_{G,M})$ there is a directed edge from u to v in $D_{G,M}$ whenever there is a colour u edge from some $x \in X$ to the vertex $m_v \cap Y$.

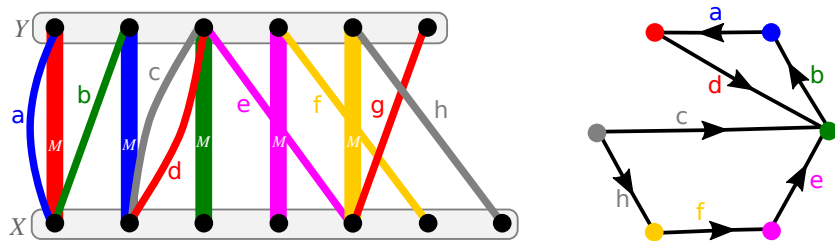


Figure 1: A graph G , with a matching M , and the corresponding directed graph $D_{G,M}$. The thick vertical edges labelled “ M ” are the rainbow matching M . All other edges are denoted by $a - h$ to show which edge of $D_{G,M}$ corresponds to which edge of G . Notice that the edge g of G doesn’t have a corresponding edge in $D_{G,M}$ —this is because g doesn’t go through $Y \cap V(M)$.

See Figure 1 for a diagram of a bipartite multigraph and the corresponding directed graph $D_{G,M}$. Consider the directed path in the $D_{G,M}$ with edge sequence (h, f, e) and vertex sequence (grey, yellow, pink, green). Notice that deleting the yellow, pink, and green edges from M and replacing them with h, f , and e produces a new rainbow matching of the same size as M . In addition this new matching misses a different colour (green rather than grey.) This demonstrates that directed paths in $D_{G,M}$ can give the kinds of local manipulations we are interested in.

However not all directed paths in $D_{G,M}$ correspond to sequences of local manipulations. For example in Figure 1, the directed path c, b, a doesn’t work since the three edges c, b, a in G do not form a matching. In fact it is easy to check that the only directed paths in Figure 1 which correspond to the kinds of manipulations we’re interested in are sub-paths of (h, f, e) .

The previous paragraphs show that while paths in $D_{G,M}$ can capture the kind of local manipulations we’re looking for, not all paths do so. We will add labels to the edges of $D_{G,M}$ in order to be able to describe exactly the kind of paths we’re interested in. The set of labels for edges of $D_{G,M}$ is $X_0 \cup C_M$

(where $X_0 = X \setminus V(M)$ and C_M is the set of colours of the edges of M .) If there is a colour u edge in G from $x \in X$ to $m_v \cap Y$, then we label the corresponding edge $uv \in D_{G,M}$ by the following rule.

- If $x \in X_0$ then the edge uv is labelled by x .
- If $x \in m_c \in M$ then uv is labelled by c , the colour of m_c .

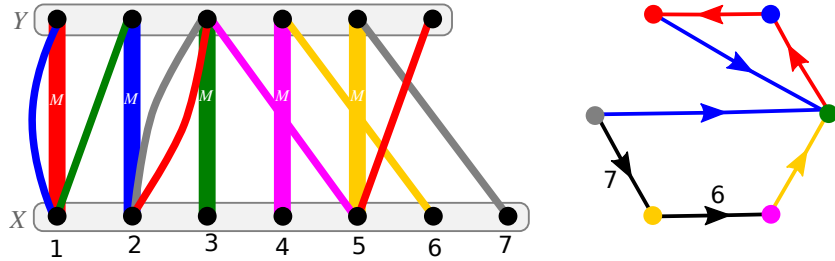


Figure 2: The same graphs G and $D_{G,M}$ as in Figure 1, but now with the edge-labels on $D_{G,M}$. The coloured edges in $D_{G,M}$ are ones labelled by elements of C_M (or equivalently the ones labelled by something in $V(D_{G,M})$.) The black edges are ones labelled by elements of X_0 (or equivalently ones labelled by something not in $V(D_{G,M})$.)

See Figure 2 for an example of this labelling. One key point to notice is that the set of labels $X_0 \cup C_M$ is not just an ambient set—since $V(D_{G,M}) = C_G$ an element of C_M can simultaneously be a vertex of $D_{G,M}$ and a label of edges in $D_{G,M}$. Formally, an edge-labelled directed graph is defined to be a directed graph D together with a set X_0 with $X_0 \cap V(D) = \emptyset$ and a labelling function $f : E(D) \rightarrow V(D) \cup X_0$. The set X_0 is called the set of *non-vertex labels* in D . We call $X_0 \cup V(D)$ the *set of labels* in D (regardless of whether D actually has edges labelled by all elements of $X_0 \cup V(D)$).

Having equipped $D_{G,M}$ with a labelling, we can define the kinds of paths we are interested in.

Definition 2.2 (Switching path). *A path $P = (p_0, \dots, p_d)$ in an edge-labelled, directed graph D is a switching path if the following hold.*

- P is rainbow i.e. the edges of P have different labels.
- If $p_i p_{i+1}$ is labelled by a vertex $v \in V(D)$, then $v = p_j$ for some $1 \leq j \leq i$.

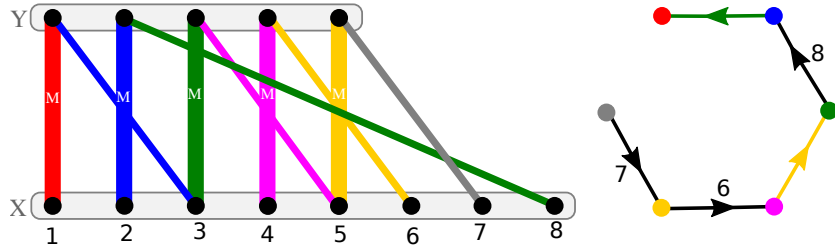


Figure 3: A switching path in a graph $D_{G,M}$. Replacing the edges of M for the other pictured edges of G produces a new matching of the same size as M .

In other words a switching path is a rainbow path with a kind of “consistency” property for its edge-labels which are vertices: For every edge $e \in P$ which is labelled by a vertex v , P must pass through v before it reaches e . Notice that this vertex v is not allowed to be p_0 , the starting vertex of P . A consequence of this is that the first edge p_0p_1 of P cannot be labelled by a vertex of D (in the case of $D_{G,M}$ this means that the first edge of any switching path must be labelled by something in X_0).

See Figure 3 for an example of a switching path. Notice that this path does correspond to the kinds of local manipulations of M which we are interested in i.e if we exchange the edges of M for the edges in G corresponding to the switching path, then we obtain a new rainbow matching of the same size as M .

When looking at a switching path in the graph is $D_{G,M}$, the vertices of P correspond to edges of G which we want to remove from the matching M , and the edges of P correspond to edges of G which we want to add to M . The two conditions in the definition of “switching path” then have natural interpretations when one seeks to obtain a new rainbow matching by switching the edges along P . Asking for the switching path to be rainbow is equivalent asking for the edges we want to add to M not intersecting in X (which is needed to get a matching). The second part of Definition 2.2 ensures that when we add an edge to M , its colour was previously removed from M .

The following exercise makes precise how to modify a matching M using a switching path in $D_{G,M}$ starting from a colour outside M .

Exercise 2.3. Let M be a rainbow matching in a graph G , p_0 a colour not in M , and $P = (p_0, p_1, \dots, p_d)$ a switching path in $D_{G,M}$. For $i \geq 1$, let m_i be the colour p_i edge of M , and for $i \geq 0$, let e_i be the edge of G corresponding

to $p_i p_{i+1}$. Show that the following is a rainbow matching missing the colour p_d :

$$M + e_0 - m_1 + e_1 \cdots - m_{d-1} + e_{d-1} - m_d.$$

For a solution to the above exercise, see Claim 3.6. Exercise 2.3 is exactly what we use to try and extend M into a larger matching. If M was chosen to be maximum, then Exercise 2.3 can be used to show that $D_{G,M}$ possesses a certain degree property. This and other properties of $D_{G,M}$ will be discussed in the next section.

2.2. Properties of the directed graph

The labelled directed graph $D_{G,M}$ ends up having several properties which we use in the proof of Theorem 1.3. In this section we go through the properties which we need. See Figure 4 for examples of some of the features that $D_{G,M}$ can have.

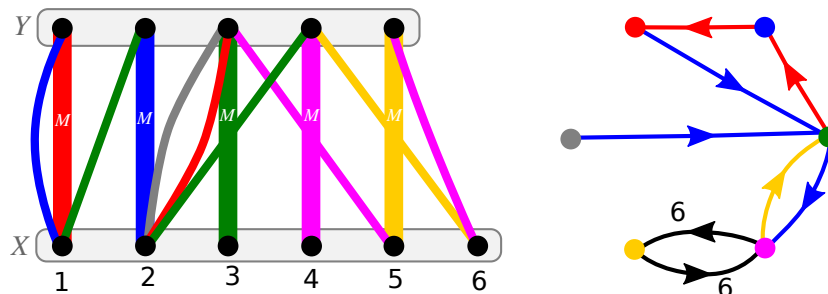


Figure 4: Some of the features $D_{G,M}$ has. The directed graph $D_{G,M}$ doesn't have multiple edges, unless they go in different directions (like the two edges labelled 6). At a vertex v , $D_{G,M}$ never has out-going edges with the same label, but it may have in-going edges with the same label (For example the green vertex has two in-going blue edges.)

For two vertices $u, v \in DGM$ it is possible for uv and vu to both be present in $D_{G,M}$. For example the two edges between the green and pink vertices in Figure 4. However it is impossible for the edge uv to appear twice with different labels i.e. the directed graph $D_{G,M}$ is simple.

Exercise 2.4. Using the fact that G is properly coloured, show that for $u, v \in V(D_{G,M})$, there is at most one edge from u to v in $D_{G,M}$.

For a solution to this exercise, see Lemma 3.2. The labelling on the directed graph $D_{G,M}$ is far from a general labelling. We make the following definitions which generalize proper colouring to directed graphs.

Definition 2.5. Let D be a labelled directed graph.

- D is out-properly labelled if for any $u \in V(D)$, all out-going edges uv have different labels.
- D is in-properly labelled if for any $u \in V(D)$, all in-going edges vu have different labels.

It turns out that the labelling on $D_{G,M}$ is always out-proper.

Exercise 2.6. Using the fact that G is properly coloured, show that $D_{G,M}$ is out-properly labelled.

For a solution to this exercise, see Lemma 3.2. The labelling on $D_{G,M}$ is not always in-proper. For example, in Figure 4, the green vertex has two in-going blue edges. Notice that in Figure 4 this happened because of the multiple edge in G . It turns out that this is the only way to have in-going edges with the same label in $D_{G,M}$.

Exercise 2.7. Suppose that G is properly coloured, simple, and M is a matching in G . Show that $D_{G,M}$ is in-properly labelled.

Recall that the special case of Theorem 1.3 when G is simple was proved in the author's earlier paper [8]. The case when G is simple turns out to be much easier to prove precisely because the directed graph $D_{G,M}$ associated to G is both in-properly and out-properly labelled. The reason for the difficulty of the multigraph case is that dense directed graphs which are not in-properly labelled do not necessarily have certain connectivity properties. This difficulty is explained in more detail in Section 2.3.

The other main property of $D_{G,M}$ which we will need is a degree property i.e. we will want to know that all vertices in $D_{G,M}$ have a suitably large degree. Let $Y_0 = Y \setminus S$ be the set of vertices in Y disjoint from the matching M . From the definition of $D_{G,M}$, notice that every edge $e \in G$ corresponds to an edge of $D_{G,M}$ unless e passes through Y_0 or $e \in M$ ². A consequence of this is that $e(D_{G,M}) = e(G) - |M| - e(X, Y_0)$. Recall that every colour c in

²The edges of M could be naturally thought of corresponding to loops in $D_{G,M}$, but to keep our analysis to loopless graphs, we won't do this.

G has $(1 + \epsilon)n$ edges. For a colour c , let c_{Y_0} be the number of colour c edges going through Y_0 . From the definition of $D_{G,M}$ we have

$$|N^+(c)| = (1 + \epsilon)n - c_{Y_0} - 1 \text{ if } M \text{ has a colour } c \text{ edge,} \quad (1)$$

$$|N^+(c)| = (1 + \epsilon)n - c_{Y_0} \text{ if } M \text{ has no colour } c \text{ edge.} \quad (2)$$

Here $N^+(c)$ denotes the out-neighbourhood of c i.e. the set of $x \in V(D_{G,M})$ with $cx \in E(D_{G,M})$. Notice that (1) and (2) do not by themselves imply that $|N^+(c)|$ is large for any colour c . It is possible that most of the edges of G go through Y_0 , making the c_{Y_0} term dominant in (1) and (2). However the fact that M is a *maximum size* rainbow matching does force some colours in G to have a large out-degree in $D_{G,M}$. In particular if c_0 is a colour which does not appear on M , then notice that there cannot be any edges in G between X_0 and Y_0 —indeed if such an edge existed then it could be added to M to give a rainbow matching larger than M . Recall that from the assumption of Theorem 1.3 there are $\geq (1 + \epsilon)n$ colour c_0 edges in G , and at most $|M| \leq n$ of these can intersect $X \cap V(M)$. The other ϵn colour c_0 edges must go between X_0 and $Y \cap V(M)$, giving $|N^+(c_0)| \geq \epsilon n$.

The above discussion shows that all colours not on M have a high out-degree in $D_{G,M}$. Can we get something similar for the other colours in G ? Recall from Exercise 2.3 that switching paths can be used to give new rainbow matchings with the same size as M . Using this it is easy to show that any colour close to $c_0 \notin M$ in $D_{G,M}$ has a large degree in $D_{G,M}$ as well.

Exercise 2.8. *Let c_0 be a colour not on M , and c some other colour. Let P be a switching path from c_0 to c in $D_{G,M}$. Then $|N^+(c)| \geq \epsilon n - |P|$.*

The above exercise is a special case of Lemma 3.7 which we prove later. So far we have looked at only edges labelled by X_0 and found that vertices close to missing colours have many such edges leaving them. For a set of labels L , define $N_L^+(v)$ to be the set of $x \in N(v)$ with vx labelled by some $\ell \in L$. Under the assumptions of Exercise 2.8, it is easy to show that $|N_{X_0}^+(c)| \geq \epsilon n - |P|$.

We would like to have information about how big N_L^+ is for sets of labels L other than X_0 . Where could we get such information? In Figure 5, notice that if M is a maximum matching, then there cannot be any red edges going from $\{2, 4, 9\}$ to Y_0 . Indeed if there was such an edge e then we could look at the rainbow matching M' as in Exercise 2.3 (corresponding to the switching path in Figure 5) and then add e to M' to get a larger rainbow matching. Thus if there are red edges in G touching $\{2, 4, 9\}$, then they must go through

$Y \cap V(M)$, and hence must have corresponding edges in $D_{G,M}$. Since these edges go from v to $N^+(v)$, this would tell us that $N^+(v)$ is slightly bigger than the estimate we have in Exercise 2.8. For just the single path P in Figure 5, this increase is very small. But if we had a large collection of switching paths P like the one in Figure 5, then the gains may add up to give a large improvement on the bound in Exercise 2.8. The next definition captures what kind of information about the path P in Figure 5 we were interested in.

Definition 2.9 (Amidst). *Let u and v be two vertices in an edge-labelled, directed graph D , and ℓ a label. We say that ℓ is amidst u and v if there is a switching path $P = (u, p_1, \dots, p_d, v)$ from u to v such that the following hold.*

- There are no edges of P labelled by ℓ .
- If ℓ is a vertex of D then $\ell \in \{p_1, \dots, p_d, v\}$.

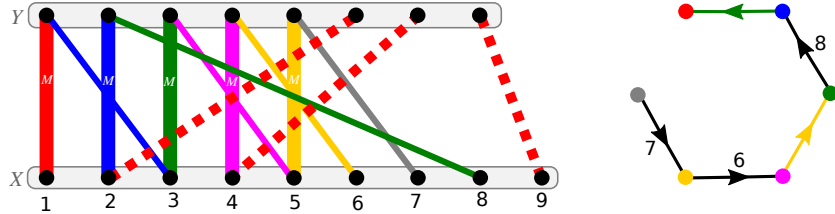


Figure 5: A switching path P in a graph $D_{G,M}$, and three edges that can be added to get a larger matching. Notice that the three labels {blue, pink, 9} in $D_{G,M}$ are amidst grey and red, as witnessed by the switching path P . The vertices in X of blue and pink are 2 and 4—which are the X -vertices of the corresponding dashed red edges. This shows how amidstness is used to identify vertices of X through which we can add edges to extend M . This is the essence of Exercise 2.10.

Notice that in the above diagram, labels which are not *amidst* a pair of labels cannot be used for augmenting the matching M . For example, if there was a red edge f from vertex 5 to Y_0 , then one might hope to switch some edges to free up the red colour and vertex 5 in order to extend the matching by adding f . However this cannot be done because freeing up vertex 5 and colour red would require yellow to be amidst grey and red. In the above diagram yellow is not amidst grey and red. (intuitively because in order to free up red, the yellow edge starting at 6 must be used).

Notice a parallel between each of the two parts of the definitions of “switching path” and “amidst”: The first parts are about forbidding edges of a path from having particular labels, whereas the second parts are about

paths passing through a particular vertex. This similarity is no coincidence — a path $P = (p_0, \dots, p_d)$ is a switching path if, and only if, the path p_0, \dots, p_{d-1} is a switching path witnessing the label of $p_{d-1}p_d$ being amidst p_0 and p_{d-1} .

If P is a path as in Definition 2.9, then we say that P *witnesses* ℓ being amidst u and v . As an example, the path P in Figure 5 witnesses each of the labels {blue, pink, 9} being amidst grey and red. Suppose that $\ell \in C_M$ is the colour of some edge m in M . By an argument similar to the one in the previous paragraph, it is possible to show that if ℓ is a label amidst u and v , and u is not present on M , then there is no colour v edge from $m \cap X$ to Y_0 .

Exercise 2.10. *Let ℓ, u, v be colours in G with u not in M and ℓ the colour of an edge $m \in M$. If ℓ is amidst u and v , then there is no colour v edge from $m \cap X$ to Y_0 in G .*

For a solution to the above exercise see Lemma 3.5. The essence of the solution is in Figure 5 — the dashed red edges are exactly the kind of edges that Exercise 2.10 is about. If any of them were present in the graph then they could be augmented to the matching. We now have that given a set of vertices $X' \subseteq X$, if all the corresponding labels are amidst u and v , then all the colour v edges touching X' in G must contribute to $N^+(v)$ in $D_{G,M}$. The following exercise is a strengthening of Exercise 2.8 which takes into account vertices in X outside X_0 .

Exercise 2.11. *Suppose that M misses a colour c^* , v is a colour in G , and A is a set of labels in $D_{G,M}$ which are amidst c^* and v . Then $|N_A^+(v)| \geq |A| - |X_0| + \epsilon n - 1$.*

For a solution to this exercise, see Lemma 3.7. As remarked before, this is actually a strengthening of Exercise 2.8. Indeed given a path P as in Exercise 2.8, notice that if $x \in X_0$ is a label which does not occur on edges of P , then x is amidst u and v (witnessed by the path P .) Applying Exercise 2.11 with A the set of labels in X_0 and not on P we get $|N_A^+(v)| \geq |A| - |X_0| + \epsilon n - 1 \geq \epsilon n - |P|$.

Exercise 2.11 allows us to finally state the method we use to prove Theorem 1.3. We prove that for any $\epsilon > 0$, there cannot be arbitrarily large labelled digraphs satisfying the degree condition of Exercise 2.11. The following is an intermediate theorem we prove, which implies Theorem 1.3.

Theorem 2.12. *For all ϵ with $0 < \epsilon \leq 0.9$, there is an $N_0 = N_0(\epsilon)$ such that the following holds. Let D be any out-properly edge-labelled, simple, directed graph on $n \geq N_0$ vertices. Let X_0 be the set of labels which are not vertices of D*

Then for all $u \in V(D)$, there is a vertex v and a set of labels A amidst u and v , such that $|N_A^+(v)| < |A| - |X_0| + \epsilon n$.

We remark that the set A can be an arbitrary subset of $V(D) \cup X_0$ and that D might not have edges labelled by all elements of A .

Modulo the discussion in this section, it is easy to see that this theorem implies Theorem 1.3. Indeed suppose that there was a sufficiently large graph G as in Theorem 1.3. Suppose that a maximum matching M in G doesn't use every colour. By Exercises 2.4 and 2.6 we know that the corresponding digraph $D_{G,M}$ is out-properly labelled and simple. Let c^* be some colour outside M . By Exercise 2.11 we know that for any $v \in V(D_{G,M})$, we have $|N_{A^*}^+(v)| \geq |A| - |X_0| + \epsilon n - 1 \geq |A| - |X_0| + 0.9\epsilon n$ for any set of labels A amidst c^* and v . But this contradicts Theorem 2.12.

We conclude this section by explaining how amidstness can be used to build switching paths. Recall that a path $P = (p_0, \dots, p_d)$ is a switching path if, and only if, the path p_0, \dots, p_{d-1} is a switching path witnessing the label of $p_{d-1}p_d$ being amidst p_0 and p_{d-1} . Because of this, labels which are amidst two vertices u and v have potential to be used to extend switching paths. The following exercise makes this precise.

Exercise 2.13. *If a label ℓ is amidst x and y and there is some vertex z such that the edge yz is present and labelled by ℓ , then there is a switching path from x to z .*

A version of this exercise is proved in Lemma 4.25. Exercise 2.13 is important because it is one of the tools we will use to build longer and longer switching paths.

2.3. The right notion of connectedness

Theorem 2.12 is proved studying connectivity properties of subgraphs of D . It is not immediately apparent why connectivity is useful here. One hint of it being useful comes from the definition of “amidst”. The first part of the definition of “amidst” asks for a u to v path avoiding all edges of label ℓ . If there are $< k$ colour ℓ edges then this is a property k -edge-connected graphs have. The second part of the definition of “amidst” asks for a path going from

u to v via some other vertex ℓ . This is a property which 2-vertex-connected undirected graphs have (as a consequence of Menger’s Theorem).

The purpose of connectivity in the proof is to find sets of vertices $C \subseteq V(D_{G,M})$ which are highly connected in the following sense—for any pair $u, v \in C$ we have c amidst u and v for most $c \in C$. We can then plug C into the assumption of Theorem 2.12 in order to deduce that v has a high out-degree. Knowing that vertices in C have high out-degree is then used to find a set C' which is also highly connected and substantially larger than C . Iterating this process we get larger and larger highly connected sets, which can eventually be used to get a contradiction to these sets being smaller than $V(D)$.

What notion of connectivity should we use? In [8], the following notion was used.

Definition 2.14. *Let W be a set of vertices in a labelled digraph D . We say that W is (k, d) -rainbow connected in D if, for any set of at most k labels S and any vertices $x, y \in W$, there is a rainbow x to y path of length $\leq d$ in D avoiding colours in S .*

This kind of connectivity is useful when the graph G is a simple graph rather than a multigraph. Recall that if G is a simple graph then the labelling on $D_{G,M}$ is both in-proper and out-proper. In [8] it is proved that in any labelled digraph D there is a highly (k, d) -connected set C with $|C| \geq \delta^+(G) - o(n)$ which is a key intermediate result in proving Theorem 1.3 in the case when G is simple.

When G is a multigraph, then we know that $D_{G,M}$ is out-properly labelled, but not necessarily in-properly labelled. Definition 2.14 isn’t the right notion of connectivity for studying such graphs. It is possible to have an out-proper labelling of the complete directed graph in which any vertex can be isolated by deleting just one label. See Figure 6 for an example of such a graph. This graph is a complete directed graph where every edge xy is labelled by ℓ_y (for some label ℓ_y which only appears on in-going edges to y .) This graph has a high out-degree but doesn’t have any $(1, \infty)$ -connected subgraphs. This is the issue with using (k, d) -connectedness since we would like high out-degree graphs to have highly connected subsets.

We introduce a different kind of connectedness, for which the graph in Figure 6 is highly connected. The following is at the heart of the notion of connectivity which we use.

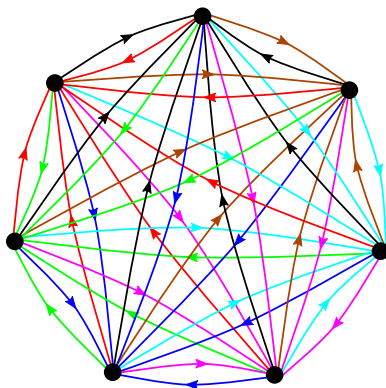


Figure 6: A labelled directed graph whose labelling is out-proper, but not in-proper. Here all the edge-labels are not vertices. The edge-labelling is such that every vertex v has a “chosen colour” with all edges directed towards v having the chosen colour. Notice that deleting all edges having a particular label reduces the in-degree of some vertex to 0, effectively isolating it.

Definition 2.15 (Reaching). *For a vertex $v \in V(D)$ and a set $R \subseteq V(D)$, we say that v (k, d, Δ) -reaches R if for any set S of $\leq k$ labels, there are length $\leq d$ switching paths avoiding S to all, except possibly at most Δ , vertices $x \in R$.*

Standard notions of connectedness are based on studying when two vertices are connected by a path. “Reaching” is fundamentally different from these since it is of no use to know that a vertex u reaches another vertex v . In fact any vertex u $(\infty, \infty, 1)$ -reaches any singleton $\{v\}$ (since $\Delta = 1$, we can let $\{v\}$ be the set of Δ vertices in S to which we don’t need to find a path in the definition of reaching. More generally, there is nothing to check in the definition of “reaching” when $\Delta \geq |R|$.) Thus “reaching” is only meaningful when we talk about a vertex reaching a reasonably large set of vertices R . Notice that the graph in Figure 6 has good connectivity properties with our new definition.

Exercise 2.16. *For the labelled directed graph D in Figure 6 and any $k \leq \Delta$, show that every $v \in V(D)$ $(k, 1, \Delta)$ -reaches $V(D)$.*

To prove Theorem 2.12, we will need to have a fairly deep understanding of “reaching”. This involves first proving several basic consequences of the definition such as showing that reaching is monotone under change of pa-

rameters, preserved by unions, and has a kind of transitivity property. These properties are proved in Section 4.1.

Our main goal when studying “reaching” will be to show that some analogue of connected components exists for the new notion of connectedness. Recall that a strongly connected component C in a directed graph is a maximal set of vertices in a graph such that for any two vertices x and y in C there is a path from x to y . Analogously, in a labelled graph we would like to find a maximal set C such that any $x \in C$ reaches all of C for suitable parameters. This notion of a maximal reached set seems a bit hard to work with, so we will instead deal with the following approximate version.

Definition 2.17 ($(k, d, \Delta, \gamma, \hat{k}, \hat{d}, \hat{\Delta})$ -component). *A set $C \subseteq V(D)$ is a (k, d, Δ, γ) -component if for any vertex $v \in C$, there is a set R_v with $|R_v \Delta C| \leq \epsilon^3 n$ such that the following hold.*

(i) v (k, d, Δ) -reaches R_v .

(ii) v doesn't $(\hat{k}, \hat{d}, \hat{\Delta})$ -reach any set R disjoint from R_v with $|R| \geq \gamma n$.

In other words a $(k, d, \Delta, \gamma, \hat{k}, \hat{d}, \hat{\Delta})$ -component is a set C such that every vertex $v \in C$ reaches most of C and doesn't reach any large set outside C . It is not at all obvious that $(k, d, \Delta, \gamma, \hat{k}, \hat{d}, \hat{\Delta})$ -components exist for particular parameters $k, d, \Delta, \gamma, \hat{k}, \hat{d}, \hat{\Delta}$. An important intermediate lemma we prove in Section 4.2, is that for given k, d, Δ, γ , there is a $(k', d', \Delta', \gamma', \hat{k}', \hat{d}', \hat{\Delta}')$ -component for new parameters $k', d', \Delta', \gamma', \hat{k}', \hat{d}', \hat{\Delta}'$ close to k, d, Δ, γ .

We make a remark about how constants will be dealt with throughout this paper. Looking at the definitions of “ (k, d, Δ) -reaches” and “ $(k, d, \Delta, \gamma, \hat{k}, \hat{d}, \hat{\Delta})$ -component”, they look a bit scary because of the large number of parameters there are in each definition. In the actual proof of Theorem 2.12 in Section 4 this won't be the case because we introduce a single parameter, λ , which will control each of the parameters $k, d, \Delta, \gamma, \hat{k}, \hat{d}, \hat{\Delta}$. Formally, in Section 4 we define seven explicit functions $k_\epsilon(\lambda)$, $d_\epsilon(\lambda)$, $\Delta_\epsilon(\lambda)$, $\gamma_\epsilon(\lambda)$, $\hat{k}_\epsilon(\lambda)$, $\hat{d}_\epsilon(\lambda)$, and $\hat{\Delta}_\epsilon(\lambda)$ depending on ϵ (which is the constant given in the statement of Theorem 2.12.) Then we say that v λ -reaches a set R if v $(k_\epsilon(\lambda), d_\epsilon(\lambda), \Delta_\epsilon(\lambda))$ -reaches R , and that a set C is a λ -component if C is a $(k_\epsilon(\lambda), d_\epsilon(\lambda), \Delta_\epsilon(\lambda), \gamma_\epsilon(\lambda), \hat{k}_\epsilon(\lambda), \hat{d}_\epsilon(\lambda), \hat{\Delta}_\epsilon(\lambda))$ -component. The advantage of this is that it means that only one parameter, λ , needs to be kept track of between the various lemmas that we prove. This makes the high level structure of the proof of Theorem 2.12 easier to follow.

We mention a final definition which we use in the paper.

Definition 2.18 (Bypassing). *For a vertex $v \in V(D)$ and a set $B \subseteq V(D)$, we say that v $(\hat{k}, \hat{d}, \hat{\Delta}, \gamma)$ -bypasses B if v doesn't $(\hat{k}, \hat{d}, \hat{\Delta})$ -reach any set R contained in B with $|R| \geq \gamma n$*

The significance of the above definition is that part (ii) of the definition of $(k, d, \Delta, \gamma, \hat{k}, \hat{d}, \hat{\Delta})$ -component can be now rephrased as “ v $(\hat{k}, \hat{d}, \hat{\Delta}, \gamma)$ -bypasses $V(D) \setminus R_v$ ”. Thus the notion of bypassing is important because it eases the study of $(k, d, \Delta, \gamma, \hat{k}, \hat{d}, \hat{\Delta})$ -components.

Recall that at the start of the section we said that the reason for using connectedness is to be able to study “amidstness”. It is not immediately apparent how the definitions we introduce do this. With a bit of work it is possible to prove that in a $(k, d, \Delta, \gamma, \hat{k}, \hat{d}, \hat{\Delta})$ -component C , most triples $(u, c, v) \in C \times C \times C$ have c amidst u and v .

Lemma 2.19. *For $\epsilon > 0$, D a sufficiently large labelled directed graph, and C a $(k, d, \Delta, \gamma, \hat{k}, \hat{d}, \hat{\Delta})$ -component in D for suitable $k, d, \Delta, \gamma, \hat{k}, \hat{d}, \hat{\Delta}$, there are at least $|C|^3 - (\epsilon n)^3$ triples $(u, c, v) \in C \times C \times C$ with c amidst u and v .*

The above lemma is an easy consequence of Lemma 4.26 which we prove in Section 4.3. The full Lemma 4.26 will say a bit more, giving information about the structure of triples $(u, c, v) \in C \times C \times C$ with c amidst u and v .

2.4. An overview of the proof of Theorem 2.12

Here we give a high level overview of the strategy of the proof of Theorem 2.12. The proof begins by supposing for the sake of contradiction that there is a vertex $u \in V(D)$ such that for every vertex v and a set of labels A amidst u and v we have $|N_A^+(v)| < |A| - |X_0| + \epsilon n$. The proof of the theorem naturally splits into three parts.

1. Find $(k_i, d_i, \Delta_i, \gamma_i, \hat{k}_i, \hat{d}_i, \hat{\Delta}_i)$ -components C_0, \dots, C_m for suitable parameters such that $C_i \cap C_{i+1} \neq \emptyset$. In addition we find a short switching path from u to each C_i . This is done as follows:
 - 1.1 Prove lemmas along the lines of “for any vertex v and parameters k, d, Δ, γ there are complementary sets R_v and B_v such that v (k', d', Δ') -reaches R_v and doesn't reach anything in B_v for suitable parameters k', d', Δ' close to k, d, Δ . This is performed in Lemmas 4.13 and 4.14.

- 1.2 Show that reaching has a transitivity property: If v reaches a sufficiently large set R and every vertex in R reaches a set R' , then v reaches R' . This is performed in Lemma 4.11.
 - 1.3 Choose a vertex v with the set R_v from part 1.1 as small as possible. Using transitivity, it is possible to show that for most vertices $u \in R_v$ we have that $|R_u \Delta R_v|$ is small. By letting $C = R_v$ minus a few vertices it is possible to get a single component of the sort we want. This is performed in Lemma 4.18.
 - 1.4 By iterating 1.3, we can get the sequence of components C_0, \dots, C_m which we need. This is performed in Lemma 4.22.
2. We show that if C is a $(k, d, \Delta, \gamma, \hat{k}, \hat{d}, \hat{\Delta})$ -component close to u , then **either** any $v \in C$ $(k', d', \Delta', \gamma')$ -reaches some set R with $|R| \geq |C| + (\epsilon - o(1))n$ for suitable parameters **or** the conclusion of Theorem 2.12 holds for some $A \subseteq C \cup X_0$. The formal statement of this is Lemma 4.32. This step is performed as follows:
- 2.1 We show that for most triples $(u, c, v) \subseteq C \times C \times C$, c is amidst u and v . This is performed in Lemma 4.26.
 - 2.2 Let R be the set of $z \in V(D)$ for which there are a lot of triples (u, c, v) such that vz is an edge labelled by c and c is amidst u and v .
 - 2.3 The vertex v ends up (k', d', Δ') -reaching R as a consequence of 2.1. This is performed in Claim 4.34.
 - 2.4 Use the assumption of Theorem 2.12 applied to a suitable subset of $C \cup X_0$ together with Lemma 2.19 to show that $|R| \geq |C| + (\epsilon - o(1))n$. This is performed in Claims 4.35 and 4.36.
3. Combining parts 1 and 2 and the definition of $(k, d, \Delta, \gamma, \hat{k}, \hat{d}, \hat{\Delta})$ -component we obtain that $|C_{i+1}| \geq |C_i| + (\epsilon - o(1))n$ for every C_i from part 1. If the number of components $m \gg \epsilon^{-1}$ this gives a contradiction to $|C_m| \leq |V(D)| = n$.

2.5. An example

In this section we give an illustrative labelled directed graph and explain how the proof of Theorem 2.12 works for that particular graph.

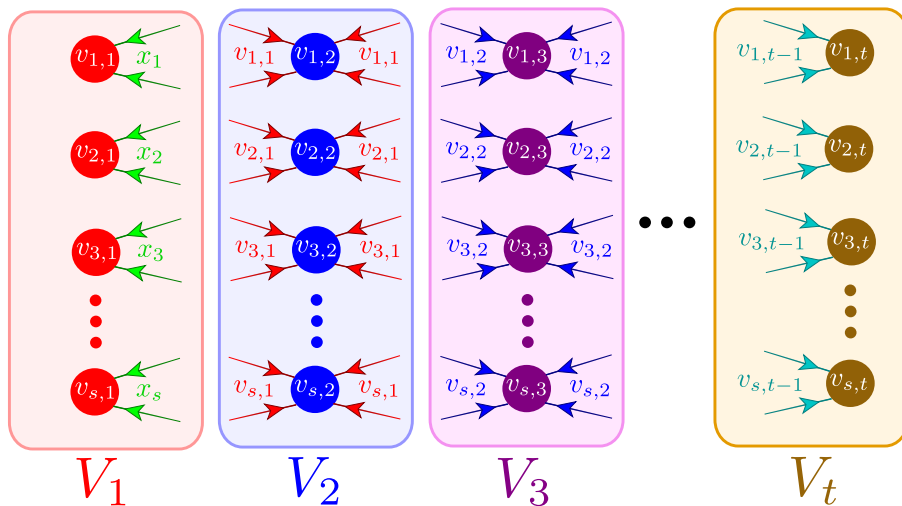


Figure 7: The labelled, directed graph $D_{s,t}$.

For a fixed $\epsilon > 0$ and $t \ll \epsilon^{-1}$, we define a directed graph $D_{s,t}$ as follows: $D_{s,t}$ has $n = st$ vertices split into t disjoint classes V_1, \dots, V_t each of size s with $V_i = \{v_{1,i}, \dots, v_{s,i}\}$. The set of non-vertex labels in $D_{s,t}$ is $X_0 = \{x_1, \dots, x_s\}$. All the edges are present in G going in both directions. Each vertex in G has its own “chosen label” with all edges directed towards the vertex having that label (much like the graph in Figure 6.) For $i \neq 1$ and any vertex u , the edge $uv_{j,i}$ has label $v_{j,i-1}$. For a vertex u and a vertex $v_{j,1} \in V_1$, the edge $uv_{j,1}$ has label x_j .

Because of the simple structure the graphs $D_{s,t}$ have, it is possible to describe all switching paths, reached sets, bypassed sets, and components in these graphs.

Exercise 2.20 (Switching paths in $D_{s,t}$). *A directed path $P = (p_0, p_1, \dots, p_d)$ is a switching path if, and only if, “ $v_{j,i} = p_k \in V(P) \implies v_{j,i-1}, \dots, v_{j,1} \in \{p_1, \dots, p_{k-1}\}$ ”.*

Exercise 2.21 (Amidstness in $D_{s,t}$).

- A vertex $v_{a,b}$ is amidst vertices $v_{c,d}$ and $v_{e,f}$ if, and only if, $b, d \neq f$ and $v_{a,b} \neq v_{c,d}$.
- A label x_b is amidst vertices $v_{c,d}$ and $v_{e,f}$ if, and only if, $b, d \neq f$.

In each of the next three exercises we give a necessary condition and a sufficient condition for a set to be reached, bypassed, or be a component. Although the necessary conditions and a sufficient conditions which we give are not exactly the same, they are always quite similar. Thus the next three exercises should be seen as giving a near-characterization of sets which are reached, bypassed, and components in $D_{s,t}$.

Exercise 2.22 (Reaching in $D_{s,t}$). *Let $v \in V(D_{s,t})$ and $R \subseteq V(D_{s,t})$.*

- *If v (k, d, Δ) -reaches R then $|R \setminus (V_1 \cup \dots \cup V_d)| \leq \Delta$.*
- *If $|R \setminus (V_1 \cup \dots \cup V_d)| \leq \Delta - (k + 1)d$, then v (k, d, Δ) -reaches R .*

Exercise 2.23 (Bypassing in $D_{s,t}$). *Let $v \in V(D_{s,t})$, $B \subseteq V(D_{s,t})$, and $\hat{\Delta} \geq (\hat{k} + 1)\hat{d}$.*

- *If v $(\hat{k}, \hat{d}, \hat{\Delta}, \gamma)$ -bypasses B then $|B \cap (V_1 \cup \dots \cup V_d)| \leq \gamma n$.*
- *If $|B \cap (V_1 \cup \dots \cup V_d)| < \gamma n - \hat{\Delta}$, then v $(\hat{k}, \hat{d}, \hat{\Delta}, \gamma)$ -bypasses B .*

Exercise 2.24 (Components in $D_{s,t}$). *Let $C \subseteq V(G)$, $\epsilon^3 n \geq \max(\gamma n, \Delta, \hat{\Delta})$, $\Delta \geq (k + 1)d$, $\hat{\Delta} \geq (\hat{k} + 1)\hat{d}$, and $\hat{d} \leq d$*

- *If C is a $(k, d, \Delta, \gamma, \hat{k}, \hat{d}, \hat{\Delta})$ -component then $|C \Delta (V_1 \cup \dots \cup V_d)| \leq 8\epsilon^3 n$.*
- *If $|C \Delta (V_1 \cup \dots \cup V_d)| \leq \epsilon^3 n$ then C is a $(k, d, \Delta, \gamma, \hat{k}, \hat{d}, \hat{\Delta})$ -component.*

Using Exercise 2.21 we can check that Theorem 2.12 holds for the graphs $D_{s,t}$. To see this notice that for any pair of vertices $v_{a,b}$ and $v_{c,d}$ for $b \neq d$ and any set of labels A with $A \cap \{v_{i,b}, v_{i,d} : i = 1, \dots, s\} = \emptyset$ we have A amidst $v_{a,b}$ and $v_{c,d}$. Notice also that $|N_A^+(v)| = |A|$ or $|A| - 1$ for every vertex v and set of labels A . Finally, recall that $|X_0| = s \geq \epsilon n$. Thus we see that for a given vertex $u = v_{a,b}$ the conclusion of Theorem 2.12 holds by choosing $v = v_{c,d}$ for $d \neq b$ and A any set of labels disjoint from $\{v_{i,b}, v_{i,d} : i = 1, \dots, s\}$ with $|A| > n$ (for example we could take $A = V(D_{s,t}) \cup X_0 \setminus \{v_{i,b}, v_{i,d} : i = 1, \dots, s\}$.)

This example teaches us an important lesson. It is not hard to see that if A is amidst u and v and satisfies $|N_A^+(v)| < |A| - |X_0| + \epsilon n$ then $|A| > n$ must hold. From this we see that any proof of Theorem 2.12 for the graph $D_{s,t}$ must “find” a very large set of labels A amidst some pair of vertices. In the remainder of this section, we explain how the strategy in Section 2.4 finds such a set A .

In Part 1 of the overview in Section 2.4, the proof finds $(k_i, d_i, \Delta_i, \gamma_i, \hat{k}_i, \hat{d}_i, \hat{\Delta}_i)$ -components C_0, \dots, C_m for suitable parameters such that $C_i \cap C_{i+1} \neq \emptyset$. Here “suitable parameters” means that $m \gg \epsilon^{-1}$, that $d_i, \Delta_i, \gamma_i, \hat{d}_i$, and $\hat{\Delta}_i$ increase with i while k_i and \hat{k}_i decrease with i . For the sake of argument, let us consider what happens when $m = \epsilon^{-4}$, $d_i = \hat{d}_i = i$, $\gamma_i = \epsilon^9 i$, $\Delta_i = \hat{\Delta}_i = 2im$, and $k = \hat{k} = m - i$. For these values, Exercise 2.24 tells us that for $i = 1, \dots, t$ we must have $|C_i \Delta (V_1 \cup \dots \cup V_i)| \leq 8\epsilon^3 n$ and that for $i = t + 1, \dots, m$ we have $|C_i \Delta (V_1 \cup \dots \cup V_t)| \leq 8\epsilon^3 n$.

In Parts 2 and 3 of the overview in Section 2.4, it is shown that if C is a component close to u , then *either* any $v \in C$ $(k', d', \Delta', \gamma')$ -reaches some set R with $|R| \geq |C| + (\epsilon - o(1))n$ for suitable parameters *or* the conclusion of Theorem 2.12 holds for some $A \subseteq C \cup X_0$ and $v \in C$. Testing this for the components C_0, \dots, C_m from Part 1, we see that for $i = 0, \dots, t - 1$, there is a set R such that v $(k_i, d_i + 1, \Delta_i, \gamma_i)$ -reaches some set R with $|R| \geq |C| + (\epsilon - o(1))n$ (namely we can take $R = V_1 \cup \dots \cup V_{i+1}$.) On the other hand for $i = t, \dots, m$ such a set R doesn't exist, so Part 2 would imply that the conclusion of Theorem 2.12 holds for some $A \subseteq C_i \cup X_0$ and $v \in C$. If $u = v_{a,b}$ then we see that this is indeed the case with e.g., $A = C_i \cup X_0 \setminus \{v_{1,b}, \dots, v_{t,b}, v_{1,c}, \dots, v_{t,d}\}$ and $v = v_{a,c}$.

Notation

For standard notation we follow [12]. A path $P = (p_0, p_1, \dots, p_d)$ in a directed graph D is a sequence of vertices p_0, p_1, \dots, p_d such that $p_i p_{i+1}$ is an edge for $i = 0, \dots, d - 1$. The *order* of P is the number of vertices it has, and the *length* of P is the number of edges it has. We'll use additive notation for concatenating paths i.e. if $P = (p_1, p_2, \dots, p_i)$ and $Q = (p_i, p_{i+1}, \dots, p_d)$ are two internally vertex-disjoint paths, then we let $P + Q$ denote the path $(p_1 p_2 \dots p_d)$. Throughout the paper, all directed graphs will be simple meaning that an edge xy appears only at most once. We do allow both of the edges xy and yx to appear in the directed graphs we consider. For clarity we will omit floor and ceiling signs where they aren't important.

Our digraphs are always simple i.e. they never have two copies of an edge going from a vertex u to a vertex v . A digraph is out-properly labelled if all out-going edges at a vertex have different labels. For a vertex v in a digraph, the out-neighbourhood of v , denoted $N^+(v)$ is the set of $w \in V(D)$ with vw an edge of D .

Throughout the paper we will deal with edge-*coloured* undirected graphs and edge-*labelled* directed graphs. The difference between the two concepts

is that in an edge-coloured graph, the set of possible colours is just some ambient set, whereas in an edge-labelled digraph D the set of possible labels is $V(D) \cup X_0$ where $V(D)$ is the set of vertices of D (and X_0 is some ambient set unrelated to D .) Formally, an edge-labelled directed graph is defined to be a directed graph D together with a set X_0 with $X_0 \cap V(D) = \emptyset$ and a labelling function $f : E(D) \rightarrow V(D) \cup X_0$. The set X_0 is called the set of *non-vertex labels* in D . We call $X_0 \cup V(D)$ the *set of labels* in D (regardless of whether D actually has edges labelled by all elements of $X_0 \cup V(D)$).

Throughout the paper we will always use “ G ” to denote a coloured bipartite graph with parts X and Y and M a rainbow matching in G . We’ll use C_G to denote the set of colours in G and C_M to denote the set of colours in M . We’ll use $V(M)$ to mean the set of vertices contained in edges of M , and $X_0 = X \setminus V(M)$ and $Y_0 = Y \setminus V(M)$ to denote the vertices in X and Y outside M . For a colour $c \in C_M$, we use m_c to denote the colour c edge of M .

3. From bipartite graphs to directed graphs

In this section we show how go from the Aharoni-Berger Conjecture to a problem about edge-labelled digraphs. We define a directed, edge-labelled digraph $D_{G,M}$ corresponding to a coloured bipartite graph G and a rainbow matching M in G .

Definition 3.1 (The directed graph $D_{G,M}$). *Let G be a coloured bipartite graph with parts X and Y and n colours. Let M be a rainbow matching in G . Let $X_0 = X \setminus V(M)$ be the subsets of X disjoint from M . Let C_G be the set of colours used in G and $C_M \subset C_G$ be the set of colours used on edges in M . For a colour $c \in C_M$, we let m_c denote the colour c edge of M . The labelled digraph $D_{G,M}$ corresponding to G and M is defined as follows:*

- *The vertex set of $D_{G,M}$ is the set C_G .*
- *The edges of $D_{G,M}$ are be labelled by elements of the set $X_0 \cup C_M$.*
- *For two colours u and $v \in V(D_{G,M})$ and a vertex $x \in X$, there is a directed edge from u to v in $D_{G,M}$ whenever $v \in C_M$ and there is a colour u edge from x to $m_v \cap Y$.*
 - *If $x \in X_0$ then the edge uv is labelled by x .*

– If $x \in m_c \in M$ then uv is labelled by c , the colour of m_c .

Notice that every edge $e \in E(G)$ corresponds to at most one edge of $D_{G,M}$. There are two types of edges in G which do not correspond to edges of $D_{G,M}$: Edges going through Y_0 do not appear in $D_{G,M}$, and also the edges of M do not appear in $D_{G,M}$ either. Thus the edges of $D_{G,M}$ are naturally in bijection with the edges of $G[X \cup V(M)] \setminus M$. Also notice that if $c \notin C_M$ is a colour which doesn't appear in M , then the in-degree of c in $D_{G,M}$ is zero.

For any set L of labels in $D_{G,M}$ we define a corresponding set $(L)_X$ of vertices in X as follows. For a colour $c \in C_M$ we define $(c)_X$ to be $m_c \cap X$ where m_c is the colour c edge of M . For any vertex $x \in X_0$, we set $(x)_X = \{x\}$. For L a set of labels of $D_{G,M}$, we define $(L)_X = \bigcup_{\ell \in L} (\ell)_X$ i.e. $(L)_X$ is the subset of L consisting of vertices in X_0 together with $M' \cap X$ where M' is the subset of M consisting of edges whose colour is in L .

Notice that with the above definition, if xy is an edge of G and ℓ is the label of the corresponding edge of $D_{G,M}$, then we always have $(\ell)_X = x$. Conversely if uv is an edge of $D_{G,M}$ labelled by ℓ , then the corresponding edge of G goes from $(\ell)_X$ to $m_v \cap Y$. Also notice that $|(L)_X| = |L|$ for any set of labels of $D_{G,M}$.

It turns out that if G is properly coloured, then $D_{G,M}$ is out-properly labelled and simple.

Lemma 3.2. *Let G be a properly edge-coloured bipartite graph and M a matching in G . Then the directed graph $D_{G,M}$ is out-properly labelled and simple.*

Proof. Suppose that uv and uv' are two distinct edges of $D_{G,M}$ with the same label ℓ . By definition of $D_{G,M}$ they correspond to two edges of the form $(\ell)_X y$ and $(\ell)_X y'$ of G having colour u , where $y = m_v \cap Y$ and $y' = m_{v'} \cap Y$. But this contradicts the colouring of G being proper.

Suppose that $D_{G,M}$ is not simple i.e. an edge uv occurs twice with different labels ℓ and ℓ' . This corresponds to two edges of the form $(\ell)_X y$ and $(\ell')_X y$ of G having the same colour u (where $y = m_v \cap Y$.) But this contradicts the colouring of G being proper. \square

We now come to the central objects of study in this paper—switching paths. Switching paths in a labelled digraph D are rainbow paths which have a kind of “consistency” property for the edges they contain which are labelled by vertices of D .

Definition 3.3 (Switching path). *A path $P = (p_0, \dots, p_d)$ in an edge-labelled, directed graph D is a switching path if the following hold.*

- P is rainbow i.e. the edges of P have different labels.
- If $p_i p_{i+1}$ is labelled by a vertex $v \in V(D)$, then $v = p_j$ for some $1 \leq j \leq i$.

Another key definition in this paper is of a label being *amidst* two vertices.

Definition 3.4 (Amidst). *Let u and v be two vertices in an edge-labelled, directed graph D , and c a label. We say that c is amidst u and v if there is a switching path $P = (u, p_1, \dots, p_d, v)$ from u to v such that the following hold.*

- There are no edges of P labelled by c .
- If c is a vertex of D then $c \in \{p_1, \dots, p_d, v\}$.

If P is a path as in Definition 3.4, then we say that P *witnesses* c being amidst u and v . Notice that like in the definition of “switching path”, in the second part of the definition of “amidst” the vertex c is required to be a *non-starting* vertex of P . Also notice that if there is a switching path P from u to v with $|P| \geq 2$, then v is amidst u and v , as witnessed by P .

The following lemma establishes a link between a matching M being maximum in a graph G and the behavior of switching paths in the corresponding digraph $D_{G,M}$.

Lemma 3.5. *Let G be a properly coloured bipartite graph with parts X and Y and M a maximum rainbow matching in G .*

Suppose that M misses a colour c^ and a is a label in $D_{G,M}$ which is amidst c^* and some $v \in V(D_{G,M})$. Then there is no colour v edge in G from $(a)_X$ to $Y_0 = Y \setminus V(M)$.*

Proof. Suppose for the sake of contradiction that a colour v edge $(a)_X y$ exists for $y \in Y_0$. Let P be a switching path witnessing a being amidst c^* and v . Let p_0, p_1, \dots, p_k be the vertex sequence of P with $p_0 = c^*$ and $p_k = v$. For $1 \leq i \leq k$, let m_i be the edge of M with colour p_i . Such edges exist since the in-degree of p_i is positive for $i \geq 1$. For $0 \leq i \leq k-1$ let ℓ_i be the label of $p_i p_{i+1}$ and define $x_i = (\ell_i)_X$. For $0 \leq i \leq k-1$ let e_i be the edge of G corresponding to the edge $p_i p_{i+1}$ of $D_{G,M}$ i.e. e_i is the colour p_i edge going from x_i to $m_{i+1} \cap Y$.

Claim 3.6. *Let $M' = M \cup \{e_0, \dots, e_{k-1}\} \setminus \{m_1, \dots, m_k\}$. Then M' is a rainbow matching in G of size $|M|$ missing the colour v .*

Proof. First we show that M' is a rainbow set of edges missing the colour v . Notice that for each $i \geq 1$, e_i and m_i both have colour p_i . Also the edge e_0 has colour $p_0 = c^*$. Since M is rainbow and missed colour c^* , $M \setminus \{m_1, \dots, m_k\}$ is rainbow and misses the colours c^*, p_1, \dots, p_k . Therefore M' is rainbow and misses colour $p_k = v$.

It remains to show that M' is a matching. Notice that $M \setminus \{m_1, \dots, m_k\}$ is a matching as a consequence of M being a matching.

Next we show that $\{e_0, \dots, e_{k-1}\}$ is a matching. Since P is a switching path, its edges have different labels, which is equivalent to the vertices x_0, \dots, x_{k-1} being distinct. Also, since P is a path, the vertices p_1, \dots, p_k are distinct which implies that the edges m_1, \dots, m_k are also distinct. Since e_i goes from x_i to $m_{i+1} \cap Y$, these imply that for distinct i and j we have $e_i \cap e_j = \emptyset$.

Finally we show that $e_i \cap m = \emptyset$ for $0 \leq i \leq k-1$ and $m \in M \setminus \{m_1, \dots, m_k\}$. Suppose that $e_i \cap m \cap Y \neq \emptyset$. Then since $e_i \cap Y = m_{i+1} \cap Y$, we have $m = m_{i+1}$ which contradicts $m \in M \setminus \{m_1, \dots, m_k\}$. Suppose that $e_i \cap m \cap X \neq \emptyset$, or equivalently $e_i \cap m \cap X = \{x_i\}$. Then $x_i \in V(M)$ which is equivalent to $\ell_i \in V(D_{G,M})$. In particular we find out that ℓ_i is a colour in G (rather than a vertex of X_0 .) Recall that ℓ_i is the label of the edge $p_i p_{i+1}$ of P . Using the definition of P being a switching path, we get that $\ell_i = p_j$ for some $1 \leq j \leq i$. Then m_j is the colour ℓ_i edge of M which gives $x_i = (\ell_i)_X \in m_j \cap X$. This implies that $m = m_j$ which contradicts $m \in M \setminus \{m_1, \dots, m_k\}$. \square

We claim that $(a)_X \notin V(M')$. Since a is amidst c^* and v , we have that a doesn't appear on edges of P , which implies $(a)_X \neq x_i$ for $0 \leq i \leq k-1$. This shows that $(a)_X$ is disjoint from e_0, \dots, e_{k-1} . If $a \notin C_M$, then we have $a \in X_0$ and so $(a)_X \notin V(M)$ which gives $(a)_X \notin V(M')$. If $a \in C_M$, then since a is amidst c^* and v , we have that $a = p_i$ for some $1 \leq i \leq k$. This gives $(a)_X = m_i \cap X$ which implies that $(a)_X \notin V(M \setminus \{m_1, \dots, m_k\})$ and hence $(a)_X \notin V(M')$.

Now we have that neither of the vertices $(a)_X$ or y are in M' , and also M' misses colour v . Thus $M' + (a)_X y$ is a rainbow matching of size $|M'| + 1$ contradicting the maximality of M . \square

For a set of labels L in a labelled digraph D , define

$$N_L^+(v) = \{w \in N^+(v) : vw \text{ is labelled by some } \ell \in L\}.$$

Notice that from the definition of $D_{G,M}$, we have that $|N_L^+(v)|$ is exactly the number of colour v edges in G going from $(L \setminus \{v\})_X$ to $Y \cap V(M)$.

The following corollary of the above lemma shows that if we have a graph G with a maximum matching missing some colour, then the corresponding digraph $D_{G,M}$ satisfies a degree condition.

Lemma 3.7. *Let G be a properly coloured bipartite graph with parts X and Y , M a maximum rainbow matching in G , and $X_0 = X \setminus V(M)$.*

Suppose that M misses a colour c^ , v is a colour in G with $|M| + k$ edges, and A is a set of labels in $D_{G,M}$ which are amidst c^* and v . Then $|N_A^+(v)| \geq |A| - |X_0| + k - 1$.*

Proof. Suppose for the sake of contradiction that $|N_A^+(v)| < |A| - |X_0| + k - 1$. Since $|(A)_X| = |A|$, there are exactly $|M| + |X_0| - |A|$ vertices in X outside $(A)_X$. The number of colour v edges touching $(A \setminus \{v\})_X$ in G is $\geq |M| + k - |X \setminus (A \setminus \{v\})_X| \geq |M| + k - |X \setminus (A)_X| - 1 = |A| - |X_0| + k - 1 > |N_A^+(v)|$. Since $|N_A^+(v)|$ equals the number of colour v edges between $(A \setminus \{v\})_X$ and $Y \cap V(M)$ we obtain that there is a colour v edge from some $(a)_X \in (A \setminus \{v\})_X$ to $y \in Y_0 = Y \setminus V(M)$. But, by definition of A we have a amidst c^* and v , contradicting Lemma 3.5. \square

The above lemma produces a directed graph with a degree condition. In the remainder of the paper we show that this degree condition is almost too strong to hold. We show that if k is linear in $|G|$ for every vertex v and $|G|$ is sufficiently large, then no digraphs satisfying the conclusion of Lemma 3.7 exists. This is equivalent to there being no graphs with a maximum matching M satisfying the assumptions of Lemma 3.7 i.e. we obtain that any maximum matching in such a graph must use every colour.

4. Connectivity of labelled, directed graphs

The goal of this section is to prove the following theorem. Together with Lemmas 3.2 and 3.7, it immediately implies Theorem 1.3.

Theorem 4.1. *For all ϵ with $0 < \epsilon \leq 0.9$, there is a $N_0 = N_0(\epsilon)$ such that the following holds. Let D be any out-properly edge-labelled, simple, directed*

graph on $n \geq N_0$ vertices. Let X_0 be the set of labels which are not vertices of D

Then for all $u \in V(D)$, there is a vertex v and a set of labels A amidst u and v , such that $|N_A^+(v)| < |A| - |X_0| + \epsilon n$.

Throughout this section, for a set of vertices S in a graph D we denote the vertex-complement of S by $\bar{S} = V(D) \setminus S$.

For a path P , define a corresponding set of labels \underline{P} consisting of labels which are either vertices of P or labels of edges of P . Formally $\underline{P} = V(P) \cup \{\ell : \ell \text{ is the label of some } e \in E(P)\}$ denotes the set of labels consisting of $V(P)$ together with the set of labels of edges of P . For a path of length d , we will often use the bound $|\underline{P}| \leq 2d + 1 \leq 3d$. For a set of labels S and a path P starting at a vertex v , we say that P *avoids* S if $S \cap \underline{P} \subseteq \{v\}$ i.e. P has no edges labelled by elements of S and P has no vertices in S *except possibly the starting vertex* v .

The condition that S is allowed to contain the starting vertex of P in the definition of “avoids” may seem strange. We have this condition since it makes many of the arguments in this paper neater. In particular, it allows us to cleanly concatenate switching paths with the following lemma.

Lemma 4.2. *Let $P = (p_0, p_1, \dots, p_t)$ and $Q = (p_t, p_{t+1}, \dots, p_s)$ be two switching paths in a labelled digraph D . If Q avoids \underline{P} then $P + Q = (p_1, p_2, \dots, p_s)$ is also a switching path.*

Proof. To see that $P + Q$ is rainbow, notice that P and Q are rainbow and that Q shares no edge-labels with P since Q avoids \underline{P} . To see the second part of the definition of $P + Q$ being a switching path notice that if $p_i p_{i+1}$ is labelled by $v \in V(D)$, then depending on whether $v \in P$ or $v \in Q$ we have $v = p_j$ for $1 \leq j \leq i$ or $v = p_j$ for $t + 1 \leq j \leq i$. \square

Another consequence of the definition of “avoids” is that for any set of labels S , a single vertex path $P = v$ is a path from v to v avoiding S .

The proof of Theorem 4.1 involves lots of constants. The first constant which we use is ϵ which is the constant given to us by Theorem 4.1. Throughout the section it is best to fix ϵ with $0 < \epsilon \leq 0.9$, and to read everything that we do as a proof of Theorem 4.1 for that particular ϵ .

Next we introduce three numbers N_0 , λ_{\max} and δ depending on ϵ whose relationship is $N_0^{-1} \ll \lambda_{\max}^{-1} \ll \delta \ll \epsilon$. The number N_0 will be the N_0 in Theorem 4.1, while λ_{\max} and δ are just two numbers with no special meaning.

We set $\lambda_{\max} = 4^{\epsilon^{-9}}$ and $\delta = \epsilon^3$. For an integer x let $\text{twr}(x) = \lambda_{\max}^{\lambda_{\max}^{\dots^{\lambda_{\max}}}}$ be the tower function, where there are x exponentiations. Set $N_0 = \text{twr}(2\lambda_{\max})$.

Next for any $\lambda \in \mathbb{N}$, we define four numbers $d_\lambda, k_\lambda, \Delta_\lambda$, and γ_λ . These numbers will control the variables in our definition of connectedness and will allow us to define “reaching” and “components” using just one parameter (rather than using four as in Section 2.3.) The specific definitions of $d_\lambda, k_\lambda, \Delta_\lambda$, and γ_λ are not too important—the intuition is that for any $\lambda \in [1, \lambda_{\max}]$ we have

$$\begin{aligned} \lambda_{\max} &\ll d_\lambda \ll k_\lambda \ll \Delta_\lambda \ll \gamma_\lambda^{-1} \ll N_0, \\ d_{\lambda+1} &\gg d_\lambda, & k_{\lambda+1} &\ll k_\lambda, \\ \Delta_{\lambda+1} &\gg \Delta_\lambda, & \gamma_{\lambda+1} &\ll \gamma_\lambda. \end{aligned}$$

Notice that some sort of upper bound on λ is necessary for all the above to hold since $d_{\lambda+1} \gg d_\lambda$, $k_{\lambda+1} \ll k_\lambda$, and $d_\lambda \ll k_\lambda$ cannot simultaneously hold for all $\lambda \in \mathbb{N}$. Because of this, in all our lemmas we will make sure that λ is in the range $1 \leq \lambda \leq \lambda_{\max}$.

For specific $d_\lambda, k_\lambda, \Delta_\lambda$, and γ_λ with which our proofs work, define

$$\begin{aligned} d_\lambda &= \lambda_{\max} \cdot 4^\lambda, \\ k_\lambda &= \lambda_{\max}(4^{\lambda_{\max}^4} - 4^{\lambda_{\max} \cdot \lambda}), \\ \Delta_\lambda &= \text{twr}(\lambda), \\ \gamma_\lambda &= \text{twr}(\lambda + 2)^{-1}. \end{aligned}$$

To prove Theorem 4.1 we will need a careful understanding of the switching paths in a labelled digraph. We will study switching paths via a new notion of connectedness which we now introduce. The following is the heart of the notion of connectedness that we study.

Definition 4.3 (λ -reaching). *For a vertex v in a labelled digraph D and a set $R \subseteq V(D)$, we say that v λ -reaches R if for any set S of $\leq k_\lambda$ labels, there are length $\leq d_\lambda$ switching paths avoiding S to all, except possibly at most Δ_λ , vertices $x \in R$.*

Notice that v λ -reaches a set R exactly when it $(k_\lambda, d_\lambda, \Delta_\lambda)$ -reaches R , as defined in Section 2.3. To complement the notion of “reaching” we introduce

a notion of “bypassing”. Informally a set B is bypassed by a vertex v if v doesn’t reach anything big inside B .

Definition 4.4 (λ -bypassing). *For a vertex v in a labelled digraph D and a set $B \subseteq V(D)$, we say that v λ -bypasses B if v doesn’t λ -reach any $R \subseteq B$ with $|R| \geq \gamma_\lambda |D|$.*

The third key definition is that of a λ -component. Recall that when studying ordinary undirected graphs a connected component C is a set where every pair $x, y \in C$ is connected by a path, and no pair $x \in C, z \notin C$ is connected by a path. Intuitively a λ -component is similar to this, with “every pair” replaced by “almost every pair” and “no pair” replaced by “almost no pair”.

Definition 4.5 (λ -component). *A set of vertices C in a labelled digraph D is a λ -component if for any vertex $v \in C$, there is a set $R_v \subseteq V(D)$ with $|R_v \triangle C| \leq \delta n$ such that the following hold.*

- (i) v λ -reaches R_v .
- (ii) v $(\lambda - 3)$ -bypasses $\overline{R_v}$.

Notice that for a labelled digraph D it is far from clear that λ -components exist in D . Section 4.2 will be devoted to proving that every properly labelled digraph D , has λ -components for suitable λ .

4.1. Basic properties

Here we establish many basic properties of λ -reaching, λ -bypassing, and λ -components. The first property is that reaching or bypassing a set W is preserved by passing to a subset of W , and by moving λ in a suitable direction.

Observation 4.6. *Let D be a labelled digraph, $v \in V(D)$, $R, B \subseteq V(D)$, and $\lambda \in \mathbb{N}$.*

- (i) **Monotonicity of reaching:** *Let $\lambda^+ \geq \lambda$ and $R^- \subseteq R$. Then v λ -reaches $R \implies v$ λ^+ -reaches R^- .*
- (ii) **Monotonicity of bypassing:** *Let $\lambda^- \leq \lambda$ and $B^- \subseteq B$. Then v λ -bypasses $B \implies v$ λ^- -bypasses B^- .*

Proof. For (i), let S be a set of k_{λ^+} labels. Since $k_{\lambda^+} \leq k_\lambda$ and v λ -reaches R , there are length $\leq d_\lambda$ switching paths from v to all except at most $\leq \Delta_\lambda$ vertices in R . Since $d_{\lambda^+} \geq d_\lambda$ and $\Delta_{\lambda^+} \geq \Delta_\lambda$, these same paths give length $\leq d_{\lambda^+}$ switching paths from v to all except at most $\leq \Delta_{\lambda^+}$ vertices in $R^- \subseteq R$.

For (ii), let R be a subset of B^- which is λ^- -reached by v . Since $\lambda \geq \lambda^-$, by part (i) we know that v λ -reaches R . Since $R \subseteq B^- \subseteq B$ and v λ -bypasses B , we get that $|R| \leq \gamma_\lambda |D|$. Since $\gamma_{\lambda^-} \geq \gamma_\lambda$ we get that $|R| \leq \gamma_{\lambda^-} |D|$. Since R was an arbitrary subset of B^- which is λ^- -reached by v , we have proved that v λ^- -bypasses B^- . \square

Since the above observation is extremely fundamental and basic we will not always explicitly refer to it throughout its many applications. The next observation provides trivial conditions for sets to be reached or bypassed by a vertex.

Observation 4.7. *Let D be a labelled digraph, X_0 the set of non-vertex labels in D , $v \in V(D)$, $R, B \subseteq V(D)$, and $\lambda \in \mathbb{N}$.*

- (i) **Reaching small sets:** $|R| \leq \Delta_\lambda \implies v$ λ -reaches R .
- (ii) **Bypassing small sets:** $|B| < \gamma_\lambda |D| \implies v$ λ -bypasses B .
- (iii) **Reaching neighborhoods:** *If D is out-properly labelled then v 3-reaches $N_{X_0}^+(v)$.*

Proof. For (i), we can take the family of paths for the definition of v λ -reaching R to be empty. For (ii), notice that every subset $R \subseteq B$ has $|R| \leq |B| < \gamma_\lambda |D|$ regardless of whether R is reached by v or not.

For (iii), notice that $\Delta_3 \geq k_3$. Let S be a set of $\leq k_3$ labels. Notice that for every $y \in N_{X_0 \setminus S}^+(v) \setminus S$, the edge vy is a length $1 \leq d_3$ switching path from v to y avoiding S . Since D is properly coloured we have $|N_{X_0}^+(v) \setminus (N_{X_0 \setminus S}^+(v) \setminus S)| \leq |S| \leq k_3 \leq \Delta_3$, and so we have enough paths for the definition of v λ -reaching $N_{X_0}^+(v)$. \square

Observation 4.6 shows that if R is λ -reached by v , we can pass to a subset of R and still have it λ -reached. We will sometimes want to increase the size of a set R and know that it is still reached by v . The following lemma shows that we can add the vertex v itself to R and still know that $R \cup \{v\}$ is reached by v with the same parameter λ .

Observation 4.8 (Reaching one more vertex). v λ -reaches $R \implies v$ λ -reaches $R \cup \{v\}$.

Proof. Let S be a set of $\leq k_\lambda$ labels. Recall that $\{v\}$ is a length $0 \leq d_\lambda$ switching path from v to v avoiding S (using the fact that the first vertex of a path is allowed to be in S in the definition of “avoids”.) Also, since v λ -reaches R , there are length $\leq d_\lambda$ switching paths avoiding S to all except at most Δ_λ vertices of R . These paths, together with $\{v\}$, give the required paths to show that v λ -reaches $R \cup \{v\}$. \square

Consider a set R_v as in the definition of λ -component i.e. R_v is λ -reached by v and $\overline{R_v}$ is $(\lambda - 3)$ -bypassed by v . By Observation 4.8 $R_v \cup \{v\}$ is λ -reached by v , and by the monotonicity of bypassing $\overline{R_v \cup \{v\}}$ is $(\lambda - 3)$ -bypassed by v . This shows that without affecting anything we could have added the condition “ $v \in R_v$ ” to the definition of λ -component.

The next two lemmas show that reaching and bypassing are preserved by unions, as long as we weaken the parameter λ slightly.

Lemma 4.9 (Reaching unions). For $m \leq \gamma_\lambda^{-1}$, suppose that a vertex v λ -reaches sets $R_1, \dots, R_m \subseteq V(D)$. Then v $(\lambda + 3)$ -reaches $\bigcup_{i=1}^m R_i$.

Proof. Let S be a set of $k_{\lambda+3}$ labels. Since $k_{\lambda+3} \leq k_\lambda$ and v λ -reaches R_i , there are length $\leq d_\lambda$ switching paths avoiding S to all, except possibly Δ_λ , vertices $x \in R_i$ for each i . Therefore there are length $\leq d_\lambda \leq d_{\lambda+3}$ switching paths avoiding S to all, except possibly $m\Delta_\lambda \leq \gamma_\lambda^{-1}\Delta_\lambda = \text{twr}(\lambda+2)\text{twr}(\lambda) \leq \text{twr}(\lambda+3) = \Delta_{\lambda+3}$ vertices in $\bigcup_{i=1}^m R_i$. \square

A similar lemma holds for bypassing.

Lemma 4.10 (Bypassing unions). For $m \leq \gamma_{\lambda-1}^{-1}$, suppose that a vertex v λ -bypasses sets $B_1, \dots, B_m \subseteq V(D)$. Then v $(\lambda - 1)$ -bypasses $\bigcup_{i=1}^m B_i$.

Proof. Suppose that v $(\lambda - 1)$ -reaches a set $R \subseteq \bigcup_{i=1}^m B_i$. Without loss of generality we can suppose that B_1, \dots, B_m are ordered so that $|R \cap B_1| \geq |R \cap B_j|$ for $j > 1$. Since $R = \bigcup_{i=1}^m R \cap B_i$ we have $|R \cap B_1| \geq |R|/m$. From the monotonicity of reaching, v λ -reaches $R \cap B_1$. Since v λ -bypasses B_1 , this implies $|R \cap B_1| < \gamma_\lambda |D|$. This gives $|R| \leq m|R \cap B_1| \leq m\gamma_\lambda |D| \leq \gamma_{\lambda-1}^{-1} \gamma_\lambda |D| = \text{twr}(\lambda+1)\text{twr}(\lambda+2)^{-1}|D| \leq \text{twr}(\lambda+1)^{-1}|D| = \gamma_{\lambda-1} |D|$. Since R was arbitrary, we have proved that v $(\lambda - 1)$ -bypasses $\bigcup_{i=1}^m B_i$. \square

Recall that “two vertices u and v being connected by a path” is a transitive relation on vertices in an graph. This transitivity is used to show that connected components in a graph are equivalence classes. The following lemma shows that “reaching” also has a kind of transitive property. The lemma plays a similar role in showing that λ -components exist.

Lemma 4.11 (Transitivity of reaching). *For a labelled digraph D and $1 \leq \lambda \leq \lambda_{\max}$, suppose that we have a vertex $v \in V(D)$, and R such that v λ -reaches R . Suppose that we have distinct vertices $x_0, \dots, x_{\Delta_\lambda} \in R$ and a set W such that x_i λ -reaches W for each i . Then v $(\lambda + 1)$ -reaches W .*

Proof. Let S be a set of $k_{\lambda+1}$ labels. Since v λ -reaches R and $k_{\lambda+1} \leq k_\lambda$, there is some $i \in 0, \dots, \Delta_\lambda$ such that there is a length $\leq d_\lambda$ switching path P from v to x_i avoiding S . Since x_i λ -reaches W and $|S| + |\underline{P}| \leq k_{\lambda+1} + 3d_\lambda \leq k_\lambda$, there is a length $\leq d_\lambda$ switching path P_w avoiding S and \underline{P} from x_i to all, except Δ_λ vertices of $w \in W$.

Using Lemma 4.2, the paths $P + P_w$ are length $\leq 2d_\lambda \leq d_{\lambda+1}$ switching paths avoiding S to all except at most $\Delta_\lambda \leq \Delta_{\lambda+1}$ vertices $w \in W$. This proves the lemma. \square

A consequence of Observation 4.7 (iii) is that components cannot be much smaller than the neighborhoods of vertices they contain.

Lemma 4.12 (Components are larger than neighbourhoods). *Let D be a out-properly labelled, directed graph on n vertices with X_0 the set of non-vertex labels in D , and $6 \leq \lambda \leq \lambda_{\max}$. For any λ -component C and $v \in C$ we have $|C| \geq |N_{X_0}^+(v)| - \delta n - \gamma_{\lambda-3}n$.*

Proof. By the definition of C being a λ -component, there is a set R_v with $|R_v \setminus C| \leq \delta n$ such that v $(\lambda - 3)$ -bypasses $\overline{R_v}$. By Observation 4.7 (iii) and the monotonicity of reaching, v $(\lambda - 3)$ -reaches $N_{X_0}^+(v)$. This gives $|N_{X_0}^+(v) \cap \overline{R_v}| \leq \gamma_{\lambda-3}n$ which implies the lemma:

$$\begin{aligned} |C| &\geq |N_{X_0}^+(v)| - |N_{X_0}^+(v) \cap \overline{C}| \\ &\geq |N_{X_0}^+(v)| - |N_{X_0}^+(v) \cap \overline{C} \cap \overline{R_v}| - |\overline{C} \setminus \overline{R_v}| \geq |N_{X_0}^+(v)| - \delta n - \gamma_{\lambda-3}n. \end{aligned}$$

\square

4.2. Constructing λ -components

The goal of this section is to show that λ -components exist for suitable λ . The first step towards this is to show that for any vertex v and number λ , there is a set $R_v \subseteq V(D)$ possessing the two properties R_v has in the definition of λ -component.

Lemma 4.13. *For all vertices v in a labelled digraph D and $1 \leq \lambda \leq \lambda_{\max}$, there is a set $R \subseteq V(D)$ such that v $(\lambda + 3)$ -reaches R and v λ -bypasses \overline{R} .*

Proof. We define sets of vertices $R_0, R_1, R_2, \dots, R_m$ recursively as follows.

- Let $R_0 = \emptyset$.
- For each $i \geq 1$, if possible, choose R_i to be any set disjoint from $R_0 \cup \dots \cup R_{i-1}$ which is λ -reached by v , and also $|R_i| \geq \gamma_\lambda |D|$.
- Otherwise, if no such R_i exists, we stop with $m = i - 1$.

Notice that the sets R_1, \dots, R_m are all disjoint and satisfy $|R_i| \geq \gamma_\lambda |D|$ which implies that $m \leq \gamma_\lambda^{-1}$. Set $R = R_1 \cup \dots \cup R_m$.

By definition of m , v λ -bypasses \overline{R} —indeed otherwise we could choose a set R_{m+1} of size $\gamma_\lambda |D|$ disjoint from R which is λ -reached by v , contradicting the fact that we stopped at m . By Lemma 4.9 v $(\lambda + 3)$ -reaches R . This completes the proof. \square

Notice that in the above lemma would be stronger if it produced a set R with R λ -reached by v and \overline{R} λ' -bypassed by v for $\lambda' > \lambda$ (rather than $\lambda' < \lambda$ as Lemma 4.13 gives us.) The next lemma tries to prove something like this—it produces two sets R and B which are “nearly complementary” such that v λ -reaches R and λ' -bypasses B for $\lambda' > \lambda$.

Lemma 4.14. *Let D be a labelled digraph on n vertices, $\lambda_0 \in \mathbb{N}$ with $43\delta^{-1} \leq \lambda_0 \leq \lambda_{\max}$, and $v \in V(D)$. There are two sets of vertices R and B satisfying the following.*

(i) $|V(D) \setminus (R \cup B)| \leq \delta n / 3$.

(ii) *There is a λ with $\lambda_0 - 42\delta^{-1} \leq \lambda \leq \lambda_0$ such that*

- v $(\lambda - 4)$ -reaches R .
- v λ -bypasses B .

Proof. Define $\lambda_1, \dots, \lambda_{6\delta-1}, R_1, \dots, R_{6\delta-1}$ as follows.

- For each i , set $\lambda_i = \lambda_{i-1} - 7$.
- Let R_i be a set which is $(\lambda_i + 3)$ -reached by v and with $\overline{R_i}$ λ_i -bypassed by v . Such a set exists by Lemma 4.13.

We show that there is some index m satisfying a property like part (i) of the lemma.

Claim 4.15. *There is some $m \in \{1, \dots, 6\delta-1\}$ for which $|V(D) \setminus (R_m \cup \overline{R_{m-1}})| \leq \delta n/3$.*

Proof. Suppose for the sake of contradiction that $|V(D) \setminus (R_i \cup \overline{R_{i-1}})| > \delta n/3$ for all $i = 1, \dots, 6\delta-1$. Notice that we have $|R_i \cap \overline{R_{i-1}}| \leq \gamma_{\lambda_{i-1}} n$ (since v λ_{i-1} -bypasses $\overline{R_{i-1}}$, v $(\lambda_i + 3)$ -reaches $R_i \cap \overline{R_{i-1}}$, and $\lambda_i + 3 \leq \lambda_{i-1}$.) We also have $|R_{i-1} \setminus R_i| = |V(D) \setminus (R_i \cup \overline{R_{i-1}})| > \delta n/3$ for $i < 6\delta-1$. Combining these, we get the following

$$\begin{aligned} |R_i| &= |R_i \cap R_{i-1}| + |R_i \cap \overline{R_{i-1}}| \\ &= |R_{i-1}| - |R_{i-1} \setminus R_i| + |R_i \cap \overline{R_{i-1}}| < |R_{i-1}| - (\delta/3 - \gamma_{\lambda_{i-1}})n. \end{aligned}$$

Notice that for all $i \leq 6\delta-1$ we have $\lambda_i \geq 1$, and so $\gamma_{\lambda_i} \leq \lambda_{\max}^{-1} \leq \delta/12$ which implies $|R_i| < |R_{i-1}| - \delta n/4$. Iterating this gives $0 \leq |R_i| < |R_1| - (i-1)\delta n/4 \leq n - (i-1)\delta n/4$. This is a contradiction for $i = 6\delta-1$. \square

Set $R = R_m$, $B = \overline{R_{m-1}}$ and $\lambda = \lambda_{m-1}$. Then v $(\lambda - 4)$ -reaches R_i and λ -bypasses B by the constructions of R_i and R_{i-1} . We have $|V(D) \setminus (R \cup B)| \leq \delta n/3$ by choice of m . Finally we have $\lambda_0 \geq \lambda \geq \lambda_0 - 7m \geq \lambda_0 - 42\delta-1$. \square

As a prelude to constructing components we give a condition under which a singleton $\{v\}$ is a λ -component.

Lemma 4.16. *Let D be a labelled digraph on $n \geq N_0$ vertices and $4 \leq \lambda \leq \lambda_{\max}$. Suppose that v λ -bypasses B with $|\overline{B}| \leq \delta n/2$. Then $\{v\}$ is a λ -component.*

Proof. To prove the lemma we need to choose a set R_v and show that it satisfies all the properties of the set R_v in the definition of λ -component. Apply Lemma 4.13 to get a set R_v which is λ -reached by v and with $\overline{R_v}$ $(\lambda - 3)$ -bypassed by v . Notice that since v λ -bypasses B , we must have $|R_v \cap B| \leq \gamma_{\lambda} n$. This gives $|R_v \triangle \{v\}| \leq |R_v| + 1 \leq |R_v \cap B| + |\overline{B}| + 1 \leq \gamma_{\lambda} n + \delta n/2 + 1 \leq \delta n$. \square

The following lemma is a purely technical tool which we will need. A r -uniform multihypergraph \mathcal{H} with n vertices and m edges is a family of m size r subsets of $[n]$ with the possibility of \mathcal{H} containing several copies of the same subset.

Lemma 4.17. *Let \mathcal{H} be a γn -uniform multihypergraph with n vertices and m edges. Then, for any t with $\gamma/2 \geq 2t/m$, there are t edges $T_1, \dots, T_t \in \mathcal{H}$ with $|T_1 \cap \dots \cap T_t| \geq (\frac{\gamma}{2})^t n$.*

Proof. Let \mathcal{T} be a set of t distinct edges of \mathcal{H} chosen uniformly at random from all such sets. To prove the lemma it is sufficient to show that the expected size of the intersection of the edges in \mathcal{T} is at least $(\frac{\gamma}{2})^t n$. For any vertex $v \in V(\mathcal{H})$, let $d(v)$ be the number of edges of \mathcal{H} containing v . Let $V_{\geq t} = \{v \in V(\mathcal{H}) : d(v) \geq t\}$. By linearity of expectation we have the following.

$$\begin{aligned} \mathbb{E} \left(\left| \bigcap_{E \in \mathcal{T}} E \right| \right) &= \sum_{v \in V(\mathcal{H})} \mathbb{P} \left(v \in \bigcap_{E \in \mathcal{T}} E \right) = \sum_{v \in V_{\geq t}} \mathbb{P} \left(v \in \bigcap_{E \in \mathcal{T}} E \right) \\ &= \sum_{v \in V_{\geq t}} \frac{\binom{d(v)}{t}}{\binom{m}{t}} \geq \sum_{v \in V_{\geq t}} \left(\frac{(d(v) - t)}{m} \right)^t. \end{aligned}$$

The inequality comes from “ $\binom{d}{t}/\binom{m}{t} \geq ((d-t)/m)^t$ for $d \geq t$.” Using convexity of $f(x) = x^t$ we can prove the lemma.

$$\begin{aligned} \mathbb{E} \left(\left| \bigcap_{E \in \mathcal{T}} E \right| \right) &\geq \sum_{v \in V_{\geq t}} \left(\frac{(d(v) - t)}{m} \right)^t \geq \left(\sum_{v \in V_{\geq t}} \frac{(d(v) - t)}{nm} \right)^t n \\ &\geq \left(\frac{\gamma m - 2t}{m} \right)^t n \geq \left(\frac{\gamma}{2} \right)^t n. \end{aligned}$$

The third inequality uses $\sum_{v \in V(G)} d(v) = \gamma nm$ and $\sum_{v \in V(G) \setminus V_{\geq t}} d(v) \leq tn$. The last inequality comes from $\gamma/2 \geq 2t/m$. \square

The following lemma is the main result of this section. It implies that for a given λ_0 , there is a λ -component C for some λ which is close to λ_0 . In addition the lemma gives some control over where the component C is located—given any set B_0 which is λ_0 -bypassed, we can choose C to be outside B_0 .

Lemma 4.18. *Let D be a labelled digraph on $n \geq N_0$ vertices and $87\delta^{-2} \leq \lambda_0 \leq \lambda_{\max}$. Suppose we have $v_0 \in V(D)$ and $B_0 \subseteq V(D)$ such that $v_0 \in \overline{B_0}$ and v_0 λ_0 -bypasses B_0 . Then there is a nonempty $C \subseteq \overline{B_0}$ such that C is a λ -component with $\lambda_0 - 87\delta^{-2} \leq \lambda \leq \lambda_0$.*

Proof. We start with the following claim.

Claim 4.19. *There is a vertex $v' \in \overline{B_0}$, $\lambda' \in [\lambda_0 - 86\delta^{-2}, \lambda_0]$, and a set $B' \supseteq B_0$ with the following properties.*

- v' λ' -bypasses B' .
- For every $u \in \overline{B'}$, if there is a set $B_u \supset B'$ such that u $(\lambda' - 43\delta^{-1})$ -bypasses B_u , then $|B_u| < |B'| + \delta n/2$.

Proof. Using B_0 , v_0 , and λ_0 from the lemma, we define B_1, \dots, B_m , v_1, \dots, v_m , and $\lambda_1, \dots, \lambda_m$ as follows.

- For each i , set $\lambda_{i+1} = \lambda_i - 43\delta^{-1}$.
- For each i , if possible, choose a vertex $v_{i+1} \in \overline{B_i}$ and a set $B_{i+1} \supset B_i$ such that v_{i+1} λ_{i+1} -bypasses B_{i+1} and $|B_{i+1}| \geq |B_i| + \delta n/2$.
- Otherwise, if no such pair of v_{i+1} and B_{i+1} exists, then stop with $m = i$.

Notice that since $|B_{i+1}| \geq |B_i| + \delta n/2$ for $i < m$, we stop with $m \leq 2\delta^{-1}$. Let $\lambda' = \lambda_m$, $v' = v_m$, and $B' = B_m$. Since $m \leq 2\delta^{-1}$, we have $\lambda' = \lambda_0 - 43\delta^{-1}m \geq \lambda_0 - 86\delta^{-2}$. We have $B' = B_m \supseteq B_{m-1} \supseteq \dots \supseteq B_0$ and $v' \in \overline{B_{m-1}} \subseteq \overline{B_0}$ as required. The vertex v' λ' -bypasses B' by choice of v_m and B_m . The fact that “for every $u \in \overline{B'}$, if there is a set $B_u \supset B'$ such that u $(\lambda' - 43\delta^{-1})$ -bypasses B_u , then $|B_u| < |B'| + \delta n/2$ ” is equivalent to the fact that we stopped at m . \square

Apply Lemma 4.14 to v' and λ' in order to obtain sets R and B and $\lambda'' \in [\lambda' - 42\delta, \lambda']$ such that $|V(D) \setminus (R \cup B)| \leq \delta n/3$, v' $(\lambda'' - 4)$ -reaches R , and v' λ'' -bypasses B . By Lemma 4.10 and the monotonicity of bypassing, v' $(\lambda'' - 1)$ -bypasses $B \cup B'$.

We make the following definition

$$S = \{x \in R : x \text{ } (\lambda'' - 2)\text{-reaches some } T \subseteq B \cup B' \text{ with } |T| \geq \gamma_{\lambda''-3}n\}.$$

From the definition of S and the monotonicity of reaching, we have that for every $v \in R \setminus S$ the vertex v $(\lambda'' - 3)$ -bypasses $B \cup B'$. Using Lemma 4.17 and the “transitivity of reaching” we show that S is small.

Claim 4.20. $|S| \leq 4\Delta_{\lambda''}\gamma_{\lambda''}^{-1}$.

Proof. Suppose for the sake of contradiction that $|S| > 4\Delta_{\lambda''}^2\gamma_{\lambda''}^{-1}$. For each $s \in S$, choose some set $T_s \subseteq B \cup B'$ with $|T_s| = \gamma_{\lambda''-3}n$ which is $(\lambda'' - 2)$ -reached by s . Let $\mathcal{H} = \{T_s : s \in S\}$. Notice that \mathcal{H} is an $(\gamma_{\lambda''-3}n)$ -uniform multihypergraph with $|S|$ edges. Notice that $|S| > 4\Delta_{\lambda''}\gamma_{\lambda''}^{-1}$ implies $\gamma_{\lambda''-3}/2 \geq 2(\Delta_{\lambda''-2} + 1)/|S|$. Therefore we can apply Lemma 4.17 to \mathcal{H} with $t = \Delta_{\lambda''-2} + 1$ and $\gamma = \gamma_{\lambda''-3}$ in order to find $\Delta_{\lambda''-2} + 1$ sets $T'_0, \dots, T'_{\Delta_{\lambda''-2}} \in \mathcal{H}$ with

$$\left| \bigcap_{i=0}^{\Delta_{\lambda''-2}} T'_i \right| \geq \left(\frac{\gamma_{\lambda''-3}}{2} \right)^{\Delta_{\lambda''-2}+1} n > \gamma_{\lambda''-1}n.$$

The second inequality comes from the fact that $\gamma_{\lambda''-1}^{-1} = \text{twr}(\lambda'' + 1) \geq (2\text{twr}(\lambda'' - 1))^{\text{twr}(\lambda''-2)+1} = (2/\gamma_{\lambda''-3})^{\Delta_{\lambda''-2}+1}$. By Lemma 4.11 applied with $\lambda = \lambda'' - 2$, $v = v'$, $R = R$, $W = \bigcap_{i=0}^{\Delta_{\lambda''-2}} T'_i$, and x_i the vertex of S which $(\lambda'' - 2)$ -reaches T'_i , we get that v' $(\lambda'' - 1)$ -reaches $\bigcap_{i=0}^{\Delta_{\lambda''-2}} T'_i$. This contradicts v' $(\lambda'' - 1)$ -bypassing $B \cup B'$. \square

Let $C = R \setminus (S \cup B' \cup B)$ and $\lambda = \lambda'' - 2$. Since $C \subseteq \overline{B'}$ and $B' \supseteq B_0$ we have $C \subseteq \overline{B_0}$. From the definitions of λ' and λ'' we have that $\lambda_0 - 87\delta^{-2} \leq \lambda \leq \lambda_0$.

Claim 4.21. C is a λ -component.

Proof. For each $v \in C$ apply Lemma 4.13 to get a set R_v which is λ -reached by v and with $\overline{R_v}$ $(\lambda - 3)$ -bypassed by v . To prove the claim, it is enough to show that $|R_v \triangle C| \leq \delta n$. We'll do this by showing $|R_v \setminus C| \leq \delta n/2$ and $|C \setminus R_v| \leq \delta n/2$.

First we show that $|R_v \setminus C| \leq \delta n/2$. Notice that $C = R \cap \overline{S} \cap \overline{B'} \cap \overline{B} = (R \cup B) \cap \overline{S} \cap \overline{B'} \cap \overline{B}$. Notice that since $v \in R \setminus S$ and $\lambda = \lambda'' - 2$ we have $|R_v \cap (B \cup B')| \leq \gamma_{\lambda''-3}n$. Combining these with $|S| \leq 4\Delta_{\lambda''}\gamma_{\lambda''}^{-1}$ and $|V(D) \setminus (R \cup B)| \leq \delta n/3$ we get

$$\begin{aligned} |R_v \setminus C| &= |R_v \setminus ((R \cup B) \cap \overline{S} \cap \overline{B'} \cap \overline{B})| \\ &= |R_v \cap ((\overline{R \cup B}) \cup S \cup B' \cup B)| \\ &\leq |R_v \cap (\overline{R \cup B})| + |R_v \cap S| + |R_v \cap (B \cup B')| \\ &\leq |V(D) \setminus (R \cup B)| + |S| + |R_v \cap (B \cup B')| \\ &\leq \delta n/3 + 4\Delta_{\lambda''}\gamma_{\lambda''}^{-1} + \gamma_{\lambda''-3}n \leq \delta n/2. \end{aligned}$$

The last inequality comes from $n \geq N_0$ and $\lambda \leq \lambda_{\max}$.

Next we show that $|C \setminus R_v| \leq \delta n/2$. Using $v \in R \setminus S$ and the monotonicity of bypassing we get that v $(\lambda - 3)$ -bypasses B' . By Lemma 4.10, v $(\lambda - 4)$ -bypasses $\overline{R_v} \cup B'$. Now using the monotonicity of bypassing we have a vertex $v \in B'$ and a set $\overline{R_v} \cup B' \supseteq B'$ such that v $(\lambda' - 43\delta^{-1})$ -bypasses $\overline{R_v} \cup B'$. From Claim 4.19 we have $|\overline{R_v} \cup B'| \leq |B'| + \delta n/2$ which implies $|\overline{R_v} \setminus B'| \leq \delta n/2$. This gives us

$$|C \setminus R_v| = |C \cap \overline{R_v}| \leq |\overline{R_v} \setminus B'| + |C \cap B'| = |\overline{R_v} \setminus B'| \leq \delta n/2.$$

□

We have now proved that C satisfies all the requirements of the lemma aside from “ C is nonempty”. Thus, for the remainder of the proof we can assume that C is empty, or equivalently $R \subseteq S \cup B' \cup B$. We'll show that $\{v'\}$ is a λ -component satisfying the conditions of the lemma. Notice that $\{v'\} \subseteq \overline{B_0}$ holds as a consequence of the definition of v' in Claim 4.19.

By $|S| \leq 4\Delta_{\lambda'}\gamma_{\lambda'}^{-1} \leq \gamma_{\lambda+1}n$ and Observation 4.7 (ii), v $(\lambda + 1)$ -bypasses S . By Lemma 4.10, v' λ -bypasses $S \cup B' \cup B$. From $R \subseteq S \cup B' \cup B$ we obtain $|V(D) \setminus (S \cup B' \cup B)| = |V(D) \setminus (R \cup S \cup B' \cup B)| \leq |V(D) \setminus (R \cup B)| \leq \delta n/3$. By Lemma 4.16 applied with $v = v'$, $\lambda = \lambda$, and $\hat{B} = S \cup B' \cup B$, we have that $C' = \{v'\}$ is a λ -component which satisfies the conditions of the lemma. □

We can iteratively apply Lemma 4.18 to find a sequence of λ -components for decreasing λ . We'll use this sequence in the proof of Theorem 4.1.

Lemma 4.22. *Fix $m = 4\delta^{-1}$ and let D be a labelled digraph on $n \geq N_0$ vertices. For any $v \in V(D)$ we can choose C_1, \dots, C_m and $\lambda_1, \dots, \lambda_m$ such that for each i , C_i is a λ_i -component with $|C_i \setminus C_{i+1}| \leq \delta n$ and $1 \leq \lambda_{i+1} - 88\delta^{-2} \leq \lambda_i \leq \lambda_{i+1} - 9 \leq \lambda_{\max}$ for $i = 1, \dots, m$.*

In addition there is a length $\leq d_{\lambda_{\max}/2}$ switching path which starts at v and passes through all of C_1, \dots, C_m .

Proof. We will choose vertices v_1, \dots, v_{m+1} , paths P_1, \dots, P_{m+1} , sets R_1, \dots, R_{m+1} , numbers $\lambda_1, \dots, \lambda_{m+1}$, and components C_1, \dots, C_m . They will have the following properties.

- (i) For $i \leq m$, $1 \leq \lambda_{i+1} - 88\delta^{-2} \leq \lambda_i \leq \lambda_{i+1} - 9 \leq \lambda_{\max}/4$.
- (ii) For $i \leq m + 1$, v_i λ_i -reaches R_i and v_i $(\lambda_i - 3)$ -bypasses $\overline{R_i}$.

- (iii) For $i \leq m + 1$, P_i is a switching path from v to v_i passing through v_{i+1}, \dots, v_m . Also $|P_i| \leq d_{\lambda_{i+1}} + d_{\lambda_{i+2}} + \dots + d_{\lambda_m}$.
- (iv) For $i \leq m$, C_i is a λ_i -component with $C_i \subseteq R_{i+1}$, $v_i \in C_i$, and $|C_i \triangle R_i| \leq \delta n$.

Once we have constructed these sequences, then it is easy to see that the components C_1, \dots, C_m , the numbers $\lambda_1, \dots, \lambda_m$, and the path P_1 satisfy the conditions of the lemma. Indeed C_i is a λ_i -component with $|C_i \setminus C_{i+1}| \leq |R_{i+1} \setminus C_{i+1}| \leq |R_{i+1} \triangle C_{i+1}| \leq \delta n$ by (iv). By (i), (iii), and (iv) the path P_1 is a length $\leq md_{\lambda_m} \leq md_{(\lambda_{\max}/4)} \leq d_{\lambda_{\max}/2}$ switching path starting from v which passes through the vertex $v_i \in C_i$ for $i = 1, \dots, m$.

We will construct v_i, P_i, R_i, λ_i , and C_i in reverse order starting with $i = m+1$ and ending with $i = 1$. Let $v_{m+1} = v$, $P_{m+1} = \{v\}$, and $\lambda_{m+1} = \lambda_{\max}/4$. Use Lemma 4.13 to find a set R_{m+1} such that v_{m+1} (λ_{m+1})-reaches R_{m+1} and ($\lambda_{m+1} - 3$)-bypasses $\overline{R_{m+1}}$. By construction, conditions (i) – (iii) are satisfied for $v_{m+1}, P_{m+1}, R_{m+1}$, and λ_{m+1} . Condition (iv) doesn't need to be checked since we do not have a component C_{m+1} .

For each $i \leq m + 1$, suppose that we have constructed v_i, P_i, R_i , and λ_i . We build $v_{i-1}, P_{i-1}, R_{i-1}, \lambda_{i-1}$, and C_{i-1} as follows.

By the monotonicity of bypassing and Observation 4.8, v_i λ_i -reaches $R_i \cup \{v_i\}$ and ($\lambda_i - 3$)-bypasses $\overline{R_i \cup \{v_i\}}$. Since v_i λ_i -reaches $R_i \cup \{v_i\}$ and $|P_i| \leq 3id_{\lambda_{m+1}} \leq k_{\lambda_i}$, there is a subset $R'_i \subseteq R_i \cup \{v_i\}$ of order at least $|R_i| - \Delta_{\lambda_i}$ such that there are length $\leq d_{\lambda_i}$ switching paths from v_i to all $r \in R'_i$ avoiding $\underline{P_i}$. Without loss of generality we can assume $v_i \in R'_i$ (since there is a length 0 switching path from v_i to v_i avoiding any set of labels.) By $\Delta_{\lambda_i} \leq \gamma_{\lambda_i-3}n$, Observation 4.7 (ii), and Lemma 4.10, v_i ($\lambda_i - 4$)-bypasses $\overline{R'_i}$. Apply Lemma 4.18 with $v_0 = v_i$, $\lambda_0 = \lambda_i - 9$, and $B_0 = \overline{R'_i}$ in order to find a nonempty λ -component C_{i-1} contained in R'_i for some $\lambda_i - 9 - 87\delta^{-2} \leq \lambda \leq \lambda_i - 9$. Let $\lambda_{i-1} = \lambda$. Let v_{i-1} be any vertex in C_{i-1} . Since $v_{i-1} \in C_{i-1} \subseteq R'_i$, there is a length $\leq d_{\lambda_i}$ switching path Q from v_i to v_{i-1} which avoids $\underline{P_i}$. Let $P_{i-1} = P_i + Q$. Let R_{i-1} be the set from the definition of “ λ_{i-1} -component” such that $|R_{i-1} \triangle C_{i-1}| \leq \delta n$, v_{i-1} λ_{i-1} -reaches R_{i-1} , and ($\lambda_{i-1} - 3$)-bypasses $\overline{R_{i-1}}$.

We claim that $v_{i-1}, P_{i-1}, R_{i-1}, \lambda_{i-1}$, and C_{i-1} satisfy properties (i) – (iv). For property (i) we have $\lambda_i - 88\delta^{-2} \leq \lambda_i - 9 - 87\delta^{-2} \leq \lambda_{i-1} \leq \lambda_i - 9$ and also $\lambda_{i-1} \geq \lambda_{m+1} - (m-i+1)88\delta^{-2} \geq \lambda_{\max} - 252\delta^{-3} \geq 1$. Property (ii) holds since we chose R_{i-1} so that v_{i-1} λ_{i-1} -reaches R_{i-1} and ($\lambda_{i-1} - 3$)-bypasses $\overline{R_{i-1}}$. To

see that property (iii) holds, notice that it held for P_i , and that Q is a length $\leq d_{\lambda_{i-1}}$ switching path avoiding \underline{P}_i from v_i to v_{i-1} . Using Lemma 4.2 we have that $P_{i-1} = P_i + Q$ is a length $e(P_i) + e(Q) \leq d_{\lambda_{i-1}} + \dots + d_{\lambda_m}$ switching path from v to v_{i-1} . For property (iv), C_{i-1} being a λ_{i-1} -component with $C_{i-1} \subseteq R_i$ comes from our application of Lemma 4.18, $v_{i-1} \in C_{i-1}$ comes from our choice of v_{i-1} , and $|C_{i-1} \triangle R_{i-1}| \leq \delta n$ comes from the choice of R_{i-1} . \square

4.3. Growth of λ -components

Notice that so far in Section 4, “amidstness” has only come up in the statement of Theorem 4.1. In this section we build a link between amidstness and λ -components. First we will need a more precise notion of amidstness which incorporates the parameter λ .

Definition 4.23. *Let u and v be two vertices in an edge-labelled, directed graph D , c a label, and S a set of labels. We say that c is (λ, S) -amidst u and v if $c \notin S$ and there is a length $\leq 2d_\lambda$ switching path $P = (u, p_1, \dots, p_d, v)$ avoiding S from u to v such that the following hold.*

- (i) *There are no edges of P labelled by c .*
- (ii) *If c is a vertex of D then c is in $\{p_1, \dots, p_d, v\}$.*

The following is an extension of Lemma 4.2. It shows that when we concatenate two switching paths, then the vertex at which we concatenate automatically becomes amidst the endpoints of the concatenated path.

Lemma 4.24. *In a labelled digraph D , let S be a set of vertices, P a length ≥ 1 and $\leq d_\lambda$ switching path from u to x which avoids S , and Q a length $\leq d_\lambda$ switching path from x to v which avoids S and \underline{P} . Then $P+Q$ witnesses x being (λ, S) -amidst u and v .*

Proof. Using Lemma 4.2 we have that $P+Q$ is a length $\leq 2d_\lambda$ switching path avoiding S from u to v which passes through x . We have $x \notin S$ since P avoids S , P ends with x , and $|P| \geq 2$. Part (ii) of the definition of (λ, S) -amidst holds since $x \in P + Q$ and x is not the starting vertex of P . It remains to show that part (i) of the definition of (λ, S) -amidst holds. Note that P has no edges labelled by x since P is a switching path ending with x , and Q has no edges labelled by x since Q avoids $\underline{P} \ni x$. These imply that $P + Q$ has no edges labelled by x . \square

Recall that the idea behind “amidstness” was to identify labels which can be used to extend switching paths. The next lemma makes this precise.

Lemma 4.25. *Let x be a label in a labelled digraph D which is (λ, S) -amidst v and y . Suppose that we have a vertex z with $z \notin S$ and the edge yz present and labelled by x . Then there is a length $\leq 2d_\lambda + 1$ switching path P from v to z avoiding S .*

Proof. Since x is (λ, S) -amidst u and v , there is a length $\leq 2d_\lambda$ switching path Q from v to y avoiding S and having no edges labelled by x . In addition if x is a vertex then $x \in V(Q) \setminus \{v\}$. If $z \in Q$, then we are done by choosing P to be the subpath of Q ending with z . Otherwise we take $P = Q + z$ to get a path from v to z . To see that this is a switching path first notice that the label of the last edge yz is x which is not present on the edges of P . In addition if x is a vertex then $x \in V(Q) \setminus \{v\} = V(P) \setminus \{v, z\}$. Thus P is a switching path. Notice that $x \notin S$ since x is (λ, S) -amidst v and y . Combining this with $z \notin S$ and the fact that Q avoided S shows that P avoids S . Thus P is a length $\leq 2d_\lambda + 1$ switching path from v to z avoiding S as required. \square

Recall that for a label x and a vertex v , $N_{\{x\}}^+(v)$ denotes the set of vertices y with vy an edge labelled by x . In a out-properly labelled digraph we always have $|N_{\{x\}}^+(v)| \in \{0, 1\}$. For a labelled digraph D , let

$$L^+(D) = \{(x, y) : x \text{ is a label and } y \in V(D) \text{ with } N_{\{x\}}^+(y) \neq \emptyset\}.$$

Equivalently we have that $L^+(D)$ is the set of pairs (x, y) with x a label and $y \in V(D)$ such that there is an edge in D starting at y labelled by x .

The following lemma gives a connection between the notions of amidstness and λ -components. It shows that for a λ -component C , most triples of the form $(v, x, y) \subseteq C \times (X_0 \cup C) \times C$ have x amidst v and y .

Lemma 4.26. *Let D be an out-properly labelled digraph with $|D| = n \geq N_0$, $\lambda \in \mathbb{N}$ with $3 \leq \lambda \leq \lambda_{\max}$, X_0 the set of non-vertex labels of D , and C a λ -component. Fix a vertex $v \in C$. To every set of labels S with $|S| \leq k_\lambda - 3d_\lambda$ we can assign sets $C_S \subseteq C$ and $Z_S \subseteq (X_0 \cup C_S) \times C_S$ with the following properties.*

(i) *For every $(x, y) \in Z_S$, the label x is (λ, S) -amidst v and y .*

(ii) $|C_S| \geq |C| - 2\delta n$ and $|Z_S| \geq |(X_0 \cup C_S) \times C_S| - 4\delta n|C_S|$.

(iii) For two sets S and T we have $|L^+(D) \cap (Z_S \setminus Z_T)| \leq 6\Delta_\lambda n$.

Proof. Since C is a λ -component, for every $u \in C$ there is a set $R^u \subseteq C$ with $|R^u| \geq |C| - \delta n$ so that u λ -reaches R^u . For a set of labels S , let R_S^u be the set of $y \in R^u$ for which there is a length $\leq d_\lambda$ switching path from u to y avoiding S . For each u , S and $y \in R_S^u$, fix such a switching path $P_S^{u,y}$. Since u λ -reaches R^u , we have

$$|R^u \setminus R_S^u| \leq \Delta_\lambda \text{ for all } S \text{ with } |S| \leq k_\lambda. \quad (3)$$

For a pair of sets S and T with $|S|, |T| \leq k_\lambda$, we trivially have $|R_S^u \setminus R_T^u| \leq |R^u \setminus R_T^u|$ which implies $|R_S^u \setminus R_T^u| \leq \Delta_\lambda$. We now define the sets C_S and Z_S .

$$\begin{aligned} C_S &= R_S^v \setminus \{v\}, \\ Z_S^0 &= \{(x, y) \in X_0 \times C_S : P_S^{v,y} \text{ avoids } x \text{ and } x \notin S\}, \\ Z_S^1 &= \{(x, y) \in C_S \times C_S : y \in R_{S \cup \underline{P_S^{v,x}}}^x\}, \\ Z_S &= Z_S^0 \cup Z_S^1. \end{aligned}$$

Notice that we have $Z_S \subseteq (X_0 \cup C_S) \times C_S$. For S with $|S| \leq k_\lambda - 3d_\lambda$, (3) implies that $|C_S| \geq |R^v| - \Delta_\lambda - 1 \geq |C| - 2\delta n$ as required by part (ii) of the lemma. The next two claims prove “ $|Z_S| \geq |(X_0 \cup C_S) \times C_S| - 4\delta n|C_S|$ ”, completing the proof of part (ii).

Claim 4.27. For any S with $|S| \leq k_\lambda - 3d_\lambda$ we have $|(X_0 \times C_S) \setminus Z_S^0| \leq k_\lambda |C_S|$.

Proof. We have $(X_0 \times C_S) \setminus Z_S^0 \subseteq (S \times C_S) \cup \bigcup_{y \in C_S} \left(\underline{P_S^{v,y}} \times \{y\} \right)$. Using $|S| \leq k_\lambda - 3d_\lambda$ and $|\underline{P_S^{v,y}}| \leq 3d_\lambda$, this implies $|(X_0 \times C_S) \setminus Z_S^0| \leq |S||C_S| + \sum_{y \in C_S} |\underline{P_S^{v,y}}| \leq k_\lambda |C_S|$. \square

Claim 4.28. For any S with $|S| \leq k_\lambda - 3d_\lambda$ we have $|(C_S \times C_S) \setminus Z_S^1| \leq 2\delta n|C_S|$.

Proof. Notice that we have

$$|Z_S^1| = \sum_{x \in C_S} |R_{S \cup \underline{P_S^{v,x}}}^x \cap C_S| = \sum_{x \in C_S} (|C_S| - |C_S \setminus R_{S \cup \underline{P_S^{v,x}}}^x|). \quad (4)$$

We will bound the second term by the following

$$|C_S \setminus R_{S \cup \underline{P_S^v, x}}^x| \leq |C \setminus R_{S \cup \underline{P_S^v, x}}^x| \leq |C \setminus R^x| + |R^x \setminus R_{S \cup \underline{P_S^v, x}}^x| \leq \delta n + \Delta_\lambda \leq 2\delta n.$$

The second last inequality comes from $|C \setminus R^x| \leq \delta n$, $|S \cup \underline{P_S^v, x}| \leq k_\lambda$, and (3). Plugging the above into (4) we get $|Z_S^1| \geq \sum_{x \in C_S} (|C_S| - 2\delta n) = |C_S|^2 - 2\delta n|C_S|$ as required. \square

From Claims 4.27 and 4.28, $n \geq N_0$, and $\lambda \leq \lambda_{\max}$ we get $|Z_S| \geq |(X_0 \cup C_S) \times C_S| - k_\lambda|C_S| - 2\delta n|C_S| \geq |(X_0 \cup C_S) \times C_S| - 4\delta n|C_S|$, completing the proof of part (ii) of the lemma.

Next we prove part (i) of the lemma.

Claim 4.29. *Let S be a set of $\leq k_\lambda - 3d_\lambda$ labels and $(x, y) \in Z_S$. Then the label x is (λ, S) -amidst v and y .*

Proof. Suppose $x \in X_0$, or equivalently $(x, y) \in Z_S^0$. By definition of Z_S^0 , we have $x \notin S$. The path $P_S^{v, y}$ is a length $\leq d_\lambda$ switching path from v to y avoiding S . Since $(x, y) \in Z_S^0$, $P_S^{v, y}$ also avoids x , and so witnesses x being (λ, S) -amidst v and y .

Suppose $x \in C_S$, or equivalently $(x, y) \in Z_S^1$. Since $x \in C_S \subseteq R_S^v$, recall that we have a length $\leq d_\lambda$ switching path $P_S^{v, x}$ from v to x avoiding S . Since $v \notin C_S$, we have $x \neq v$, which implies that $P_S^{v, x}$ has length ≥ 1 . Since $y \in R_{S \cup \underline{P_S^v, x}}^x$, we have a length $\leq d_\lambda$ switching path $P_{S \cup \underline{P_S^v, x}}^{x, y}$ from x to y avoiding S and $\underline{P_S^v, x}$. By Lemma 4.24 we have that $Q = P_S^{v, x} + P_{S \cup \underline{P_S^v, x}}^{x, y}$ is a switching path witnessing x being (λ, S) amidst v and y . \square

The following two claims prove part (iii) of the lemma.

Claim 4.30. *For S and T with $|S|, |T| \leq k_\lambda - 3d_\lambda$ we have $|L^+(D) \cap (Z_S^0 \setminus Z_T^0)| \leq 2\Delta_\lambda n$.*

Proof. We proceed as follows.

$$\begin{aligned}
|(L^+(D) \cap (Z_S^0 \setminus Z_T^0))| &\leq |L^+(D) \cap (X_0 \times C_S \setminus Z_T^0)| \\
&\leq |L^+(D) \cap (X_0 \times C_S \setminus X_0 \times C_T)| \\
&\quad + |L^+(D) \cap (X_0 \times C_T \setminus Z_T^0)| \\
&\leq |L^+(D) \cap (X_0 \times C_S \setminus X_0 \times C_T)| + k_\lambda |C_S| \\
&= |L^+(D) \cap (X_0 \times (C_S \setminus C_T))| + k_\lambda |C_S| \\
&\leq n |C_S \setminus C_T| + k_\lambda |C_S| \\
&\leq \Delta_\lambda n + k_\lambda |C_S| \\
&\leq 2\Delta_\lambda n.
\end{aligned}$$

The first inequality comes from $Z_S^0 \subseteq X_0 \times C_S$. The second inequality is an instance of $U \setminus W \subseteq (U \setminus V) \cup (V \setminus W)$. The third inequality comes from Claim 4.27. The equality is an instance of $U \times V \setminus U \times W = U \times (V \setminus W)$. The fourth inequality comes from $|L^+(D) \cap ((V(D) \cup X_0) \times U)| = \sum_{u \in U} |N^+(u)| \leq n|U|$ which holds for any set of vertices U . The fifth inequality comes from $|R_S^u \setminus R_T^u| \leq \Delta_\lambda$. The sixth inequality holds since $k_\lambda \leq \Delta_\lambda$ for all $\lambda \geq 3$. \square

Claim 4.31. *For S and T with $|S|, |T| \leq k_\lambda - 3d_\lambda$ we have $|Z_S^1 \setminus Z_T^1| \leq 4\Delta_\lambda |C_S|$.*

Proof. Using the definitions of Z_S and C_S we have the following.

$$\begin{aligned}
|Z_S^1 \setminus Z_T^1| &= |((C_S \setminus C_T) \times C_S) \cap (Z_S^1 \setminus Z_T^1)| + |((C_S \cap C_T) \times C_S) \cap (Z_S^1 \setminus Z_T^1)| \\
&= \sum_{x \in C_S \setminus C_T} |C_S \cap R_{S \cup \underline{P_S^{v,x}}}^x| + \sum_{x \in C_S \cap C_T} |(C_S \cap R_{S \cup \underline{P_S^{v,x}}}^x) \setminus (C_T \cap R_{T \cup \underline{P_T^{v,x}}}^x)| \\
&\leq \sum_{x \in C_S \setminus C_T} |C_S| + \sum_{x \in C_S \cap C_T} |C_S \setminus C_T| + \sum_{x \in C_S \cap C_T} |R_{S \cup \underline{P_S^{v,x}}}^x \setminus R_{T \cup \underline{P_T^{v,x}}}^x| \\
&\leq 4\Delta_\lambda |C_S|.
\end{aligned}$$

The first equality comes from $Z_S^1 \setminus Z_T^1 \subseteq C_S \times C_S$. The second equality comes from $Z_S^1 = \bigcup_{x \in C_S} \{x\} \times (C_S \cap R_{S \cup \underline{P_S^{v,x}}}^x)$ and $Z_T^1 = \bigcup_{x \in C_T} \{x\} \times (C_T \cap R_{T \cup \underline{P_T^{v,x}}}^x)$. The first inequality comes from $(U \cap W) \setminus (U' \cap W') \subseteq (U \setminus U') \cup (W \setminus W')$. The second inequality comes from (3), $|S \cup \underline{P_S^{v,x}}| \leq k_\lambda$, and $|T \cup \underline{P_T^{v,x}}| \leq k_\lambda$. \square

Claims 4.30 and 4.31 imply (iii), concluding the proof of the lemma. \square

We now prove the main result of this section. The following lemma shows that if we have a digraph D which doesn't satisfy Theorem 4.1, then λ -components in D have a special property—every vertex in a λ -component C $(\lambda + 1)$ -reaches a set R which is much larger than C . This lemma will later be combined with Lemma 4.22 in order to show that all digraphs satisfy Theorem 4.1.

Lemma 4.32. *Let D be a out-properly labelled, digraph on $n \geq N_0$ vertices, X_0 the set of labels of D which are not vertices, and $3 \leq \lambda \leq \lambda_{\max}$. Let C be a λ -component in D , $u \in V(D)$, and P a length $\leq d_{\lambda_{\max}}$ switching path from u to some $v \in C$. Then one of the following holds.*

- (i) *There is a vertex $y \in C$, and a set of labels A amidst u and y such that $|N_A^+(y)| < |A| - |X_0| + \epsilon n$.*
- (ii) *v $(\lambda + 1)$ -reaches a set R with $|R| \geq |C| + (\epsilon - 7\delta)n$.*

Proof. Suppose that (i) doesn't hold.

Claim 4.33. *Let $y \in C$ and A be a set of labels which are (λ, \underline{P}) -amidst v and y . Then $|N_A^+(y)| \geq |A| - |X_0| + \epsilon n$.*

Proof. It is sufficient to show that every label in A is amidst u and y . Then the claim follows since we are assuming that (i) doesn't hold.

Fix $a \in A$. Using the definition of a being (λ, \underline{P}) -amidst v and y , there is a switching path Q from v to y avoiding \underline{P} and having no edges labelled by a . In addition if a is a vertex of D , then $a \in Q \setminus \{v\}$. We also have $a \notin \underline{P}$ since a is (λ, \underline{P}) -amidst v and y . Since Q avoids \underline{P} , Lemma 4.2 implies that $P + Q$ is a switching path from u to y . Since neither P nor Q had edges labelled by a , $P + Q$ also has no edges labelled by a . In addition, if a is a vertex, then $P + Q$ passes through a and $a \neq u$ (since $a \notin \underline{P}$.) Therefore $P + Q$ witnesses a being amidst u and y as required. \square

Notice that every $x \in X_0 \setminus \underline{P}$ is (λ, \underline{P}) -amidst v and v (witnessed by the single-vertex path v .) By Claim 4.33 applied with $A = X_0 \setminus \underline{P}$ we have $|N_{X_0 \setminus \underline{P}}^+(v)| \geq \epsilon n - |\underline{P}|$. By Lemma 4.12, $n \geq N_0$, and $\underline{P} \leq 3d_{\lambda_{\max}}$ we have

$$|C| \geq |N_{X_0}^+(v)| - \delta n - \gamma_{\lambda-3}n \geq \epsilon n - |\underline{P}| - \delta n - \gamma_{\lambda-3}n \geq \epsilon n/2. \quad (5)$$

Apply Lemma 4.26 to C and v to assign sets $C_S \subseteq C$ and $Z_S \subseteq (X_0 \cup C_S) \times C_S$ satisfying all the conclusions of Lemma 4.26 to every set of labels S with $|S| \leq k_\lambda - 3d_\lambda$. For $z \in V(D)$ and S a set of labels, let

$$E_S(z) = \{(x, y) \in L^+(D) \cap Z_S \text{ with } yz \text{ present and labelled by } x\}.$$

Since the labelling on D is out-proper, we have that $E_S(z) \cap E_S(z') = \emptyset$ for any S and $z \neq z'$. Also notice that for sets S and S' we have $E_S(z) \cap L^+(D) \cap Z_{S'} \subseteq E_{S'}(z)$. We now define the set R

$$R = \{z \in V(D) : |E_{\underline{P}}(z)| \geq \delta|C_{\underline{P}}|\}.$$

First we show that v $(\lambda + 1)$ -reaches R .

Claim 4.34. v $(\lambda + 1)$ -reaches R .

Proof. Let S be a set of at most $k_{\lambda+1}$ labels. Notice that $|S \cup \underline{P}| \leq k_{\lambda+1} + 3d_{\lambda_{\max}} \leq k_\lambda - 3d_\lambda$. Let $R_S = \{z \in R : E_{S \cup \underline{P}}(z) \neq \emptyset\}$.

We claim that for every $z \in R_S \setminus (S \cup \underline{P})$, there is a length $\leq d_{\lambda+1}$ switching path $P_{v,z}$ from v to z avoiding S . Notice that for every $z \in R_S \setminus (S \cup \underline{P})$, we have $E_{S \cup \underline{P}}(z) \neq \emptyset$, and so by the definition of $E_{S \cup \underline{P}}(z)$, there is a pair $(x, y) \in L^+(D) \cap Z_{S \cup \underline{P}}$ with yz present and labelled by x . Since $(x, y) \in Z_{S \cup \underline{P}}$, by part (i) of Lemma 4.26 we have x $(\lambda, S \cup \underline{P})$ -amidst v and y . By Lemma 4.25 applied with $S' = S \cup \underline{P}$, there is a switching path from v to z avoiding $S \cup \underline{P}$ of length $\leq 2d_\lambda + 1 \leq d_{\lambda+1}$.

To prove the claim it is sufficient to show that $|R \setminus (R_S \setminus (S \cup \underline{P}))| \leq \Delta_{\lambda+1}$. We will do this by showing $|R \setminus R_S| + |S \cup \underline{P}| \leq \Delta_{\lambda+1}$. We have

$$\begin{aligned} \delta|C_{\underline{P}}||R \setminus R_S| &\leq \sum_{z \in R \setminus R_S} |E_{\underline{P}}(z)| \\ &= \left| \bigcup_{z \in R \setminus R_S} E_{\underline{P}}(z) \right| \\ &\leq |(L^+(D) \cap Z_{\underline{P}}) \setminus (L^+(D) \cap Z_{S \cup \underline{P}})| \\ &\leq 6\Delta_\lambda n. \end{aligned}$$

The first inequality comes from the definition of R . The equality comes from $E_{\underline{P}}(z) \cap E_{\underline{P}}(z') = \emptyset$ for $z \neq z'$. For the second inequality first recall that $E_{\underline{P}} \subseteq L^+(D) \cap Z_{\underline{P}}$. Then notice that for $z \in R \setminus R_S$ we have $E_{S \cup \underline{P}}(z) = \emptyset$

and hence $E_{\underline{P}}(z) \cap L^+(D) \cap Z_{S \cup \underline{P}} \subseteq E_{S \cup \underline{P}}(z) = \emptyset$. This shows $E_{\underline{P}}(z) \subseteq (L^+(D) \cap Z_{\underline{P}}) \setminus (L^+(D) \cap Z_{S \cup \underline{P}})$ for $z \in R \setminus R_S$ which implies the second inequality. The third inequality comes from part (iii) of Lemma 4.26 and $|S \cup \underline{P}| \leq k_\lambda - 3d_\lambda$.

Rearranging and using $|S \cup P| \leq k_\lambda$ we obtain $|R \setminus R_S| + |S \cup P| \leq 6\Delta_\lambda n / \delta |C_{\underline{P}}| + k_\lambda$. From (5) and Lemma 4.26 (ii) we have $|C_P| \geq |C| - \delta n \geq \epsilon n / 3$. Combining these gives $|R \setminus R_S| + |S \cup P| \leq 18\Delta_\lambda / \delta \epsilon + k_\lambda \leq \Delta_{\lambda+1}$ as required. \square

The following claim lets us lower bound $|L^+(D) \cap Z_{\underline{P}}|$ in terms of $|Z_{\underline{P}}|$.

Claim 4.35. $|L^+(D) \cap Z_{\underline{P}}| \geq |Z_{\underline{P}}| - |X_0| |C_{\underline{P}}| + \epsilon n |C_{\underline{P}}|$.

Proof. For every $y \in C_{\underline{P}}$, define

$$\begin{aligned} A_{\underline{P}}(y) &= \{x \in X_0 \cup C_{\underline{P}} : (x, y) \in Z_{\underline{P}}\}, \\ A_{\underline{P}}^L(y) &= \{x \in X_0 \cup C_{\underline{P}} : (x, y) \in L^+(D) \cap Z_{\underline{P}}\}. \end{aligned}$$

From the definition of $L^+(D)$ we have $A_{\underline{P}}^L(y) = \{x \in A_{\underline{P}}(y) : |N_{\{x\}}^+(y)| = 1\}$. Notice that we have $Z_{\underline{P}} = \bigcup_{y \in C_{\underline{P}}} A_{\underline{P}}(y) \times \{y\}$ and $L^+(D) \cap Z_{\underline{P}} = \bigcup_{y \in C_{\underline{P}}} A_{\underline{P}}^L(y) \times \{y\}$. The following string of equalities holds.

$$|L^+(D) \cap Z_{\underline{P}}| = \sum_{y \in C_{\underline{P}}} |A_{\underline{P}}^L(y)| = \sum_{y \in C_{\underline{P}}} |N_{A_{\underline{P}}^L(y)}^+(y)| = \sum_{y \in C_{\underline{P}}} |N_{A_{\underline{P}}(y)}^+(y)|. \quad (6)$$

The first equality comes from $L^+(D) \cap Z_{\underline{P}} = \bigcup_{y \in C_{\underline{P}}} A_{\underline{P}}^L(y) \times \{y\}$. The second equality holds since $|N_{\{x\}}^+(y)| = 1$ for $x \in A_{\underline{P}}^L(y)$ which implies $|N_{A_{\underline{P}}^L(y)}^+(y)| = |A_{\underline{P}}^L(y)|$. The third equality holds because $|N_{A_{\underline{P}}(y)}^+(y)| = |N_{A_{\underline{P}}^L(y)}^+(y)| + |N_{A_{\underline{P}} \setminus A_{\underline{P}}^L(y)}^+(y)|$ and also $|N_{A_{\underline{P}} \setminus A_{\underline{P}}^L(y)}^+(y)| = 0$ which comes from “ $A_{\underline{P}}^L(y) = \{x \in A_{\underline{P}}(y) : |N_{\{x\}}^+(y)| = 1\}$ ”.

By Lemma 4.26 (i), for every $x \in A_{\underline{P}}(y)$ we have x (λ, \underline{P}) -amidst v and y . We can use Claim 4.33 to bound $\sum_{y \in C_{\underline{P}}} |N_{A_{\underline{P}}(y)}^+(y)|$.

$$\sum_{y \in C_{\underline{P}}} |N_{A_{\underline{P}}(y)}^+(y)| \geq \sum_{y \in C_{\underline{P}}} (|A_{\underline{P}}(y)| - |X_0| + \epsilon n) = |Z_{\underline{P}}| - |X_0| |C_{\underline{P}}| + \epsilon n |C_{\underline{P}}| \quad (7)$$

The first inequality comes from Claim 4.33. The equality comes from $Z_{\underline{P}} = \bigcup_{y \in C_{\underline{P}}} A_{\underline{P}}(y) \times \{y\}$. The claim follows from (6) and (7). \square

Now we show that R is large.

Claim 4.36. $|R| \geq |C| + (\epsilon - 7\delta)n$.

Proof. Since D doesn't have repeated edges, for any $z \in V(D)$ and $y \in C_{\underline{P}}$, there can be at most one label x for which $(x, y) \in E_{\underline{P}}(z)$. In particular this implies that for all $z \in V(D)$ we have $|E_{\underline{P}}(z)| \leq |C_{\underline{P}}|$. For $z \notin R$ we have $|E_{\underline{P}}(z)| \leq \delta|C_{\underline{P}}|$. These imply

$$(n - |R|)\delta|C_{\underline{P}}| + |R||C_{\underline{P}}| \geq \sum_{z \in V(D)} |E_{\underline{P}}(z)| = |L^+(D) \cap Z_{\underline{P}}|. \quad (8)$$

To see the equality, notice that by the definition of $E_{\underline{P}}(z)$ both sides equal $\left| \bigcup_{z \in V(D)} E_{\underline{P}}(z) \right|$. Combining Claim 4.35 with part (ii) of Lemma 4.26 we get

$$|L^+(D) \cap Z_{\underline{P}}| \geq |Z_{\underline{P}}| - |X_0||C_{\underline{P}}| + \epsilon n|C_{\underline{P}}| \geq |C_{\underline{P}}|^2 + \epsilon n|C_{\underline{P}}| - 4\delta n|C_{\underline{P}}|. \quad (9)$$

Combining (8) and (9) and rearranging implies the claim.

$$|R| \geq \frac{|C_{\underline{P}}|^2 + \epsilon n|C_{\underline{P}}| - 5\delta n|C_{\underline{P}}|}{|C_{\underline{P}}| - \delta|C_{\underline{P}}|} \geq |C_{\underline{P}}| + \epsilon n - 5\delta n. \geq |C| + \epsilon n - 7\delta n.$$

The last inequality is from Lemma 4.26 (ii). □

Claims 4.34 and 4.36 prove the lemma. □

4.4. Proofs of Theorems 1.3 and 4.1

In this section we prove the theorems of this paper.

Proof of Theorem 4.1. Suppose that there is a vertex u for which the theorem doesn't hold i.e. that for every vertex v and set of labels A amidst u and v we have $|N_A^+(v)| \geq |A| - |X_0| + \epsilon n$. Apply Lemma 4.22 to u in order to obtain numbers $\lambda_1, \dots, \lambda_{4\delta-1}$, components $C_1, \dots, C_{4\delta-1}$, and a path P passing through all of them. For each i , let v_i be a vertex in $P \cap C_i$.

Using Lemma 4.12 we can show that all the components C_i are large.

Claim 4.37. $|C_i| \geq \epsilon n/2$ for $i = 1, \dots, 4\delta^{-1}$.

Proof. Fix $i \leq 4\delta^{-1}$. Let P' be the subpath of P ending with v_i and $A = X_0 \setminus \underline{P}'$. Notice that P' witnesses every label in A being amidst u and v_i . Since we are assuming that the theorem doesn't hold we obtain $|N_A^+(v_i)| \geq |A| - |X_0| + \epsilon n \geq \epsilon n - |\underline{P}'| \geq \epsilon n - 3d_{\lambda_{\max}}$ (using the fact that P has length $\leq d_{\lambda_{\max}/2}$ in Lemma 4.22.) Since $N_A^+(v_i) \subseteq N_{X_0}^+(v_i)$, $n \geq N_0$, and $\lambda_i \leq \lambda_{\max}$, Lemma 4.12 implies $|C_i| \geq |N_A^+(v_i)| - \delta n - \gamma_{\lambda_i-3} n \geq \epsilon n - 3d_{\lambda_{\max}} - \delta n - \gamma_{\lambda_i-3} n \geq \epsilon n/2$. \square

Using Lemma 4.32, we can show that each C_i is much bigger than the previous one.

Claim 4.38. $|C_{i+1}| \geq |C_i| + \epsilon n/2$ for $i = 1, \dots, 4\delta^{-1}$

Proof. Recall that from Lemma 4.22 we have $|C_i \setminus C_{i+1}| \leq \delta n$ for all i . Combining this with Claim 4.37 we get $|C_i \cap C_{i+1}| = |C_i| - |C_i \setminus C_{i+1}| \geq \epsilon n/2 - \delta n \geq \epsilon n/3$. For each i , let R_{v_i} be a set from the definition λ_i -component which is λ_i -reached by v_i . Using $|C_i \Delta R_{v_i}| \leq \delta n$ and $|C_i \cap C_{i+1}| \geq \epsilon n/3$ we get $|R_{v_i} \cap C_i \cap C_{i+1}| \geq |C_i \cap C_{i+1}| - |C_i \setminus R_{v_i}| \geq \epsilon n/3 - \delta n \geq \Delta_{\lambda_i} + 1$. From the monotonicity of reaching we have that v_i λ_i -reaches $R_{v_i} \cap C_i \cap C_{i+1}$. Using $|\underline{P}| \leq 3d_{\lambda_{\max}/2} \leq k_{\lambda_i}$ and $|R_{v_i} \cap C_i \cap C_{i+1}| \geq \Delta_{\lambda_i} + 1$ this implies that there is a length $\leq d_{\lambda_i}$ switching path Q avoiding \underline{P} from v_i to a vertex $v \in R_{v_i} \cap C_i \cap C_{i+1}$. Let P' be the subpath of P from u to v_i . By Lemma 4.2, $P' + Q$ is a length $\leq d_{\lambda_{\max}}$ switching path from u to v .

Notice that all the conditions of Lemma 4.32 hold with $C = C_i$, $\lambda = \lambda_i$, $u = u$, $v = v$, and $P = P' + Q$. In addition, (i) cannot hold since we are assuming that u is a vertex for which the theorem is false. Therefore part (ii) of Lemma 4.32 occurs, i.e. we obtain a set R which is (λ_i+1) -reached by v and has $|R| \geq |C_i| + (\epsilon - 7\delta)n$. Since v is in the (λ_{i+1}) -component C_{i+1} , we have a set R_v with $|R_v \Delta C_{i+1}| \leq \delta n$ such that $\overline{R_v}$ is $(\lambda_{i+1} - 3)$ -bypassed by v . From Lemma 4.22 we have $\lambda_i + 1 \leq \lambda_{i+1} - 3$. Since v $(\lambda_i + 1)$ -reaches R , $(\lambda_{i+1} - 3)$ -bypasses $\overline{R_v}$, and $\lambda_i + 1 \leq \lambda_{i+1} - 3$ we get $|R \setminus R_v| = |R \cap \overline{R_v}| \leq \gamma_{(\lambda_{i+1}-3)} n$ (using the monotonicity of reaching.) Combining this with $|R_v \Delta C_{i+1}| \leq \delta n$, $|R| \geq |C_i| + (\epsilon - 7\delta)n$, and $n \geq N_0$ we get

$$\begin{aligned} |C_{i+1}| &\geq |R_v| - |R_v \setminus C_{i+1}| \geq |R| - |R \setminus R_v| - |R_v \Delta C_{i+1}| \\ &\geq |C_i| + (\epsilon - 7\delta)n - \gamma_{(\lambda_{i+1}-3)} n - \delta n \geq |C_i| + \epsilon n/2. \end{aligned}$$

\square

Iterating $|C_{i+1}| \geq |C_i| + \epsilon n/2$ for $i = 1, \dots, 4\delta^{-1}$ gives $|C_{4\delta^{-1}}| \geq 2\delta^{-1}\epsilon n > n$ which is a contradiction to $C_{4\delta^{-1}} \subseteq V(D)$. \square

Using Theorem 4.1 and Lemma 3.7 it is easy to prove our approximate version of Conjecture 1.2.

Proof of Theorem 1.3. Let $N_0 = N_0(0.9\epsilon)$ be the constant from Theorem 4.1, and let G be a properly coloured bipartite multigraph with $n \geq N_0$ colours having $\geq (1 + \epsilon)n$ edges of each colour. Let M be a maximum rainbow matching in G . Let X and Y be the parts of the bipartition of G and $X_0 = X \setminus V(M)$. Suppose for the sake of contradiction that M misses a colour c^* . Let $D_{G,M}$ be the labelled directed graph from Definition 3.1 corresponding to M . By Lemma 3.2, $D_{G,M}$ is out-properly labelled and simple. By Lemma 3.7 we have that for any vertex $v \in V(D_{G,M})$ and any set of labels A amidst c^* and v we have $|N_A^+(v)| \geq |A| - |X_0| + \epsilon n - 1 \geq |A| - |X_0| + 0.9\epsilon n$. This contradicts Theorem 4.1 applied with $\epsilon' = 0.9\epsilon$. \square

5. Concluding remarks

Here we make some remarks about the proof in this paper and directions for further research.

Improving the bound in Theorem 1.3

In this paper we proved an approximate version of the Aharoni-Berger Conjecture. Naturally, the main direction for further research is to improve the dependency of N_0 on ϵ in Theorem 1.3, and eventually prove the full conjecture. The dependency which follows from our proof is extremely bad— for $\epsilon > 0$, we have $N_0 = \text{twr}(2 \cdot 4^{\epsilon^{-9}})$. This dependency can surely be significantly improved by tweaking the proof in various ways. The author believes that getting a polynomial error term is out of reach of the methods in this paper.

Problem 5.1. *For some $\alpha < 1$ prove the following. Let G be a properly edge-coloured bipartite multigraph with n colours and at least $n + n^\alpha$ edges of each colour. Then G has a rainbow matching using every colour.*

Of particular interest would be to solve the above problem for some $\alpha < 1/2$. This is because there are some natural variants of the Aharoni-Berger Conjecture, where $n^{1/2}$ is the best currently known bound on the error term.

One of these is the version of the Aharoni-Berger Conjecture where not every colour needs to be used in the rainbow matching. Specifically it is known that in every properly edge-coloured bipartite multigraph with n colours and at least n edges of each colour, there is a rainbow matching of size $n - \sqrt{n}$ (see [13, 14].) Also recall that Haggkvist and Johansson proved an approximate version of the Aharoni-Berger Conjecture when the colour classes in G are all disjoint perfect matchings. In their paper [11] they say that “it will be clear from the proof that in order to reach $\epsilon \leq n^{-1/2}$ some new ideas must be found, if indeed the theorem is valid in this range.” This suggests that \sqrt{n} is a natural barrier for their techniques as well.

Finally it would be extremely interesting to show that every properly n -edge-coloured bipartite multigraph with $n + o(\log^2 n)$ edges of each colour has a rainbow matching using every colour. This would improve the best known bound on the Brualdi-Stein Conjecture [15].

Following this paper’s announcement, Gao, Ramadurai, Wanless, and Wormald [16] solved Problem 5.1 when an additional restriction is placed on the multigraph — that the multiplicity of each edge is at most $\sqrt{n}/\log^2 n$. Together with the Haggkvist-Johansson approach [11], and the author’s first approach [8], this gives three different approaches which can prove an asymptotic version of the Aharoni-Berger Conjecture when the underlying graph is simple. However, when the underlying graph is a general multigraph, then the approach in this paper remains the only known way to prove the Aharoni-Berger Conjecture asymptotically.

Improving the bound in Theorem 4.1

In this paper and [8] we introduced a directed graph based approach to the Aharoni-Berger Conjecture. It would be interesting to know how far this approach can be pushed. A specific open problem is to find out how small N_0 in Theorem 4.1 can be. Perhaps with some completely different proof technique, Theorem 4.1 can be proved with much better bounds? Of particular interest is to find out whether there are serious barriers to this approach proving the full Aharoni-Berger Conjecture. For example—for every $C \in \mathbb{N}$, are there labelled directed graphs such that for all $u, v \in V(D)$ and set of labels A amidst u and v we have $|N_A^+(v)| \geq |A| - |X_0| + C$?”

Non-bipartite graphs

The problem dealt with in this paper can be asked for non-bipartite graphs as well i.e. what is the smallest $f(n)$ so that given n matchings of size

$f(n)$ in a (not necessarily bipartite) graph G , there is a rainbow matching using every colour. The following conjecture about this appears in [16].

Conjecture 5.2 ([16]). *Let G be a properly edge-coloured bipartite multigraph with n colours having at least $n+2$ edges of each colour. Then G has a rainbow matching using every colour.*

The motivation for asking for $n+2$ edges of each colour (rather than $n+1$ like in Conjecture 1.2) is that in the non-bipartite case there is an example of graph with $n+1$ edges of each colour and no rainbow matching using each colour. This example is to take two vertex-disjoint copies of K_4 and properly edge-colour it using 3 colours. This graph has 4 edges of each colour, but it is easy to check that it has no rainbow matching of size 3.

Some progress has been made on Conjecture 5.2 since this paper has been announced. Keevash and Yepremyan [17] showed that if the multiplicities are $\leq o(n)$ and each colour appears at least $(1 + o(1))n$ times, then there is a rainbow matching using $n - O(1)$ colours. Gao, Ramadurai, Wanless, and Wormald [16] proved that if the multiplicities are $\leq \sqrt{n}/\log^2 n$ and each colour appears at least $(1 + o(1))n$ times, then there is a rainbow matching using every colour. When restricted to bipartite graphs, both of these results are qualitatively weaker than the one in this paper (since Theorem 1.3 places no restriction on the multiplicities).

It is natural to ask whether the methods in this paper can prove an approximate version Conjecture 5.2 without any restriction on the multiplicities. While the author isn't aware of any inherent barriers preventing a suitable generalization from existing, it certainly is no easy task to find one. Indeed the construction of the auxiliary directed graph $D_{G,M}$ relied heavily on the graph G being bipartite, and it is not clear what natural auxiliary directed graph could be useful in the non-bipartite case.

Different approaches based on directed graphs

Theorem 1.3 was proved by considering a directed graph $D_{G,M}$ corresponding to G and studying paths called “switching paths” in $D_{G,M}$. Neither the definition of $D_{G,M}$ nor the notion of “switching path” which we used are canonical. There are variations of these definitions which could be used to prove the same theorems. For example, instead of the directed graph $D_{G,M}$ perhaps one could consider an alternative definition where edges going through Y_0 in G somehow corresponded to edges in $D_{G,M}$. Instead of switching paths, one can use the following.

Definition 5.3 (Weakly switching path). *A path $P = (p_0, \dots, p_d)$ in an edge-labelled, directed graph D is a weakly switching path if the following hold.*

- *P is rainbow i.e. the edges of P have different labels.*
- *If $p_i p_{i+1}$ is labelled by a vertex $v \in V(D)$, then $v = p_j$ for some $1 \leq j \leq d$.*

The difference between “weakly switching path” and “switching path” is that for weakly switching paths if $p_i p_{i+1}$ is labelled by $v \in V(D)$ then we only ask for v to be a non-starting vertex of P (whereas for “switching path”, we wanted v to precede $p_i p_{i+1}$ as well.) It is not hard to check that everything in Section 4 stays true if we replace “switching path” by “weakly switching path”. Also a “weakly switching path” version of Theorem 4.1 follows from the version of Theorem 4.1 which we prove (just because switching paths are a special case of weakly switching paths.)

The above discussion suggests that weakly switching paths are perhaps a better notion to use in future research. The reason we didn’t use them in this paper is a little bit technical. If we changed “switching path” to “weakly switching path” in the definition of “ v λ -reaching R ”, then we would allow paths P from v to $x \in R$ which have an edge labelled by x . This causes a problem in the proof of Lemma 4.26 since there we want to construct paths from v through a vertex r with no edges labelled by x . However it is not hard to overcome this issue and prove a version of Theorem 4.1 whilst working directly with weakly switching paths (for example by suitably changing the definition of “reaching”.)

Acknowledgment

The author would like to thank Janos Barat for introducing him to this problem as well as Ron Aharoni, Dennis Clemens, Julia Ehrenmüller, Peter Keevash, and Tibor Szabó for various discussions related to it. Additionally, the author would like to thank a careful referee for many extremely helpful suggestions.

References

- [1] I. M. Wanless, Transversals in Latin squares: A survey, in: R. Chapman (Ed.), Surveys in Combinatorics 2011, Cambridge University Press, 2011.

- [2] H. Ryser, Neuere probleme der kombinatorik, Vorträge über Kombinatorik, Oberwolfach (1967) 69–91.
- [3] R. A. Brualdi, H. J. Ryser, Combinatorial matrix theory, Cambridge University Press, 1991.
- [4] S. K. Stein, Transversals of Latin squares and their generalizations, Pacific J. Math. 59 (1975) 567–575.
- [5] R. Aharoni, E. Berger, Rainbow matchings in r -partite r -graphs, Electron. J. Combin. 16.
- [6] R. Aharoni, P. Charbit, D. Howard, On a generalization of the Ryser-Brualdi-Stein conjecture, J. Graph Theory 78 (2015) 143–156.
- [7] D. Kotlar, R. Ziv, Large matchings in bipartite graphs have a rainbow matching, European J. Combin. 38 (2014) 97–101.
- [8] A. Pokrovskiy, Rainbow matchings and rainbow connectedness, The Electronic Journal of Combinatorics 24 (1).
- [9] D. Clemens, J. Ehrenmüller, An improved bound on the sizes of matchings guaranteeing a rainbow matching, Electron. J. Combin. 23.
- [10] R. Aharoni, D. Kotlar, R. Ziv, Representation of large matchings in bipartite graphs, SIAM Journal on Discrete Mathematics 31 (3) (2017) 1726–1731.
- [11] R. Häggkvist, A. Johansson, Orthogonal Latin rectangles, Combin. Probab. Comput. 17 (2008) 519–536.
- [12] B. Bollobás, Modern graph theory, Springer Science & Business Media, 2013.
- [13] J. Barát, A. Gyárfás, G. N. Sárközy, Rainbow matchings in bipartite multigraphs, Periodica Mathematica Hungarica 74 (1) (2017) 108–111.
- [14] 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 (1978) 235–237.
- [15] P. Hatami, P. W. Shor, A lower bound for the length of a partial transversal in a Latin square, J. Combin. Theory Ser. A 115 (2008) 1103–1113.

- [16] P. Gao, R. Ramadurai, I. Wanless, N. Wormald, Full rainbow matchings in graphs and hypergraphs, arXiv:1709.02665.
- [17] P. Keevash, L. Yepremyan, Rainbow matchings in properly-coloured multigraphs, arXiv:1710.03041.