

# On the number of symbols that forces a transversal

Peter Keevash\*

Liana Yepremyan†

May 29, 2018

## Abstract

Akbari and Alipour [1] conjectured that any Latin array of order  $n$  with at least  $n^2/2$  symbols contains a transversal. We confirm this conjecture for large  $n$ , and moreover, we show that  $n^{399/200}$  symbols suffice.

## 1 Introduction

A Latin square of order  $n$  is an  $n$  by  $n$  square with cells filled using  $n$  symbols so that every symbol appears once in each row and once in each column. A transversal in a Latin square is a set of cells using every row, column and symbol exactly once. The study of transversals in Latin squares goes back to Euler in 1776; his famous ‘36 officers problem’ is equivalent to showing that there is no Latin square of order 6 that can be decomposed into 6 transversals (this was finally solved by Tarry in 1900). An even more fundamental question is whether a Latin square always has a transversal. A quick answer is ‘no’, as shown by the addition table of  $\mathbb{Z}_{2k}$ , but it remains open whether there is always a transversal when  $n$  is odd (a conjecture of Ryser [10]) or whether there is always a partial transversal size  $n - 1$  (a conjecture of Brualdi [4] and Stein [11]). The best known positive result, due to Hatami and Shor [6], is that there is always a partial transversal size  $n - O(\log^2 n)$ .

Given the apparent difficulty of finding transversals in Latin squares, it is natural to ask if the problem becomes easier in Latin arrays with more symbols (now we fill a square with any number of symbols such that every symbol appears at most once in each row and at most once in each column). This problem was considered by Akbari and Alipour [1], who conjectured that any Latin array of order  $n$  with at least  $n^2/2$  symbols contains a transversal. Progress towards this conjecture was independently obtained by Best, Hendrey, Wanless, Wilson and Wood [3] (who showed that  $(2 - \sqrt{2})n^2$  symbols suffice) and Barát and Nagy [2] (who showed that  $3n^2/4$  symbols suffice).

We will henceforth adopt the standard graph theory translation of the problem, where we consider a Latin array of order  $n$  as a properly edge-coloured complete bipartite graph  $K_{n,n}$ , with one part corresponding to rows, the other part to columns, and the colour of an edge is the symbol in the corresponding cell of the array. In this language, the Akbari-Alipour conjecture is that if there are at least  $n^2/2$  colours then there is a rainbow perfect matching. Our main result confirms this conjecture for large  $n$  in a strong form.

**Theorem 1.1.** *Suppose the complete bipartite graph  $K_{n,n}$  is properly edge-coloured using  $dn^2$  colours, where  $n$  is sufficiently large and  $d > n^{-1/200}$ . Then there is a rainbow perfect matching.*

---

\*Mathematical Institute, University of Oxford, Oxford, UK. E-mail: keevash@maths.ox.ac.uk.

Research supported in part by ERC Consolidator Grant 647678.

†Mathematical Institute, University of Oxford, Oxford, UK. E-mail: yepremyan@maths.ox.ac.uk.

Here the constant ‘200’ could be somewhat improved, but we have sacrificed some optimisations for the sake of readability of our proof, as in any case the best bound we can obtain seems far from optimal (it might even be true that  $n^{1+o(1)}$  colours suffice!)

## 2 Proof

Here we give the proof of Theorem 1.1, assuming two lemmas that will be proved later in the paper. Consider the complete bipartite graph  $K_{n,n}$  with parts  $A$  and  $B$  both of size  $n$ , and a proper edge-colouring using at least  $dn^2$  colours, where  $n$  is sufficiently large and  $d > n^{-1/200}$ .

Let  $G$  be any subgraph of  $K_{n,n}$  obtained by including exactly one edge of each colour. Then  $e(G) \geq dn^2$ . We apply the following lemma, which will be proved in the next section, to find a pair  $(A_1, B_1)$  that satisfies Hall’s condition for a perfect matching ‘robustly’, so that it will still satisfy Hall’s condition after deleting small sets of vertices from each part. Note that as  $d > n^{-1/200}$  we obtain  $|A_1| = |B_1| > n^{0.7}/3$ .

**Lemma 2.1.** *There is  $G_1 = G[A_1, B_1]$  for some  $A_1 \subseteq A$  and  $B_1 \subseteq B$  with  $|A_1| = |B_1| \geq d^{60}n/3$  such that  $G_1$  has minimum degree at least  $10^{-3}d|A_1|$ , and for any  $S \subseteq A_1$  or  $S \subseteq B_1$  we have  $|N_{G_1}(S)| \geq \min\{2|S|, 2|A_1|/3\}$ .*

We define a random subgraph  $G^r$  of  $K_{n,n}$  of ‘reserved colours’ as follows. Choose each colour independently with probability  $p := n^{-0.32}$ . Let  $G^r$  consist of all edges of all chosen colours. By Chernoff bounds, whp  $|N_{G^r}(b) \cap A_1| = p|A_1| \pm (p|A_1|)^{2/3}$  and  $|N_{G^r}(b) \setminus A_1| = p|A \setminus A_1| \pm (p|A \setminus A_1|)^{2/3}$  for all  $b \in B$ , and similarly with  $A$  and  $B$  interchanged. Let  $G^* := (K_{n,n} \setminus G^r) \setminus (A_1 \cup B_1)$ . Then the minimum degree in  $G^*$  satisfies  $\delta(G^*) \geq (1-p)(n - |A_1|) - (pn)^{2/3}$ .

Let  $M_2$  be a maximum size rainbow matching in  $G^* := (K_{n,n} \setminus G^r) \setminus (A_1 \cup B_1)$ . Let  $A_2 = V(M_2) \cap A$  and  $B_2 = V(M_2) \cap B$ . By a result of Gyarfas and Sarkozy [5, Theorem 2] we have  $|A_2| = |B_2| \geq \delta(G^*) - 2\delta(G^*)^{2/3} \geq (1-2p)(n - |A_1|)$ , as  $pn = n^{0.68} \gg n^{2/3}$ .

Let  $G'_1$  be obtained from  $G_1$  by deleting any edges that use a colour used by  $M_2$  and restricting to some subsets  $A'_1 \subseteq A_1$  and  $B'_1 \subseteq B_1$  with  $|A'_1| = |B'_1| = (1 - 10^{-4}d)|A_1|$  so that  $G'_1$  has minimum degree at least  $(10^{-3} - 2 \cdot 10^{-4})d|A_1|$ . To see that this is possible, note that we delete at most  $n$  edges from  $G'_1$ , so each of  $A_1$  and  $B_1$  has at most  $10^4n/d|A_1| < 10^{-4}d|A_1|$  vertices at which we delete more than  $10^{-4}d|A_1|$  edges.

Let  $A_0 = A \setminus (A'_1 \cup A_2)$  and  $B_0 = B \setminus (B'_1 \cup B_2)$ . Note that  $|A_0| = |B_0| < 2 \cdot 10^{-4}d|A_1|$ , as  $pn = n^{0.68} \ll d|A_1|$ . The form of our intended rainbow matching is illustrated by the black and/or vertical edges in Figure 1 (the coloured diagonal edges illustrate the augmentation algorithm used in Section 4).

Let  $A'_0$  be the set of vertices  $a$  in  $A_0$  such that at least  $|B'_1|/2$  of the edges between  $a$  and  $B'_1$  have a colour used by  $M_2$ . Define  $B'_0$  similarly, interchanging  $A$  and  $B$ . We prove the following lemma in section 4.

**Lemma 2.2.** *Both  $A'_0$  and  $B'_0$  have size at most  $p|A_1|/4$ .*

Now we apply a greedy algorithm to construct a rainbow matching  $M_2 \cup M_0$  where each edge of  $M_0$  joins  $A_0 \cup B_0$  to  $A'_1 \cup B'_1$ . We start by choosing these edges for vertices in  $A'_0 \cup B'_0$  using colours in  $G^r$ . This is possible as the number of choices at each step is at least  $p|A_1| - (p|A_1|)^{2/3} > 3p|A_1|/4$ , and at most  $3 \cdot p|A_1|/4$  choices are forbidden due to using a colour or a vertex used at a previous step. Then we continue the greedy algorithm to choose these edges for the remainder of  $A_0 \cup B_0$ .

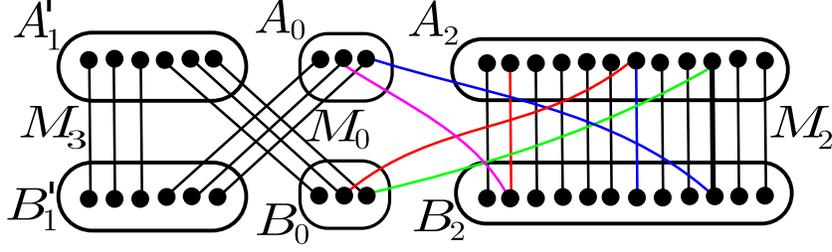


Figure 1: Proof by picture

This is possible as the number of choices at each step is at least  $|A'_1|/2$ , of which at most  $3|A_0|$  are forbidden due to using a colour or a vertex used at a previous step.

Finally, consider  $G_3 = G[A_3, B_3]$  obtained from  $G'_1$  by deleting all vertices covered by  $M_0$  and all edges that share a colour with  $M_0$ . Then  $|A_3| = |B_3| = |A'_1| - |A_0| > (1 - 10^{-3}d)|A_1|$  and  $G_3$  has minimum degree at least  $10^{-3}d|A_1| - |A_0| - |M_0| \geq 10^{-3}d|A_1|/2$ .

We claim that  $G_3$  has a perfect matching. To see this we check Hall's condition. Suppose for a contradiction there is  $S \subseteq A_3$  with  $|N(S)| < |S|$ . By the minimum degree we have  $|S| \geq 10^{-3}d|A_1|/2$ . Now we cannot have  $|S| \leq |A_1|/2$ , as then  $|N_{G_3}(S)| \geq \min\{2|S|, 2|A_1|/3\} - 2|A_0| > |S|$ . However, letting  $T = B_3 \setminus N(S)$  we have  $N(T) \subseteq A_3 \setminus S$ , so  $|N(T)| = |A_3| - |S| < |B_3| - |N(S)| = |T|$ . The same argument as for  $S$  gives  $|T| > |A_1|/2$ , contradiction. Therefore  $G_3$  has a perfect matching  $M_3$ . Now  $M_2 \cup M_0 \cup M_3$  is a rainbow perfect matching in  $G$ , which completes the proof of Theorem 1.1.

### 3 A robustly matchable pair

In this section we prove Lemma 2.1. Let  $G$  be a bipartite graph with parts  $A$  and  $B$ . We say  $G$  is  $(\varepsilon, \delta)$ -dense if for any  $A' \subseteq A$  and  $B' \subseteq B$  with  $|A'| \geq \varepsilon|A|$  and  $|B'| \geq \varepsilon|B|$  we have  $e_G(A', B') \geq \delta|A'||B'|$ . We start by applying the following result of Peng, Rödl and Ruciński [9, Theorem 1.3] with  $\varepsilon = 1/10$ ,  $c = 0.24$  and  $c' = 1/50$ .<sup>1</sup>

**Lemma 3.1.** *Suppose  $c, c' \in (0, 1)$  with  $4c + c' \leq 1$ . Then there are  $A' \subseteq A$  and  $B' \subseteq B$  with  $|A'| = |B'| \geq d^{2/\log_2(1+c\varepsilon)}n/2$  so that  $G' = G[A', B']$  is  $(\varepsilon, c'd)$ -dense.*

Let  $G_1 = G[A_1, B_1]$  be obtained from  $G'$  by the following algorithm. Initially,  $A_1 = A'$  and  $B_1 = B'$ . At any step of the algorithm, we update  $G_1$  by deleting a vertex or set of vertices of one of the following types (choosing arbitrarily if there is a choice).

- i.  $v \in A_1$  or  $v \in B_1$  with  $d_{G_1}(v) \leq 10^{-3}d|A'|$ ,
- ii.  $S \subseteq A_1$  or  $S \subseteq B_1$  with  $|S| < \varepsilon|A'|$  and  $|N_{G_1}(S)| \leq 2|S|$ .

Whenever we delete some vertices from  $A_1$  or  $B_1$  we delete an arbitrary set of the same size from the other, so that we always maintain  $|A_1| = |B_1|$ . We stop if no deletion is possible or if we have deleted at least  $2\varepsilon|A'|$  vertices from each side.

We claim that the latter option is impossible. Indeed, then without loss of generality we deleted  $\varepsilon|A'|$  vertices from  $A'$  of type (i) or (ii) as above (at least half of the deleted vertices are deleted for a reason other than maintaining equal part sizes). Let  $D^A = D_i^A \cup D_{ii}^A$  be the deleted vertices in  $A'$  according to deletions of type (i) or (ii). Note that  $|D^A| < 3\varepsilon|A'|$ , and  $|N_{G_1}(D_{ii}^A)| \leq 2|D_{ii}^A| < 6\varepsilon|A'|$ .

<sup>1</sup> This result follows from their proof; they state the case  $c = 1/8$ ,  $c' = 1/2$  and use  $\log_2(1 + \varepsilon/8) \geq \varepsilon/6$ .

Let  $B_0 = B_1 \setminus N_{G_1}(D_{ii}^A)$ , so  $|B_0| > (1 - 9\varepsilon)|A'| = \varepsilon|A'|$ . Now  $e_{G'}(D^A, B_0) \leq |D_i^A| \cdot 10^{-3}d|A'| < \frac{d}{50}|D^A||B_0|$  contradicts  $(\varepsilon, d/50)$ -density of  $G'$ , which proves the claim.

Thus the algorithm stops with  $|A_1| = |B_1| > (1 - 3\varepsilon)|A'| \geq d^{60}n/3$  (using  $2/\log_2(1.024) < 60$ ), minimum degree at least  $10^{-3}d|A'|$  and  $|N_{G_1}(S)| \geq 2|S|$  for any  $S \subseteq A_1$  or  $S \subseteq B_1$  with  $|S| < \varepsilon|A'|$ . Furthermore, for any  $S \subseteq A_1$  or  $S \subseteq B_1$  with  $|S| \geq \varepsilon|A'|$ , by  $(\varepsilon, d/50)$ -density of  $G'$  we have  $|N_{G_1}(S)| \geq |B_1| - \varepsilon|A'| \geq 2|A_1|/3$ . This proves Lemma 2.1.

## 4 Augmentation algorithm

In this section we prove Lemma 4.5, which will complete the proof of Theorem 1.1. Suppose it is not true, say  $|A'_0| > p|A_1|/4$ . We will iteratively construct  $R = R^A \cup R^B \subseteq M_2$ , where we think of  $R^A$  and  $R^B$  as ‘reachable’ from  $A_0$  and  $B_0$ . At some point  $R^A$  and  $R^B$  will intersect, which will contradict  $M_2$  being a maximum size rainbow matching in  $G^*$ . Let  $\theta := n^{-0.66}$ , and note that  $\theta|A_1| > n^{0.04}/3$ .

**Algorithm 4.1.** Let  $R^A = R^B = \emptyset$  and let  $C$  be the set of colours not used by  $M_2$ . At step  $i \geq 1$ :

- i. if  $R^A \cap R^B \neq \emptyset$  stop, otherwise let  $R_i^A$  be the set of all  $uv \in M_2$  where  $v \in B_2 \setminus V(R^B)$  such that at least  $\theta|A_1|$  edges in  $G^*$  from  $v$  to  $A_0$  use a colour in  $C$ , let  $C_i^A$  be the set of colours used by  $R_i^A$ , update  $R^A$  by adding  $R_i^A$  and  $C$  by adding  $C_i^A$ ,
- ii. if  $R^A \cap R^B \neq \emptyset$  stop, otherwise let  $R_i^B$  be the set of all  $uv \in M_2$  where  $u \in A_2 \setminus V(R^A)$  such that at least  $\theta|A_1|$  edges in  $G^*$  from  $u$  to  $B_0$  use a colour in  $C$ , let  $C_i^B$  be the set of colours used by  $R_i^B$ , update  $R^B$  by adding  $R_i^B$  and  $C$  by adding  $C_i^B$ .

**Claim 4.2.**  $|R_1^A| \geq |A_1|/4$ .

To see this, we consider the number  $X$  of edges in  $G^*$  with colour in  $C$  between  $A'_0$  and  $B_2$ . We have  $X \leq |R_1^A||A'_0| + |B_2|\theta|A_1|$  by definition of  $R_1^A$ . Also, by definition of  $A'_0$ , every vertex in  $A'_0$  has at least  $(1 - 2p)|B_2| - (|M_2| - |B'_1|/2) \geq |A_1|/3$  edges in  $G^*$  to  $B_2$  with colour in  $C$ , so  $X \geq |A'_0| \cdot |A_1|/3$ . As  $|A'_0| \geq p|A_1|/4$  and  $p^{-1}\theta n = n^{0.66} \ll pn \ll |A_1|$ , we deduce  $|R_1^A| \geq |A_1|/3 - |A'_0|^{-1}|B_2|\theta|A_1| \geq |A_1|/3 - 4p^{-1}\theta n \geq |A_1|/4$ , as claimed.

**Claim 4.3.** For  $i \geq 1$ , we have  $|R_i^B| \geq |R^A| - 3pn$  and  $|R_{i+1}^A| \geq |R^B| - 3pn$ .

To see this, we first note that as  $R^A \cap R^B = \emptyset$ , any vertex in  $B_0$  has at least  $(1 - 2p)|A_2| - |R^B| - |M_2 \setminus (R^A \cup R^B)| = |R^A| - 2p|A_2|$  edges in  $G^*$  to  $A_2 \setminus R^B$  with colour in  $C$ . Double-counting such edges as in the previous claim gives  $|B_0|(|R^A| - 2p|A_2|) \leq |R_i^B||B_0| + |B_2|\theta|A_1|$ , so  $|R_i^B| \geq |R^A| - 2p|A_2| - |B_0|^{-1}|B_2|\theta|A_1| \geq |R^A| - 3pn$ . The proof of the second inequality is similar, so the claim holds.

**Claim 4.4.** The algorithm terminates at some step  $i = i^+ < \log n$ .

To see this, we show inductively that if  $R^A \cap R^B = \emptyset$  at step  $i$  then  $|R_i^B| \geq f(i)|A_1|/3$  and  $|R_i^A| \geq (f(i) - 2^{-3})|A_1|/3$  where  $f(i) = 2^{i-4} + 2^{-1}$ . First note that  $f(1) = 5/8$  and for  $i \geq 2$  we have  $\sum_{j=1}^{i-1} (f(j) - 2^{-3}) = 2^{-4}(2^i - 1) + (i - 1)(2^{-1} - 2^{-3}) \geq f(i) - 3/16$ . At step 1 we have  $|R_1^A| \geq |A_1|/4 > (f(1) - 2^{-3})|A_1|/3$  and  $|R_1^B| \geq |R^A| - 3pn > 0.21|A_1| > f(1)|A_1|/3$ . Supposing the statement at step  $i - 1 \geq 1$ , we have  $|R_i^A| \geq (\sum_{j=1}^{i-1} |R_j^B|) - 3pn \geq (\sum_{j=1}^{i-1} f(j))|A_1|/3 - 3pn \geq (f(i) - 1/16)|A_1|/3 - 3pn \geq (f(i) - 2^{-3})|A_1|/3$  and  $|R_i^B| \geq (\sum_{j=1}^i |R_j^A|) - 3pn \geq (\sum_{j=1}^i f(j) - 2^{-3})|A_1|/3 - 3pn \geq (f(i + 1) - 3/16)|A_1|/3 - 3pn \geq f(i)|A_1|/3$ . Thus the required bounds hold by induction. While  $R^A \cap R^B = \emptyset$  we deduce  $(2f(i) - 2^{-3})|A_1|/3 < |M_2| < n$ , so  $i^+ < \log n$ , as claimed.

The algorithm terminates by finding some edge  $ab \in R^A \cap R^B$  where  $a \in A_2$  and  $b \in B_2$ . We will obtain a contradiction by modifying  $M_2$  to obtain a larger rainbow matching in  $G^*$ . Given two colours  $c$  and  $c'$  in  $C$ , we say that  $c$  is earlier than  $c'$  if  $c$  was added to  $C$  before  $c'$ . We start by applying the definition of  $R^A$  and  $R^B$  to find edges  $a_0b$  and  $ab_0$  of  $G^*$  with  $a_0 \in A_0$  and  $b_0 \in B_0$  where the colours of  $a_0b$  and  $ab_0$  are in  $C$  and earlier than that of  $ab$ . We modify  $M_2$  to obtain  $M'_2$  by deleting  $ab$  and adding  $a_0b$  and  $ab_0$ . Thus we obtain a larger matching, but  $M'_2$  may not be rainbow, due to repeating the colours of  $a_0b$  and  $ab_0$ . While the current matching  $M'_2$  is not rainbow, we apply the following ‘trace back’ algorithm (similar to that of [5]).

**Algorithm 4.5.** At step  $i \geq 1$  we have at most two ‘active’ edges, which are edges of  $M'_2$  having some colour in  $C$  shared with some edge that is still present from  $M_2$ . At step 1 these are  $a_0b$  and  $ab_0$ . If there is an active edge at step  $i$ , we choose one arbitrarily, call it  $a_i b_i$ , and let  $a'_i b'_i$  be the edge of  $M_2$  of the same colour  $c \in C$ . By construction of  $C$ , one of  $a'_i$  or  $b'_i$ , say  $a'_i$ , has at least  $\theta|A_1|$  edges to  $B_0$  or  $A_0$  using an earlier colour than  $c$  in  $C$ . We modify  $M'_2$  by deleting  $a'_i b'_i$  and adding some such edge  $a'_i b_0^i$  where  $b_0^i \in B_0$  is distinct from all previous choices. We say that  $a_i b_i$  is no longer active. We make  $a'_i b_0^i$  active if its colour is shared with some edge that is still present from  $M_2$ .

Algorithm 4.5 is illustrated in Figure 1: the thick black edge represents the edge  $ab \in R^A \cap R^B$ , at step 1 the green and blue diagonals are active, at step 2 the blue diagonal is active, at step 3 the red diagonal is active, at step 4 the pink diagonal is active, at step 5 there are no active edges so the algorithm terminates. To see that the algorithm succeeds, note that there are at most  $4 \log n$  steps of replacing an active edge by another, and each choice has at least  $\theta|A_1| > n^{0.04}/3 > 4 \log n$  options. Thus we obtain a rainbow matching  $M'_2$  in  $G^*$  with  $|M'_2| > |M_2|$ . This contradiction proves Lemma 4.5.

**Postscript.** The Akbari-Alipour conjecture was proved independently and simultaneously by Montgomery, Pokrovskiy and Sudakov [8]. Our proof is much simpler than theirs, and gives a better bound on the number of symbols required, whereas their proof applies in a much more general setting, and so has several further applications. Results similar to those in [8] (but not including the Akbari-Alipour conjecture) were independently and simultaneously obtained by Kim, Kühn, Kupavskii and Osthus [7].

## References

- [1] S. Akbari and A. Alipour, Transversals and multicolored matchings, *J. Combin. Designs* 12:325–332, 2004.
- [2] J. Barát and Z.L. Nagy, Transversals in generalized Latin squares, arXiv:1701.08220.
- [3] D. Best, K. Hendrey, I.M. Wanless, T.E. Wilson and D. Wood, Transversals in Latin arrays with many distinct symbols, arXiv:1612.09443.
- [4] R.A. Brualdi and H.J. Ryser, Combinatorial matrix theory, *Cambridge University Press*, 1991.
- [5] A. Gyárfás and G. Sárközy, Rainbow matchings and partial transversals of Latin squares, arXiv:1208.5670.
- [6] P. Hatami and P.W. Shor, A lower bound for the length of a partial transversal in a Latin square, *J. Combin. Theory Ser. A* 115:1103-1113, 2008.

- [7] J. Kim, D. Kühn, A. Kupavskii and D. Osthus, Rainbow structures in locally bounded colourings of graphs, arXiv:1805.08424.
- [8] R. Montgomery, A. Pokrovskiy and B. Sudakov, Decompositions into spanning rainbow structures, arXiv:1805.07564.
- [9] Y. Peng, V. Rödl and A. Ruciński, Holes in graphs, *Electron. J. Combin.* 9:R1, 2002.
- [10] H. Ryser, Neuere Probleme in der Kombinatorik, *Vorträge über Kombinatorik*, Oberwolfach, pp. 69-91, 1967.
- [11] S. K. Stein, Transversals of Latin squares and their generalizations, *Pacific J. Math.* 59:567-575, 1975.