Bachelor thesis

Unavoidable trees and forests in graphs

Georg Osang

July 19, 2012

Reviewers:	Prof. Dr. Maria Axenovich
	Prof. Dr. Dorothea Wagner
Advisors:	Prof. Dr. Maria Axenovich
	Dr. Ignaz Rutter

Faculty of Mathematics
Faculty of Computer Science

Karlsruhe Institute of Technology

Abstract

Classic results from extremal graph theory state that if certain graphs are made large enough, unavoidable substructures appear. Here we will cover this type of problem for specific graphs when these substructures are certain trees or forests. After giving a summary on related results, the following two extremal main problems are presented:

For a given family of same-order trees including the star and the path, how large can a tree be that contains none of them as subtree? We show that this value depends heavily on the number of spiders in the family: For a family of k-vertex trees consisting of p spiders and a constant number of non-spiders, we construct a tree of size $2^{\Theta(k \log^{p-1} k)}$ containing no trees from this family, and show that all asymptotically larger trees do contain some of them. Here $\log^{p-1}$ denotes the $(p-1)$-times-iterated logarithm.

For a balanced black-white-coloring of the complete bipartite graph $K_{n,n}$, we examine the size of the largest fork forest contained as subgraph. A fork is a path of length 2 consisting of a black and a white edge. It is shown that there are at least $(1-\frac{1}{\sqrt{2}})n$ vertex-disjoint forks all centered in the same partite set, and this is best possible, confirming a conjecture by Tverentina et al. An efficient algorithm finding the largest number is presented.

Zusammenfassung

Bekannte Resultate aus der extremalen Graphentheorie besagen, dass wenn bestimmte Graphen groß genug gemacht werden, unvermeidbare Strukturen auftreten. Hier werden wir dieses Problem für spezielle Graphen betrachten, wenn die Strukturen bestimmte Bäume oder Wälder sind. Nach einem Überblick über verwandte Resultate werden die folgenden zwei Hauptprobleme präsentiert:

Für eine gegebene Familie von Bäumen gleicher Größe, die Stern und Pfad beinhaltet, wie groß kann ein Baum sein, der keinen Baum aus der Familie als Unterbaum enthält? Wir zeigen, dass dieser Wert stark von der Anzahl der Spinnen in der Familie abhängt: Für eine Familie von k-Knoten–Bäumen, bestehend aus p Spinnen und einer konstanten Anzahl an Nicht–Spinnen, konstruieren wir einen Baum der Größe $2^{\Theta(k \log^{p-1} k)}$, der keinen Baum von der Familie enthält, und zeigen, dass alle asymptotisch größeren Bäume einen davon enthalten müssen. Hier ist $\log^{p-1}$ der $(p-1)$-Mal iterierte Logarithmus.

Für eine balancierte Schwarz–Weiß–Färbung des vollständigen bipartiten Graphen $K_{n,n}$ untersuchen wir die Größe des größten Gabelwaldes, der als Untergraph enthalten ist. Eine Gabel ist ein Pfad der Länge 2 bstehend aus einer schwarzen und einer weißen Kante. Es wird gezeigt, dass es mindestens $(1-\frac{1}{\sqrt{2}})n$ knoten-disjunkte Gabeln, die alle in der gleichen Hälfte des bipartiten Graphen zentriert sind, gibt, und dass dies bestmöglich ist, was eine Vermutung von Tverentina et al. bestätigt. Ein effizienter Algorithmus, der die größte Anzahl findet, wird vorgestellt.

Acknowledgements

I would like to express my deepest gratitude to my advisor from the faculty of mathematics, Prof. Maria Axenovich, for giving me perspective and guiding me into the intricacies of graph theory, for her advice on my research and interesting discussions, for being available for my inquiries even on short notice, her advice and assistance for my future endeavors, and her overall kindness. I would furthermore like to profusely thank my advisor from the faculty of computer science, Dr. Ignaz Rutter, for his support and advice on my research and for insightful discussions, and Prof. Dorothea Wagner for giving me the chance to write this thesis for computer science at her institute and agreeing to review my thesis. For our joint research work I would like to thank again Prof. Maria Axenovich, Dr. Ignaz Rutter and Dr. Marcus Krug, and I would also like to thank Olga Tverentina for making us aware of the problem about fork forests. Finally, I would like to say thanks to my colleagues from the graph theory seminar for their useful feedback on my presentation, especially Daniel Hoske for helping with proofreading.

Contents

1 Introduction

1.1 Definitions

The term *graph* always refers to an undirected graph without loops and double edges, i.e., a graph G is a pair $G = (V, E)$ consisting of a vertex set V and an edge set $E \subseteq \{\{v_1, v_2\} \colon v_1, v_2 \in V, v_1 \neq v_2\}$. For an edge $\{x, y\}$ we write xy for short. The number of vertices $|V|$ is called the *order* of the graph and also denoted as $|G|$, the number of edges $|E|$ is called the *size* of the graph and also denoted as $||G||$. We implicitly assume for a graph G that V is the vertex set, E the edge set, $n = |V|$ the order and $m = |E|$ the size of the graph, unless otherwise stated.

The *degree* of a vertex v, denoted $d(v)$, is the number of incident edges. If all vertices in a graph have the same degree d, then we call the graph d-regular. The average degree of a graph is the average over the degrees of all vertices.

We say a graph $G' = (V', E')$ is a subgraph of $G = (V, E)$ and write $G' \subseteq G$ if $V' \subseteq V$ and $E' \subseteq E$. We say that G' is an induced subgraph of G if in addition the following holds: $v_1v_2 \in E'$ if and only if $v_1, v_2 \in V'$ and $v_1v_2 \in E$.

An *edge coloring* $c \colon E \to C$, henceforth just *coloring* for short, using some set of colors C, assigns a color to each edge. If $|C| = k$, i.e., there are k colors available, we call the coloring a k-coloring. We also say for example red-blue coloring if a coloring uses only the colors red and blue. A graph is said to be monochromatic if all of its edges are assigned the same color.

A path is a graph consisting of a non-repeating finite sequence of vertices together with edges between any two consecutive vertices in the sequence. A cycle is a path where the first and last vertex are linked by an additional edge.

A graph is *connected* if between any two vertices there is a path connecting them. The maximal connected subgraphs of a graph are called connected components, or *components* for short. A forest is a graph that does not contain cycles. A tree is a connected forest.

The complete graph K_n on n vertices, also called a clique, is the graph that has an edge between any pair of vertices. A bipartite graph G is a graph whose vertex set can be partitioned into two disjoint subsets X and Y, $X \cup Y = V(G)$, called partite sets, such that there is no edge between any two vertices of X, and no edge between any two vertices of Y. For a bipartite graph G we implicitly assume the partite sets to be X and Y unless otherwise stated. The bipartite graph with partite sets of size r and s respectively that has an edge between any pair of vertices that are not in the same partite set is called the complete bipartite graph, and denoted as $K_{r,s}$.

More generally, a k-partite graph is a graph with k partite sets with no edges whose endpoints belong to the same partite set. The graph denoted by $K_{s_1,s_2,\ldots,s_k}$ is a complete k-partite graph with partite sets of size $s_1, \ldots, s_k$, and edges between any two vertices that are not in the same partite sets. We write $K_s^k = K_{s,s,\ldots,s}$ with k entries of s for the complete k-partite graph with partite sets of size s.

The path on n vertices is denoted by P_n, and $S_n := K_{1,n-1}$ is called a star.

A *matching* is a graph that consists of vertex disjoint copies of K_2, i.e., each component consists of two vertices linked by an edge. For a matching with s edges we write sK_2.

For all other standard graph theoretic definition, we refer the reader to [14, 48].

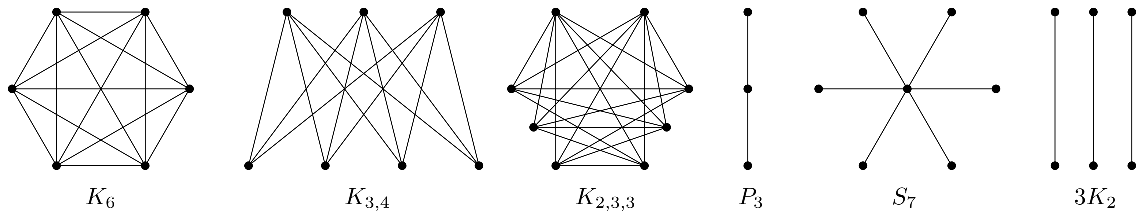


Figure 1: Various graphs.

1.2 Outline

In this thesis, we examine the topic of unavoidable trees and forests in graphs, and then proceed to present two new theorems from this area.

Section 2 gives a short overview about unavoidable trees and forests in graphs. Extremal and Ramsey-type results, both classic and recent, are summarized, and a connection of the main theorems to this kind of problems is made.

The first main theorem is presented in Section 3. It has an extremal flavor and examines unavoidable subtrees in larger trees. This theorem has resulted in a joint paper authored by Maria Axenovich and Georg Osang [6], which this section is based on.

In Section 4 the second main theorem is presented. It concerns an unavoidable substructure in balanced colorings of the complete bipartite graph $K_{n,n}$, and has characteristics of a Ramsey-type problem. It results from a joint work by Maria Axenovich, Marcus Krug, Georg Osang and Ignaz Rutter and is based on [5].

Section 5 concerns a partially solved problem about unavoidable paths that are totally multicolored. While the bounds presented here have been known previously, no simple proof specifically aimed at this problem has been published.

Section 6 gives an outlook on possible research directions related to the problems presented here, and concludes the thesis.

1.3 Unavoidable subtrees

The first main theorem of this thesis, which Section 3 is dedicated to, concerns subtrees of trees. Given a family of trees, how large can a tree be that contains no tree from the family as subtree?

There is extensive literature on unavoidable trees in tournaments, see for example [42], unavoidable trees in topological graphs [39], and more. The universal trees containing all trees of given order have been studied as well, see [11, 10, 27]. The algorithmic aspect of finding a given tree as a subtree in a larger given tree has been addressed in [43]: While subtree isomorphism in general is a generalization of the maximum clique problem and therefore NP-complete [36], for trees the problem can be solved in polynomial time. However, the extremal problem of unavoidable subtrees in trees did not receive its due attention.

Next we state the main result on unavoidable subtrees.

For a finite family $\mathcal{T}$ of k-vertex trees define $\text{ex}(\mathcal{T})$ to be the smallest integer such that any tree of at least this size contains a tree from $\mathcal{T}$ as subtree, if this integer exists, which it does exactly if $\mathcal{T}$ contains P_k and S_k.

Let $f(k,p,q)$ be the minimum of $\text{ex}(\mathcal{T})$ over all families of k-vertex trees containing P_k, S_k, p additional spiders and q additional non-spiders. Here a spider is a tree with at most one vertex of degree greater than 2, non-spiders are all other trees. We show:

Theorem 1. *$f(k,p,q) = 2^{\Theta(k \log^{p+1} k)}$, specifically,*

$$2^{4^{-q-1} k \log^{p+1} k(1+o(1))} \leq f(k,p,q) \leq 2^{k \log^{p+1} k(1+o(1))}.$$

The upper bound is guaranteed by a family $\mathcal{T} = \{S_k, P_k, Q_1, \ldots, Q_p\}$, where Q_i is a balanced spider of maximum degree $\log^i k$ for $i = 1, \ldots, p$.

Here, $\log^i$ denotes the i-times-iterated logarithm in base 2, i.e., $\log^i x = \log\log\cdots\log x$. With $f(k,p)$ being the minimum of $\text{ex}(\mathcal{T})$ over all families containing p trees on k vertices each, we get the following two corollaries:

Corollary 1. *$f(k,p) = f(k,p-2,0) = 2^{\Theta(k \log^{p-1} k)}$, specifically,*

$$2^{0.25k \log^{p-1} k(1+o(1))} \leq f(k,p) \leq 2^{k \log^{p-1} k(1+o(1))}.$$

Corollary 2. *$f(k,0,q) = 2^{\Theta(k \log k)}$, specifically,*

$$2^{0.5 \cdot 4^{-q} k \log k(1+o(1))} \leq f(k,0,q) \leq 2^{0.5k \log k}.$$

1.4 Fork forests in bi-colored complete bipartite graphs

The second main theorem, as presented in Section 4, examines fork forests in bipartite graphs.

In a graph whose edges are colored in black and white, let a fork be a path of length 2 with a black and a white edge. The question of concern is the following:

A balanced coloring is a coloring such that the number of edges in each color class differs by at most 1. In a balanced black-white coloring of $K_{n,n}$, how many vertex-disjoint forks all having their center in the same partite set can we find?

This problem was originally formulated as a lemma about matrices in [47] by Tverentina et al. They conjectured this to be $(1-\frac{1}{\sqrt{2}})n$, but only proved a weaker lower bound of $\frac{1}{2}(1-\frac{1}{\sqrt{2}})n$. We show that it is always possible to find such a fork forest containing $(1-\frac{1}{\sqrt{2}})n$ forks, and show that this is best possible. Furthermore, an algorithm finding the largest fork forest is presented.

Theorem 2. *In a balanced black-white coloring of $K_{n,n}$ there is always a fork forest containing $(1-\frac{1}{\sqrt{2}})n$ forks. There are colorings for which this is best possible.*

There is an algorithm finding a largest fork forest centered at X in any two-colored complete bipartite graph with partite sets X and Y and running in time $O(n^2 \log n\sqrt{n\alpha(n^2,n)\log n})$.

The theorem is presented more formally in section 4, including a complete proof. Solving this problem improves the exponential lower bound on resolution for ordered binary decision diagrams from [47], which was the context the problem originally appeared in. Beyond this application in formal logic, the problem also has connections to graph theory, as it belongs to a class of problems seeking color-alternating subgraphs or general large unavoidable subgraphs in two-edge colored graphs, see for example [3, 1, 15, 37].

2 Overview about unavoidable trees and forests in graphs

This section gives an overview over results on unavoidable trees and forests in graphs. Usually, the problems considered are of extremal nature.

"An extremal problem asks for the maximum or minimum value of a function over a class of objects." [48]

Here, the classes of objects are sets of certain (possibly colored) graphs that do not contain certain subgraphs. The extremal problems usually ask for the order or size of the largest graphs in such a class, meaning that any graph of greater order or size respectively contains one of the subgraphs.

The classical extremal question asks for the maximum number of edges a graph on n vertices can have without containing a certain subgraph H, or some graph from a family of subgraphs. This is equivalent to asking for the greatest size of a subgraph of K_n that does not contain H. More generally, we not only consider subgraphs of K_n, but also edge-maximal subgraphs of other graphs not containing H. The problem on unavoidable subtrees is of this flavor. Here the ground graph is not K_n, but an infinite tree, and we impose the additional restriction that the subgraph has to be connected.

Ramsey problems ask for the minimum number of vertices n that a graph from a class (in the classical version the complete graph K_n) must have such that any k-coloring of its edges has a monochromatic copy of certain other graphs. Ramsey numbers for matchings, especially in bipartite graphs, are of interest to the fork forest problem, as monochromatic matchings are part of fork forests, and matching edges of different colors can be joined into forks. There is a similar relation to paths, as crossing points of two paths of different color also imply a fork.

There are many other Ramsey variants, not only involving unavoidable monochromatic subgraphs, but also totally multicolored subgraphs and other specifically colored graphs. We might not just ask for what order a graph has to have such that a certain subgraph will appear, but also other questions like: for a graph of fixed order, the use of how many colors is required to force a certain subgraph. Other restrictions might be imposed on the coloring as well.

We dedicate a subsection to each of these three types of problems.

2.1 Extremal results

For a graph F, how many edges can a graph on n vertices have without containing F as subgraph?

The question was originally posed and solved by Turán [46] for the complete graph $F = K_k$ and earlier by Mantel for the triangle $F = K_3$.

More generally, for a family of graphs $\mathcal{F}$ we define $\text{ex}(n, \mathcal{F})$ to be the maximum number m such that there is a graph G on n vertices and m edges not containing any member of $\mathcal{F}$ as subgraph. If $\mathcal{F}$ contains only one element F, i.e., $\mathcal{F} = \{F\}$, we also write $\text{ex}(n, F)$ for $\text{ex}(n, \mathcal{F})$. Formally, $\text{ex}(n, F) = \max\{||G||\colon |G| = n, F \nsubseteq G\}$.

Turán's theorem answers the question for $F = K_k$ and shows that the edge maximal graphs not containing K_k are the complete $(k-1)$-partite graphs where the sizes of the partite sets differ by at most 1. These graphs are called Turán graphs.

The next type of graphs of interest are complete k-partite graphs. Because every graph can be embedded in K_s^k for some k and sufficiently large s, determining the behavior of $\text{ex}(n, K_s^k)$ gives us information about $\text{ex}(n, G)$ for any graph G if we know its chromatic number, i.e., the minimal number k such that G is k-partite.

The Erdős-Stone theorem [19] gives us exactly this information:

Theorem 3 (Erdős-Stone)**.**

$$\text{ex}(n, K_s^k) = \left(\frac{k-2}{k-1} + o(1)\right)\binom{n}{2}.$$

For bipartite graphs G, and therefore for trees, this means in particular that $\mathrm{ex}(n, G) = o(n^2)$. For trees however we now prove a much stronger statement. For a tree T, $\mathrm{ex}(n, T)$ is in fact linear in n. To show this, we first need the following lemma.

Lemma 1. *A graph G on n vertices and $m > 0$ edges of average degree $d = 2m/n$ contains any tree T on $d/2 + 2$ vertices.*

Proof. We first prove that there is a subgraph $H \subseteq G$ with minimum degree at least $d/2 + 1$: Iteratively remove a vertex of degree at most $d/2$, until all vertices have degree greater than $d/2$. By doing so, in each step we remove at most $d/2$ edges and one vertex, therefore reducing the sum of the degrees of all vertices by at most d. Therefore the resulting graph still has an average degree of at least d.

We can greedily embed T in H: Order the vertices of T in such a way that each vertex is adjacent to only one of the previous ones. Such an ordering can be obtained by for instance doing a breadth-first search. Embed the first vertex anywhere in H, and for each further vertex choose a vertex incident to its neighbor from the previous vertices. This is possible due to the minimum degree of H. □

Theorem 4. *For a tree T on k vertices, $\mathrm{ex}(n, T) \leq (k-2)n$, and if n is divisible by $k-1$, then $\frac{1}{2}(k-2)n \leq \mathrm{ex}(n, T)$*

Proof. For the upper bound, consider a graph G on n vertices and $(k-2)n$ edges. This graph has average degree $2(k-2)$. Due to Lemma 1, we can find any tree on $2(k-2)/2 + 2 = k$ vertices in G, and in particular T.

For the lower bound, $n/(k-1)$ vertex-disjoint copies of K_{k-1} suffice as graph that does not contain T. □

The Erdős-Sós conjecture states that for any k-vertex tree T in fact $\mathrm{ex}(n, T) \leq n(k-2)/2$. Ajtai, and Komlós, Simonovits and Szemerédi recently proved the conjecture for sufficiently large trees [2] making use of Szemerédi's regularity lemma. Previously it had been proven for instance for certain spiders [20], or for trees of diameter at most 4 [38].

Szemerédi's regularity lemma receives special mention here because while it is a very powerful tool in graph theory, it only yields meaningful statements for immensely large graphs. Gowers showed in [28] that bounds given by it cannot be improved to make it applicable to graphs of any practically relevant size.

For the star S_k, it is easy to see that $\mathrm{ex}(n, S_k) = \lfloor n(k-2)/2 \rfloor$. For even n or k, any $(k-2)$-regular graph shows the lower bound. If both are odd, an almost $(k-2)$-regular graph with one vertex of degree $k-3$ does. The upper bound follows as no vertex may have degree greater than $k-2$.

For paths, the Erdős-Sós conjecture holds and was already shown back in 1959 by Erdős and Gallai [17]. The result was refined by Faudree and Schelp [22] to take into account the cases when n is not divisible by $k-1$:

Theorem 5. *For a path P_k on k vertices, and $n = p(k-1) + q$ for $0 \le q < k-1$,*

$$\mathrm{ex}(n, P_k) = p(k-1)(k-2)/2 + r(r-1)/2.$$

The vertex-disjoint union of p copies of K_{k-1} and one copy of K_q is the only graph which achieves this number of edges without containing P_k.

Also due to Erdős and Gallai [17], the extremal number for a matching has been determined:

Theorem 6.

$$\mathrm{ex}(n, kK_2) = \max\left\{\binom{2k-1}{2}, \binom{k-1}{2} + (k-1)(n-k+1)\right\}$$

2.2 Ramsey results

The Ramsey number of two graphs G and H, $R(G, H)$, is the smallest number n such that any 2-coloring of the edges of K_n with the colors red and blue contains a red copy of G or a blue copy of H. It is also sometimes called "the Ramsey number of G versus H". More generally, $R(G_1, G_2, \ldots, G_k)$ is the smallest number n such that any k-coloring of K_n has a copy of G_i in color i for some $i = 1, \ldots, k$.

The classic Ramsey numbers, $R(K_n, K_m)$, were proven to exist by Frank P. Ramsey in 1930 [41]. (In fact, he proved a much more general version.) Only for very few of these the exact values are known. As any graph on n vertices is a subgraph of K_n, $R(G, H)$ exists as well for any two graphs G and H.

There are many results about the Ramsey numbers of a tree versus some other graphs, like complete graphs or cycles. We shall only list some results for Ramsey numbers where all the involved graphs are trees.

Ramsey numbers for complete graphs are exponential [13, 45] and only very few small ones are known exactly. The latter is not too surprising, as all colorings have to be considered, and the computational problem of determining whether a monochromatic clique of at least a certain size exists is NP-complete [36].

Conversely, Ramsey numbers for trees are linear, and the exact values are known for some classes of trees. A result by Chen and Schelp [9] implies that $R(T, T)$ is linear for trees T, i.e., there is a constant c such that for sufficiently large n, for any tree T on n vertices $R(T, T) \le cn$. First they define a term p-arrangeable as follows: A graph is p-arrangeable if there is an ordering of the vertices $v_1, \ldots, v_n$ such that for any $1 \le i \le n-1$ with $L_i = \{v_j : j \le i\}$ and $R_i = \{v_j : j > i\}$ it holds that $|N(N(\{v_i\}) \cap R_i) \cap L_i| \le p$, where $N(S)$ denotes the neighborhood of the vertex set S. Using Szemerédi's regularity lemma, they show that for any p there is some c such that for sufficiently large n the following holds: For any p-arrangeable graph G on n vertices $R(G, G) \le cn$. They show that planar graphs are p-arrangeable for some p, and specifically that trees are

1-arrangeable.

However, the linearity of Ramsey numbers for trees can also be proven more elementarily:

Theorem 7. *For trees G and H each on n vertices, $R(G, H) \leq 4(n-2)+1$*

Proof. Assume a red-blue coloring of $K_{4(n-2)+1}$, and consider the graph of the color class that has more edges. Its average degree is at least $d = 2(n-2)$. By Lemma 1 we find all trees on $d/2 + 2 = n$ vertices in this graph, so specifically G or H, depending on which color class was greater. □

For general trees T, good lower bounds on $R(T, T)$ are known as well, as mentioned in [33]. As a tree is always a bipartite graph due to its lack of cycles and therefore specifically odd cycles, we can consider a partition of its vertices into partite sets X and Y. (Algorithmically, such a partition can be obtained by a simple breadth-first search. Vertices whose depths from a root have the same parity belong in the same partite set.)

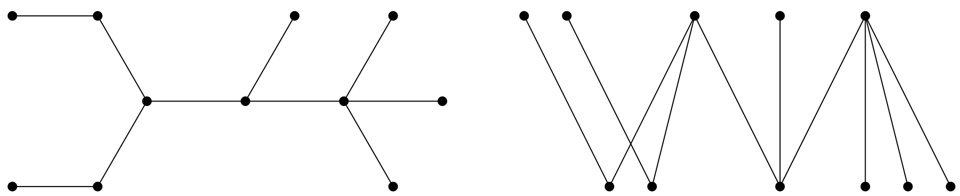


Figure 2: A tree and a bipartition of its vertices.

Theorem 8. *For a tree T with partite sets X and Y when viewed as a bipartite graph, $|X| = t_1$, $|Y| = t_2$ with $t_2 \geq t_1$, the following holds:*

$$R(T, T) \geq \begin{cases} 2t_1 + t_2 - 1 & \text{for } 2t_1 \geq t_2 \geq t_1 \geq 2, \\ 2t_2 - 1 & \text{for } 2t_1 < t_2. \end{cases}$$

Proof. In the first case, consider the following red-blue coloring of $G = K_{2t_1+t_2-1}$. Partition the vertex set of G into C_1 and C_2 with $|C_1| = t_1 + t_2 - 1$ and $|C_2| = t_1 - 1$. Color all edges within C_1 and those within C_2 in red, color edges between vertices not in the same set in blue. There is obviously no red copy of T as the connected components of the red subgraph contain fewer vertices than T each. There is no blue copy of T as X and Y each have more vertices than C_2 can accommodate.

In the second case, consider a similar coloring with the sets C_1 and C_2 both having size $t_2 - 1$. The same argument applies. □

We will see in the following survey that for some classes of trees, these lower bounds are indeed the Ramsey numbers. However there are also trees where the actual Ramsey number is greater. It has been shown [29] that for some trees called double stars the Ramsey number is greater by 1 than the lower bound given above.

Haxell et al. however managed to show, using Szemerédi's regularity lemma, that the actual Ramsey numbers cannot differ from these bounds by too large a factor for sufficiently large trees with a sufficiently small maximum degree. [33]

Theorem 9. *For any given $\eta > 0$ there are $N = N(\eta)$ and $\delta = \delta(\eta)$ such that the following holds: For every tree T with a bipartition into X and Y with $|X| = t_1$, $|Y| = t_2$ and $t_2 \geq t_1$, if $t_2 \geq N$ and the maximum degree $\Delta(T) \leq \delta t_2$, then*

$$R(T,T) \leq \begin{cases} (1+\eta)(2t_1 + t_2 - 1) & \text{for } 2t_1 \geq t_2 \geq t_1 \geq 2, \\ (1+\eta)(2t_2 - 1) & \text{for } 2t_1 < t_2. \end{cases}$$

We now review the Ramsey numbers for specific classes of trees.

Starting with Ramsey numbers for paths, Gerencsér and Gyárfás proved the following result in 1967 [26]:

Theorem 10. *For $m \geq n$,*

$$R(P_m, P_n) = m - 1 + \left\lfloor \frac{n}{2} \right\rfloor .$$

Specifically, if $m = n$,

$$R(P_n, P_n) = \left\lfloor \frac{3n-2}{2} \right\rfloor .$$

The lower bound can be seen by providing a red-blue coloring of the graph on $R(P_m, P_n) - 1$ vertices that does not contain a red P_m or blue P_n. Pick a set M of m vertices and color all edges between them in red, color the remaining edges of the graph in blue. Evidently, there is no red P_m. As every second vertex along a blue path must be in $V(G) - M$, it can consist of at most $2(R(P_m, P_n) - m) + 1 = -1 + 2(\lfloor \frac{n}{2} \rfloor) < n$ vertices.

For more than two paths, the Ramsey number has not been determined for all cases yet. However, several special cases have been analysed.

If the first path is significantly longer than the rest, and all but the first two paths have an even number of vertices, Faudree and Schelp determined the Ramsey number in [22]:

Theorem 11. *If $k \geq 1$, $r_i \geq 1$ for $1 \leq i \leq k$, $\delta_1 = 0$ or 1, and $r_0 \geq 6(\sum_{i=1}^{k} r_i)^2$, then*

$$R(P_{r_0}, P_{2r_1+\delta_1}, P_{2r_2}, \cdots, P_{2r_k}) = \left(\sum_{i=0}^{k} r_i \right) - k.$$

In the special case with 3 paths, with the constraints from above and $\delta_2 = 0$ or 1, the following holds:

$$R(P_{r_0}, P_{2r_1+\delta_1}, P_{2r_2+\delta_2}) = \left(\sum_{i=0}^{2} r_i \right) - k.$$

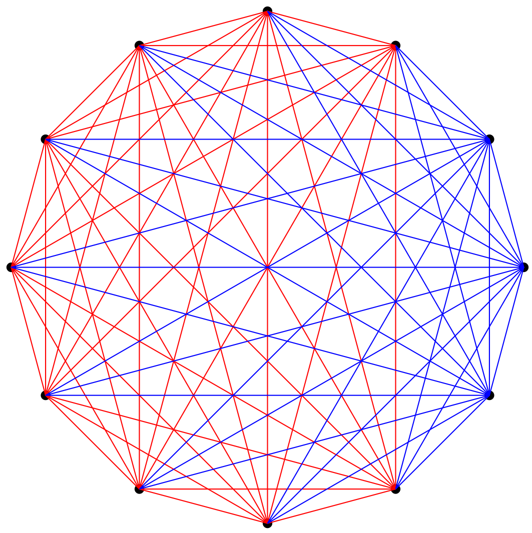


Figure 3: A coloring of the K_{12} avoiding both a red P_{10} and a blue P_8. All edges having at least one of the three rightmost vertices as endpoint are blue, the remaining edges are red. $R(P_{10}, P_8) = 13$.

Bielak managed to relax the size condition on r_0 to some extent in [7], however r_0 is still required to be quadratic in the sum of the other paths' lengths.

For three paths of equal length, the following recent result gives a solution if the paths are sufficiently long [32]:

Theorem 12. *If n is sufficiently large, then*

$$R(P_n, P_n, P_n) = \begin{cases} 2n-1 & \text{for odd } n, \\ 2n-2 & \text{for even } n. \end{cases}$$

The proof makes use of Szemerédi's regularity lemma.

Paths in the complete bipartite graph have been examined as well. Let $B(G_1, G_2) = (n, m)$ for n and m defined as follows: Any 2-coloring of $K_{r,s}$ contains a copy of G_1 in the first color or a copy of G_2 in the second color if and only if $r \geq n$ and $s \geq m$. Note that this is not necessarily well defined. For instance, if the condition about containing copies of G_1 or G_2 holds for $(r, s) = (a, a+1)$ and for $(r, s) = (a+1, a)$, but not for $(r, s) = (a, a)$ for some a, then there is no such pair (n, m).

For two paths, the above has been shown to be well defined and exact results are known due to [21, 30].

Theorem 13. *$B(P_n, P_m)$ is well defined for positive integers n and m and the following cases determine the values for all possible n and m:*

$$\begin{array}{lcll} B(P_{2n}, P_{2m}) & = & (n+m-1, n+m-1), & \\ B(P_{2n+1}, P_{2m}) & = & (n+m, n+m-1) & \textit{for } n \geq m-1, \\ B(P_{2n+1}, P_{2m}) & = & (n+m-1, n+m-1) & \textit{for } n < m-1, \\ B(P_{2n+1}, P_{2m+1}) & = & (n+m, n+m-1) & \textit{for } n \neq m, \\ B(P_{2n+1}, P_{2n+1}) & = & (2n+1, 2n-1). & \end{array}$$

Cockayne and Lorimer found the Ramsey number for c matchings in 1975 [12]:

Theorem 14. *If $m_1, \ldots, m_c$ are positive integers and $n_1 = \max\{m_1, \ldots, m_c\}$, then*

$$R(m_1K_2, \ldots, m_cK_2) = m_1 + 1 + \sum_{i=1}^{c}(m_i - 1).$$

Powers [40] investigated unavoidable matchings in r-partite graphs K_i^r with balanced partite sets. But instead of asking for the minimum value of n that enforces some monochromatic matching of the required size in any coloring, he viewed the problem from a slightly different perspective. The i^{th} Ramsey number $R_i(m_1K_2, \ldots, m_cK_2)$ asks for the minimum index r such that any c-coloring of the edges of K_i^r using colors $1, \ldots, c$ forces a copy of m_jK_2 entirely colored in color j for some $1 \leq j \leq c$.

Theorem 15. *Let $m_1, \cdots, m_c$ be positive integers and $m_1 = \max\{m_1, \ldots, m_c\}$. Define $s = \sum_{j=1}^{k}(m_j - 1)$ and $p = m_1 + s \bmod i$.*

$$R_i(m_1K_2, \ldots, m_cK_2) = \begin{cases} \lfloor (m_1+s)/i \rfloor + 1 & \textit{if } p < m_1, \\ \lfloor (m_1+s)/i \rfloor + 2 & \textit{if } p \geq m_1. \end{cases}$$

Using this, in the same paper Powers determines the Ramsey number for matchings in balanced bipartite graphs, here denoted as $R_{\text{bipartite}}$, i.e., the smallest number i such that any c-coloring of the edges of $K_{i,i}$ using colors $1, \ldots, c$ forces a copy of m_jK_2 entirely colored in color j for some $1 \leq j \leq c$.

Theorem 16. *If $m_1, \ldots, m_c$ are positive integers and $n_1 = \max\{m_1, \ldots, m_c\}$, then*

$$R_{bipartite}(m_1K_2, \ldots, m_cK_2) = \sum_{j=1}^{c}(m_j - 1) + 1.$$

A generalisation of matchings are linear forests. Linear forests are vertex-disjoint unions of paths. Specifically, an (n, j) linear forest is a forest of non-trivial paths with n vertices in total, and j of the paths having an odd number of vertices. Faudree and Schelp found the Ramsey number for two linear forests [23]:

Theorem 17. *If L_1 and L_2 are (n_1, j_1) and (n_2, j_2) linear forests respectively, then*

$$R(L_1, L_2) = \max\{n_1 + (n_2 - j_2)/2 - 1, n_2 + (n_1 - j_1)/2 - 1\}.$$

For c stars, the Ramsey numbers are known as well due to Burr and Roberts [8].

Theorem 18. *Let $m_1, \cdots, m_c$ be positive integers t of which are even, then*

$$R(S_{m_1+1}, \ldots, S_{m_c+1}) = \begin{cases} (\sum_{i=1}^{k} m_i) - k + 1 & \text{for even } t \neq 0, \\ (\sum_{i=1}^{k} m_i) - k + 2 & \text{otherwise.} \end{cases}$$

2.3 Other Ramsey-type results

There is a great variety of other Ramsey-type problems, asking for unavoidable structures in colored graphs. The results are too numerous to be listed here, but in the following some of the concepts are presented.

A graph is called rainbow if no two edges have the same color. Ramsey-type problems that involve rainbow subgraphs have been surveyed by Fujita et al. [24]

The anti-Ramsey number $\mathrm{ar}(G, H)$ is the maximum number n of colors for which there is an edge coloring of G with n colors such that there is no rainbow copy of H in G. Equivalently, $n+1$ is the smallest number such that any coloring of G using $n+1$ colors forces a rainbow copy of H. Anti-Ramsey have first been studied by Erdős, Simonovits and Sós [18], and results with regards to trees are covered in the survey [24] in Section 2.3 and Section 2.4. Specifically, for the path in the complete graph, the anti-Ramsey number has been determined in [44] under certain restrictions:

Theorem 19. *There is some constant c such that for all $t \geq 5$, $n > c \cdot t^2$ and $\varepsilon = 0, 1$,*

$$\mathrm{ar}(n, P_{2t+3+\varepsilon}) = t \cdot n - \binom{t+1}{2} + 1 + \varepsilon.$$

Another variation of similar style are mixed Ramsey numbers. The maximal (minimal) mixed Ramsey number $R_{\max}(n, G)$ ($R_{\min}(n, G)$) is the maximum (minimum) number k such that there is a coloring on k colors of K_n having neither a monochromatic nor a rainbow copy of G.

Constrained or rainbow Ramsey numbers are a different variation asking to avoid both monochromatic and rainbow subgraphs. The rainbow Ramsey number $RR(G, H)$ of two graphs is the minimum number n such that any coloring of K_n using arbitrarily many colors contains a monochromatic copy of G or a rainbow copy of H. It is shown in [35] that this number exists if and only if G is a star or H is a forest.

However, it is possible to add a third graph F such that the family of G, H and F becomes unavoidable in some sense. For this purpose, we define a coloring to be lexical if there is an ordering of the vertices such that edges having the same smaller endpoint (with regards to the ordering) are colored with the same color.

Now the canonical Ramsey theorem, first proved by Erdős and Rado [16], states the following:

Theorem 20 (Canonical Ramsey Theorem)**.** *For any graph H there is an n such that any coloring of K_n using arbitrarily many colors contains a monochromatic, a rainbow or a lexical copy of H.*

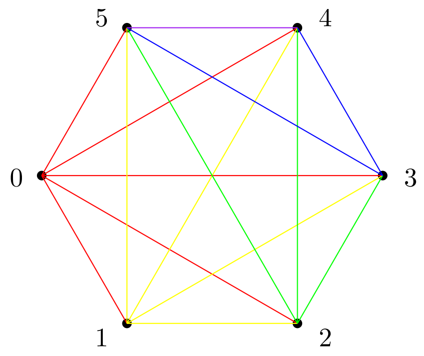


Figure 4: A lexically colored copy of K_6.

A very general notion are pattern Ramsey numbers. A pattern is a graph with colored edges. For a given pattern or family of patterns $\mathcal{F}$, the pattern Ramsey number $f(\mathcal{F})$ is the minimum number n such that any coloring of K_n with arbitrarily many colors contains a copy of some pattern. A family $\mathcal{F}$ for which $f(\mathcal{F})$ exists is called a Ramsey family. Using these terms, another formulation of the canonical Ramsey theorem is to say that a family is a Ramsey family if it contains the monochromatic colorings of a graph, the rainbow colorings of another graph, and the lexical colorings of a third graph.

Pattern Ramsey numbers also allow us to describe the rainbow path problem covered in Section 5 in a new way. A coloring where no two incident edges have the same color is called a proper coloring. This is equivalent to the coloring having no monochromatic P_3 as subgraphs. If we define the family $\mathcal{F}$ to contain the monochromatic colorings of P_3 and the rainbow colorings of P_n, then $f(\mathcal{F})$ is exactly the number the rainbow path problem asks for.

A fork forest can also be considered a pattern, and the corresponding problem asks for the smallest n such that any balanced 2-coloring of the $K_{n,n}$ contains this pattern as subgraph. Compared to the general pattern Ramsey numbers, here we have additional restrictions imposed on the coloring, and the graph we seek the pattern in is the $K_{n,n}$ instead of the K_n.

3 Unavoidable subtrees

3.1 Introduction

In this section we investigate how large trees can be without containing certain subtrees.

For a finite family $\mathcal{T}$ of k-vertex trees and a tree T, we write $T \to \mathcal{T}$ if T contains at least one of the trees from $\mathcal{T}$ as a subtree, we write $T \not\to \mathcal{T}$ otherwise. We call a family $\mathcal{T}$ *unavoidable* if there is an integer n such that for any tree T on at least n vertices $T \to \mathcal{T}$. We denote the set of all n-vertex trees $\mathcal{T}_n$. Recall that S_n denoted an n-vertex star and P_n an n-vertex path.

Clearly, not every family of trees is unavoidable. Observe that each unavoidable family $\mathcal{T}$ of trees must contain a path and a star, otherwise either an arbitrarily large star or an arbitrarily long path will avoid $\mathcal{T}$.

Here, we investigate the smallest value of n such that any tree on n vertices contains a member of $\mathcal{T}$. Let, for an unavoidable family $\mathcal{T}$ of trees,

$$\mathrm{ex}(\mathcal{T}) = \min\{n : \forall T \in \mathcal{T}_n,\ T \to \mathcal{T}\}.$$

In particular, for any tree T on at least $\mathrm{ex}(\mathcal{T})$ vertices $T \to \mathcal{T}$, but there is a tree on $\mathrm{ex}(\mathcal{T}) - 1$ vertices such that $T \not\to \mathcal{T}$.

We pick up more definitions in a separate subsection, as well as definitions only used in specific places along the way in the respective subsections.

Before working on the main theorem, we consider the special case where the family of trees to avoid consists of only the star and the path. We find the exact value of $\mathrm{ex}(S_k, P_k)$ and thereby solve this problem in Section 3.3.

We then proceed to the main theorem. Recall that

$$f(k,p,q) = \min\{\mathrm{ex}(\mathcal{T})\colon \mathcal{T} \subseteq \mathcal{T}_k \text{ is a union of } p+2 \text{ spiders and } q \text{ non-spiders}\}.$$

The theorem we aim to prove is the following:

Theorem 1. $f(k,p,q) = 2^{\Theta(k \log^{p+1} k)}$, *specifically,*

$$2^{4^{-q-1} k \log^{p+1} k (1+o(1))} \le f(k,p,q) \le 2^{k \log^{p+1} k (1+o(1))}.$$

The upper bound is guaranteed by a family $\mathcal{T} = \{S_k, P_k, Q_1, \ldots, Q_p\}$, where Q_i is a balanced spider of maximum degree $\log^i k$ for $i = 1, \ldots, p$.

Here and also in the following, the logarithms are base 2 and $\log^i x = \log\log\cdots\log x$, where log is iterated i times.

To prove it, we proceed in 3 steps.

We first show the upper bound, by taking the family of spiders from above and showing that any tree on more than $2^{k \log^{p+1} k (1+o(1))}$ vertices has one of them as a subtree.

In the second step we show that $2^{0.25 k \log^{p+1} k (1+o(1))} \le f(k,p,0)$. We do this the following way: For any family containing the star, the path and p additional spiders, we construct a tree T that avoids all the trees of T, and show that it has at least the size $2^{0.25 k \log^{p+1} k (1+o(1))}$. This does not give a bound for the main theorem yet, but provides the basis to establish the lower bound in step 3.

In the third and final step we consider general families of trees containing the star, the path, p additional spiders and q non-spiders. For each of those families we construct a tree to avoid it. We first only consider the spiders in the family, and use the construction from step 2 as a basis which avoids them. We then trim this tree in a certain way to avoid the remaining non-spiders, and estimate the effect this has on the size of the tree. The resulting tree has at least $2^{4^{-q-1} k \log^{p+1} k (1+o(1))}$ vertices, giving us the lower bound.

3.2 Definitions

Before formulating and proving the theorems in this section, we need some definitions.

Recall that a spider is a tree with at most one vertex of degree greater than 2. For a spider that is not a path, the vertex of maximum degree is the *head* or *center* of the spider; the spider is a union of paths, called *legs*, where one endpoint is a leaf and the other is the head.

We denote the maximum degree of a graph and its diameter with Δ and diam respectively.

Observe that if $Q \subseteq T$, then $\Delta(Q) \leq \Delta(T)$ and $\mathrm{diam}(Q) \leq \mathrm{diam}(T)$.

Henceforth, sequences and vectors are written in bold script, and their elements referred to in subscript, e.g., a sequence $\mathbf{b} = (b_0, \ldots)$, or a vector $\mathbf{u} = (u_0, u_1, \ldots, u_m)$ for some $m \geq 0$. We shall always index the elements of the vectors starting with 0. Most of our results use balanced (stable) rooted trees where the vertices at the same distance from the root have the same degree. Formally, for a vector $\mathbf{u}$, with $u_i \geq 2$, $i = 1, \ldots, m$ and $u_0 \geq 1$, let a *balanced tree with vector* $\mathbf{u}$, $B(\mathbf{u})$, be a rooted tree of depth $m+1$ in which all vertices at distance i from the root have degree u_i, $i = 0, \ldots, m$. The vertices of distance $m+1$ from the root are the leaves. The diameter of $B(\mathbf{u})$ is $2m+2$. We have that

$$\begin{aligned} |V(B(\mathbf{u}))| &= 1 + u_0 + u_0(u_1 - 1) + \cdots + u_0(u_1-1)(u_2-1)\cdots(u_m-1) \\ &\geq u_0 \prod_{i=1}^{m}(u_i - 1). \end{aligned} \tag{1}$$

A complete k-ary tree, T, of depth r is a balanced tree with vector $(k, k+1, \ldots, k+1)$, where $k+1$ is repeated $r-2$ times.

In all calculations we omit floors and ceilings when their usage is clear from the context.

3.3 Star and path

We start off with just the star and the path to avoid. In this situation we can find the exact size of the largest tree that avoids both of these trees as subtrees, and subsequently $\mathrm{ex}(\{S_{k+1}, P_{k+1}\})$, which is larger by 1:

Proposition 21. $\mathrm{ex}(\{S_{k+1}, P_{k+1}\}) = \begin{cases} 2 + \frac{k-1}{k-3}((k-2)^{(k-1)/2} - 1) & \textit{if } k \textit{ is odd} \\ 3 + 2\frac{k-2}{k-3}((k-2)^{(k-2)/2} - 1) & \textit{if } k \textit{ is even} \end{cases}$ $= 2^{0.5k \log k (1+o(1))}$.

Proof. Observe first that if T is a tree of largest order avoiding S_{k+1} and P_{k+1}, then the longest path in G has length $k-1$, otherwise one can subdivide an edge in a longest path of T to obtain a larger such graph.

If k is odd, then $B(k-1, \ldots, k-1)$ of depth $(k-1)/2$ has diameter $k-1$, maximum degree $k-1$ and $1 + \frac{k-1}{k-3}((k-2)^{(k-1)/2} - 1)$ vertices. Consider some tree T avoiding P_{k+1}

and S_{k+1}. Let $\{c\}$ be the center of a longest path P of length $k-1$. Any other vertex of T is at distance at most $(k-1)/2$ from c because otherwise a path of length at least $(k+1)/2$ from c and a sub-path of P with endpoint c together give a longer path than P. Thus T is a tree rooted at c with depth at most $(k-1)/2$ and maximum degree at most $k-1$, so $|V(T)| \leq 1 + \frac{k-1}{k-3}((k-2)^{(k-1)/2} - 1)$.

If k is even, then two graphs isomorphic to $B(k-2, k-1, \dots, k-1)$ of depth $(k-2)/2$, linked together at the roots by another edge, form a graph of diameter $k-1$, maximum degree of $k-1$ and with $2 + 2\frac{k-2}{k-3}((k-2)^{(k-2)/2} - 1)$ vertices. Consider some tree T avoiding P_{k+1} and S_{k+1}. Let $\{r, l\}$ be the center of the longest path P of length at most $k-1$. Define L and R to be the trees rooted at r and l, respectively, and obtained from T by deleting the edge rl. As in the previous case, any $v \in V(L)$ has distance at most $(k-2)/2$ from l, and the degree of l in L is at most $k-2$ within L. So, L, and by symmetry, R are rooted trees of depth at most $(k-2)/2$, the degree of each vertex is at most $k-1$ and the degree of the root is at most $k-2$. So, $|V(T)| = |V(L)| + |V(R)| \leq 2 + 2\frac{k-2}{k-3}((k-2)^{(k-2)/2} - 1)$.

□

3.4 The upper bound

As the first part of the main theorem, we prove the upper bound:

Theorem 1 (Part 1)**.** *Let $\mathcal{T}$ be the family $\{S_k, P_k, Q_1, \dots, Q_p\}$, where Q_i is a balanced spider of maximum degree $\log^i k$ for $i = 1, \dots, p$. Then any tree on more than $2^{k \log^{p+1} k(1+o(1))}$ vertices will contain a tree from $\mathcal{T}$ as subtree.*

We need the following definitions for the proof:

For a tree T, and a root $r \in V(T)$, define a partial order on $V(T)$ naturally with $v' \leq v$ if v is on the v'-r-path. Intuitively, the closer to the root a vertex is, the greater it is with regards to this partial order. All subtrees are rooted respecting the original order. We say that a subtree is *inherited* by a vertex v if its vertex set consists of all vertices u such that $u \leq v$. The children of a vertex v are those vertices v' adjacent to v with $v' \leq v$. The parent of a vertex v, $v \neq r$ is a vertex u adjacent to v, $v \leq u$. The *inherited subtree depth* of a vertex v in a rooted tree is the largest distance to a vertex u with $u \leq v$, i.e., the depth of the tree inherited by v. For a rooted tree we always assume that the partial order with respect to this root is implied.

We now prove the upper bound:

Proof. Let $\mathcal{T} = \{S_k, P_k, Q_1, \dots, Q_p\}$, where Q_i is a balanced spider of maximum degree $\log^i k$ for $i = 1, \dots, p$. Let T be a tree that avoids all $\mathcal{T}$. We have to show that $|V(T)| \leq 2^{k \log^{p+1} k(1+o(1))}$. Observe first that $\Delta(T) \leq k-2$ and $\mathrm{diam}(T) \leq k-2$. Fix some vertex r to be the root of T and consider the partial order of vertices with respect to this root. We say that a vertex is ***i-small*** if its inherited subtree has depth at most $k/\log^{i+1} k$.

Claim: For $i = 0, \dots, p-1$ a tree inherited by an i-small vertex in T has at most $s_i := 2^{(i+1)k(1+o(1))}$ vertices.

We proceed by induction on i. If v is a 0-small vertex and S is the tree inherited by v, then $\Delta(S) \le \Delta(T) \le k$ and $\text{depth}(S) \le k/\log k$ by definition. Thus $|V(S)| \le k^{k/\log k} = 2^k$.

Consider an $(i+1)$-small vertex v and its inherited subtree S. Obtain S' from S by removing all i-small vertices. Observe that each of the remaining vertices has inherited subtree depth greater than $k/\log^{i+1} k$ in T. If there is a vertex u in S' of degree at least $\log^{i+1} k$ in S', then each child of u in S' has inherited subtree depth greater than $k/\log^{i+1} k$ in T due to being in S', so u is a center of a copy of Q_{i+1}, a contradiction. Thus $\Delta(S') < \log^{i+1} k$ in S'. As $\text{depth}(S') \le \text{depth}(S) \le k/\log^{i+2} k$ by definition, we get that $|V(S')| \le (\log^{i+1} k)^{k/\log^{i+2} k} = 2^k$. We have that a tree S is a union of S' and i-small trees inherited by children of some vertices of S'. Each vertex in S' has at most k children in S, each of which inherits (in S) at most one such a subtree of size at most s_i. This yields

$$|V(S)| \le |V(S')|ks_i \le k \cdot 2^k \cdot 2^{(i+1)k(1+o(1))} = 2^{(i+2)k(1+o(1))} = s_{i+1},$$

and proves the claim.

Now, we shall consider a tree T and apply an argument almost identical to the one used in the claim. Delete all $(p-1)$-small vertices from T. The resulting tree T' has maximum degree at most $\log^p k$ and it has depth at most k since $\text{diam}(T) \le k$. Thus $|V(T')| \le (\log^p k)^k = 2^{k \log^{p+1} k}$. As before, each vertex of T' has at most k neighbors, each of which inherits (in T) at most one such a tree of size at most $s_{p-1} = 2^{pk(1+o(1))}$. Thus,

$$|V(T)| \le |V(T')|ks_{p-1} \le 2^{k \log^{p+1} k} \cdot k \cdot 2^{pk(1+o(1))} = 2^{k \log^{p+1} k(1+o(1))},$$

concluding the proof of the upper bound.

□

3.5 Construction of a tree avoiding a family of spiders

We first find a simplified family $\mathcal{T}''$ of spiders such that if we avoid $\mathcal{T}''$, we can guarantee that we also avoid $\mathcal{T}$. We then construct a tree which avoids $\mathcal{T}''$.

For a given family $\mathcal{T}$ of k-vertex spiders of diameter at most $k/2$ with maximum degrees $\Delta_1, \dots, \Delta_x$, for some x, define a *reduced family of spiders* $\mathcal{T}'' = \mathcal{T}''(\mathcal{T})$ to consist of x balanced spiders, where the i^{th} spider has maximum degree Δ_i and leg length $(k/2)/\Delta_i$, $i = 1, \dots, x$.

Lemma 2. *For any balanced tree B of diameter at most $k/2$, and any family $\mathcal{T}$ of k-vertex spiders, $B \not\to \mathcal{T}''$ implies that $B \not\to \mathcal{T}$.*

Proof. Let B be a balanced tree of diameter at most $k/2$. Assume that $Q \subseteq B$, for some $Q \in \mathcal{T}$ with Δ legs of lengths $\ell_1 \geq \ell_2 \geq \cdots \geq \ell_\Delta$. Since B is balanced, it contains a balanced spider with Δ legs of length ℓ_2. Because $\ell_1 \leq k/2$, we have that $\ell_2 \geq (k-\ell_1)/(\Delta-1) \geq (k/2)/\Delta =: \ell$. The balanced spider with Δ legs of length ℓ is in $\mathcal{T}''$. □

Construction 1 (Construction avoiding a family of spiders)**.** *Let $\mathcal{T}$ be a given family of k-vertex spiders containing S_k and P_k. We will construct a tree of diameter $k/2$, so let $\mathcal{T}'$ be a subfamily of $\mathcal{T}$ consisting only of spiders of diameter at most $k/2$. Let $\mathcal{T}'' = \mathcal{T}''(\mathcal{T}')$ be the reduced family of spiders of maximum degrees Δ_i and leg-length ℓ_i, respectively, $i = 0,\dots,x$, for some x, $x \leq |\mathcal{T}|-2$. Let $k-1 = \Delta_0 > \Delta_1 > \Delta_2 > \cdots > \Delta_x \geq 3$. Note that $1 = \ell_0 < \ell_1 < \ell_2 < \ldots < \ell_x$, $\ell_i = (k/2)/\Delta_i$, and define $\ell_{x+1} := k/4-1$. Let*

$$\mathbf{u} = (\Delta_x - 1, \underbrace{\Delta_x - 1,\dots,\Delta_x - 1}_{\ell_{x+1}-\ell_x}, \dots, \underbrace{\Delta_i - 1,\dots,\Delta_i - 1}_{\ell_{i+1}-\ell_i}, \dots, \underbrace{\Delta_0 - 1,\dots,\Delta_0 - 1}_{\ell_1-\ell_0})$$

be a vector with a total of $k/4-1$ entries. Define the intervals $I_i := \{j \geq 1 \colon u_j = \Delta_i - 1\}$, i.e., the set of positions (except for position 0) occupied by $\Delta_i - 1$, for $i = 0,\dots,x$. Let our desired tree to avoid the family of spiders $\mathcal{T}$ (including S_k and P_k) be

$$T = T_s(\mathcal{T}) := B(\mathbf{u}).$$

Note that, using (1), $|V(T)| \geq \prod_{i=0}^{x}(\Delta_i - 2)^{|I_i|} = \prod_{i=0}^{x}(\Delta_i - 2)^{\ell_{i+1}-\ell_i}$.

Theorem 1 (Part 2)**.** *The constructed tree T avoids all trees of the family $\mathcal{T}$ and has size at least $2^{\frac{k}{4}\log^{p+1} k(1+o(1))}$.*

Proof. T avoids all trees of the family:

Any vertex v of degree at least Δ_i in T has inherited subtree depth of less than ℓ_i in T, $i = 1,\dots,q-2$, that is, any path leading from v away from the root has length less than ℓ_i. So if v were chosen as the center to embed the spider from $\mathcal{T}''$ of maximum degree Δ_i in, it would be impossible to fit in the legs of the spider in T. Thus $T \not\to \mathcal{T}''$. By Lemma 2 the tree T avoids the trees of diameter at most $k/2$, and by the fact that T has diameter less than $k/2$, we have that $T \not\to \mathcal{T}$.

Analysis of the size of T:

Recall that the number of spiders in the reduced family x is at most p. Recall further that $\ell_{x+1} := k/4-1$ and formally define $\Delta_{x+1} := (k/2)/\Delta_{x+1}$ so the property $\ell_i = (k/2)/\Delta_i$ is also fulfilled for $i = x+1$. Recall that $T = B(\mathbf{u})$, where $\mathbf{u}$ has blocks of indices $I_x,\dots,I_0$ with entries of $\Delta_i - 1$ in the block corresponding to I_i, $i = 0,\dots,x$. We bound the number of vertices in T from below by the number of leaves:

$$
\begin{aligned}
|V(B(\mathbf{u}))| &\overset{(1)}{\geq} u_0 \prod_{i=1}^{m}(u_i - 1) \\
&\geq \prod_{i=0}^{x}(\Delta_i - 2)^{|I_i|} \\
&= \prod_{i=0}^{x}(\Delta_i - 2)^{\ell_{i+1}-\ell_i} \\
&= \prod_{i=0}^{x}(\Delta_i - 2)^{\frac{k}{2}(\Delta_{i+1}^{-1}-\Delta_i^{-1})}. \qquad (2)
\end{aligned}
$$

We compare two monotone sequences Δ_i and $f_i := \log^i k/\log^{i+1} k$, $i = 1, \ldots, x$ and define a corresponding index ι.

- Case 1. $\Delta_1 \leq f_1 = \frac{\log k}{\log\log k}$.
  Then the first term in the product of (2) is $(\Delta_0 - 2)^{\frac{k}{2}(\Delta_1^{-1}-\Delta_0^{-1})} \geq (k - 3)^{\frac{k}{2}\left(\frac{\log\log k}{\log k}-\frac{1}{k-1}\right)} = 2^{\frac{k}{2}\log\log k(1+o(1))}$. Set $\iota := 0$.
- Case 2. $\Delta_x \geq f_x = \frac{\log^x k}{\log^{x+1} k}$.
  Then the last term in the product of (2) is $(\Delta_x - 2)^{\frac{k}{4}-1-\frac{k}{2}\Delta_x^{-1}} \geq \left(\frac{\log^x k}{\log^{x+1} k} - 2\right)^{\frac{k}{4}-1-\frac{k}{2}\frac{\log^{x+1} k}{\log^x k}} = 2^{(\log^{x+1} k-\log^{x+2} k)(\frac{k}{4}-\frac{k}{2}\frac{\log^{x+1} k}{\log^x k})(1+o(1))} = 2^{\frac{k}{4}\log^{x+1} k(1+o(1))}$.
  Set $\iota := x$.
- Case 3. There is some i, $1 \leq i \leq x-1$ with $\Delta_i \geq f_i$ and $\Delta_{i+1} \leq f_{i+1}$.
  In this case we bound the i^{th} term in the product of (2): $(\Delta_i - 2)^{\frac{k}{2}(\Delta_{i+1}^{-1}-\Delta_i^{-1})} \geq \left(\frac{\log^i k}{\log^{i+1} k} - 2\right)^{\frac{k}{2}\left(\frac{\log^{i+2} k}{\log^{i+1} k}-\frac{\log^{i+1} k}{\log^i k}\right)} = 2^{\frac{k}{2}\log^{i+2} k(1+o(1))}$. Set $\iota := i$.

So, we not only bound the number of vertices in T, thereby showing the lower bound for spiders, but more specifically show the following fact that we will need for the general lower bound:

$$
\exists \iota \in \{0, 1, \ldots, x\} \qquad (\Delta_\iota - 2)^{|I_\iota|} \geq 2^{\frac{k}{4}\log^{x+1} k(1+o(1))} \geq 2^{\frac{k}{4}\log^{p+1} k(1+o(1))}. \qquad (3)
$$

□

3.6 Construction of a tree avoiding a general family of trees

Before beginning the next construction, we need some more definitions and lemmas:

For a tree Q that is not a spider, let the *span* of Q, denoted span(Q), be the set of distances between vertices of degree at least 3.

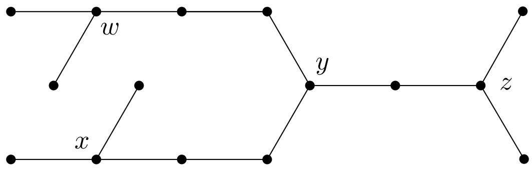


Figure 5: A non-spider whose span is $\{2, 3, 5, 6\}$. The distances come from the vertex pairs (y, z), (x, y), (x, y) and (w, x) respectively.

Observe that if $Q \subseteq T$, then $\mathrm{span}(Q) \subseteq \mathrm{span}(T)$.

We will use this property to avoid non-spiders. We trim the previously constructed tree avoiding the spiders in such a way that we can ensure that for each non-spider Q in the family, some element in $\mathrm{span}(Q)$ is not in $\mathrm{span}(T)$.

So we need a way to trim the tree without reducing its size too much. To do this, we map the property of a (balanced) tree not having a certain element d in its span to binary sequences in some sense. We then define an operation of a binary sequence on a balanced tree (or more exactly, on a vector $\mathbf{u}$ defining a balanced tree $B(\mathbf{u})$), which ensures that if the binary sequence avoids the value d, then the tree will not contain d in its span after the operation.

For each value d, we find some binary sequence avoiding d which does not affect the size of the tree being operated on by too much. To avoid multiple values simultaneously, we find a way to combine binary sequences while still not affecting the size of the tree too much. This eventually allows us to avoid all the non-spiders by avoiding an element of its span for each of them.

Let's shape these roughly outlined ideas into definitions:

Let $\mathbf{b}$ be a binary sequence, and D a set of positive integers. We say that $\mathbf{b}$ *avoids* D if we have $|y - x| + |z - y| \notin D$ for any three indices x, y, z such that $b_x = b_y = b_z = 1$ (here x, y, z do not need to be distinct). If a set D consists just of one element d, instead of writing that $\mathbf{b}$ avoids $\{d\}$ we simply write that $\mathbf{b}$ avoids d. We define the *relative frequency* of `1`s of a binary sequence $\mathbf{b}$ in the interval $I = [s, t]$, or *frequency* $\mathrm{freq}(\mathbf{b}, I)$ for short, as the number of `1`s in the sequence $(b_s, b_{s+1}, \ldots, b_t)$, divided by the total amount of integers in the interval, i.e., by $t - s + 1$. For a binary sequence $\mathbf{b}$ and a vector $\mathbf{u}$, define $\mathbf{b} \circ \mathbf{u} := \mathbf{u}'$ as follows: $u'_i = \begin{cases} u_i & \text{if } b_i = 1, \\ 1 & \text{if } b_i = 0 \text{ and } i = 0, \\ 2 & \text{if } b_i = 0 \text{ and } i \neq 0. \end{cases}$

Lemma 3. *If a binary sequence* $\mathbf{b}$ *avoids* d*, then for any vector* $\mathbf{u}$*,* $d \notin \mathrm{span}(B(\mathbf{b} \circ \mathbf{u}))$.

Proof. Let $B = B(\mathbf{b} \circ \mathbf{u})$. Assume $d \in \mathrm{span}(B)$. Then there is a path P of length d in B with its endpoints x' and z' of degree at least 3. Let y' be the vertex in this path closest to the root of B. If y' is the root, then it must have degree at least 2. If y' is not the root, y' must have degree at least 3, as it is either x' or z', or has two vertices in P

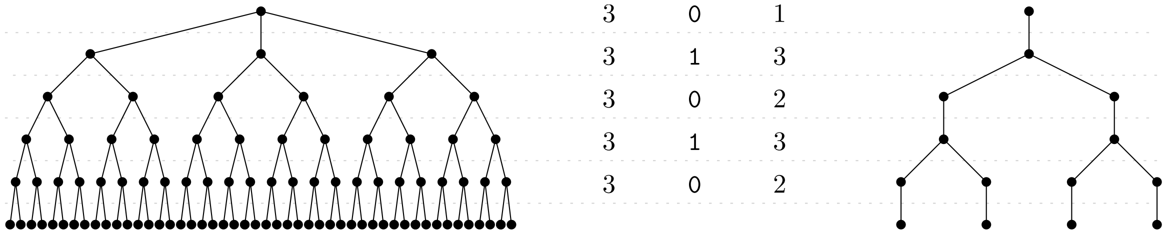


Figure 6: A tree $B(\mathbf{u})$, a sequence $\mathbf{u}$, a binary sequence $\mathbf{b}$ avoiding the value 3, a sequence $\mathbf{b} \circ \mathbf{u}$, and a tree $B(\mathbf{b} \circ \mathbf{u})$, where the entries of the sequences correspond to the layers of the trees depicted on the same horizontal line. Here $\mathbf{u} = (3, 3, 3, 3, 3)$, $\mathbf{b} = (\mathtt{0}, \mathtt{1}, \mathtt{0}, \mathtt{1}, \mathtt{0})$, $\mathbf{b} \circ \mathbf{u} = (1, 3, 2, 3, 2)$.

adjacent to it. Let x, y and z be the distances of x', y' and z' from the root respectively. It holds that $|x - y| + |y - z| = |x - z| = d$. However due to the degree of x', y' and z' we must have $b(x) = b(y) = b(z) = 1$, showing that b doesn't avoid d. □

The frequency of a binary sequence gives us a sense of how much the size of tree is be affected under the operation of the binary sequence. Now that we have the means of avoiding non-spiders through their span, we show that we can do so without reducing the size of the tree being operated on too much.

Lemma 4. *For any finite set of positive integers D, and any interval I, there is a binary sequence $\mathbf{b}$ avoiding D, with a frequency at least $4^{-|D|}$ in I.*

Proof. For two binary sequences $\mathbf{b}$ and $\mathbf{b}'$, let $\mathbf{b} \otimes \mathbf{b}'$ be the binary sequence whose i^{th} element is the product of the i^{th}s elements in $\mathbf{b}$ and $\mathbf{b}'$, i.e., element-wise logical "and". For a positive integer s, a shift, $\mathbf{b}^s$ of a sequence $\mathbf{b}$ is defined by $b^s_x := b_{x+s}$, $x = 0, 1, \ldots$. For a periodic binary sequence $\mathbf{b}$ with period p, we define the frequency $\mathrm{freq}(\mathbf{b})$ as $\mathrm{freq}(\mathbf{b}) := \mathrm{freq}(\mathbf{b}, [0, p - 1])$.

Recall that for a set of positive integers D, a binary sequence $\mathbf{b}$ avoids D if there are no three indices x, y, z with $b_x = b_y = b_z = \mathtt{1}$ such that $|x - y| + |y - z| = d$ for some $d \in D$. Note that if $\mathbf{b}$ avoids D, then a shift $\mathbf{b}^s$ avoids D as well for any s; if furthermore $\mathbf{b}'$ avoids D', then $\mathbf{b} \otimes \mathbf{b}'$ avoids $D \cup D'$.

First we shall prove the following claim by induction on m.

Claim 1. Let $\mathbf{b}_i$ be periodic binary sequences with period p_i and relative frequency $f_i = \mathrm{freq}(\mathbf{b}_i)$, $i = 1, \ldots, m$. Then for every interval I, there are shift values $s_i \in [0, p_i - 1]$, such that $\mathbf{b} = \mathbf{b}_1^{s_1} \otimes \cdots \otimes \mathbf{b}_m^{s_m}$ has frequency $f := \mathrm{freq}(\mathbf{b}, I) \geq \prod_{i=1}^{m} f_i$.

Let l be the number of integers in the interval.
Induction step: Let $\mathbf{b}'$ be the sequence obtained for $\mathbf{b}_1, \ldots, \mathbf{b}_m$, with the period p' and a frequency $f' = \mathrm{freq}(\mathbf{b}', I) \geq \prod_{i=1}^{m-1} f_i$ in the interval I. We have that $\mathbf{b}'$ has $f' \cdot l$ entries $\mathtt{1}$ in the interval I, and $\mathbf{b}_m$ has $f_m \cdot p_m$ entries $\mathtt{1}$ in any interval of length p_m.

For each `1` of $\mathbf{b}'$ in I, there are $f_m \cdot p_m$ shifts $s \in [0, p_m - 1]$ such that the `1` from $\mathbf{b}'$ gets matched up with a `1` from $\mathbf{b}_m$. Summing up the amount of `1`s that $\mathbf{b}' \otimes \mathbf{b}_m^s$ has in I over all shifts $s \in [0, p_m - 1]$, we get $f_m \cdot p_m \cdot f' \cdot l$ as the number of `1`s in total. That means that on average over all s, there are $f_m \cdot f' \cdot l$ `1`s. So there is at least one shift value s_m such that the interval contains at least $f_m \cdot f' \cdot l$ entries of `1` for $\mathbf{b} := \mathbf{b}' \otimes \mathbf{b}_m^{s_m}$. As the length of the interval is l, this means that $\mathrm{freq}(\mathbf{b}, I)$ is at least $f_m \cdot f' \geq \prod_{i=1}^m f_i$.

Induction base: To get the statement for $m = 1$, apply the induction step to the sequence $\mathbf{b}'$ consisting only of `1`s, and note that $\mathbf{b}_1 = \mathbf{b}_1 \otimes \mathbf{b}'$. For this sequence $\mathbf{b}'$, the statement is obviously true. This proves the Claim 1.

Claim 2. For every d, there is a periodic binary sequence $\mathbf{b}$ avoiding d with frequency at least $\frac{1}{4}$.

For $d = 1, 3, 5$, the repeated sequence of `10` serves the purpose, as any two `1`s have a distance divisible by 2. For $d = 2, 4$, the repeated sequence of `100` serves the purpose, as any two `1`s have a distance divisible by 3. For $d \geq 6$, consider a periodic binary sequence $\mathbf{b}$ formed by a block of $\lfloor \frac{d-1}{2} \rfloor$ `1`s followed by d `0`s. We show that $\mathbf{b}$ avoids d by showing that the contrary is false: Assume that there is a set of indices x, y, z such that $b_x = b_y = b_z = 1$ and $|x - y| + |z - y| = d$. Then we see that all three x, y and z must correspond to the same block of 1s. However, the difference in indices within such a block is at most $\lfloor \frac{d-1}{2} \rfloor$, so $|x - y| + |z - y| \leq 2\lfloor \frac{d-1}{2} \rfloor < d$. This concludes the proof of Claim 2.

The lemma now follows from these two claims. □

We now have all the means to construct the tree to avoid a family containing spiders and non-spiders, and to analyze its size.

Construction 2 (Construction of a tree avoiding a general family of trees)**.** *Let* $\mathcal{T} = \mathcal{Q}_s \cup \mathcal{Q}_n$, $|\mathcal{Q}_s| = p+2$, $|\mathcal{Q}_n| = q$ *be a given family of* k*-vertex trees, where* $\mathcal{Q}_s$ *is a family of spiders containing* P_k *and* S_k *and* $\mathcal{Q}_n$ *is a family of non-spiders. Let* $T_s = T_s(\mathcal{Q}_s)$ *be the tree from Construction 1, i.e., a tree avoiding* $\mathcal{Q}_s$*. We have that* T_s *is a balanced tree* $T = B(\mathbf{u})$*, for some vector* $\mathbf{u} = (u_0, \ldots, u_{k/4-2})$*. We shall construct a tree avoiding* $\mathcal{T}$ *by trimming* T_s *in such a way that its span avoids some element of the span of each non-spider in* $\mathcal{T}$*. For that, we need parameters* ι, D *and* $\mathbf{b}$*.*

- *Choose* ι *from* $i = 0, \ldots, x$ *as the index for which the product of elements in the interval* I_i*, i.e.,* $\prod_{j \in I_i}(u_j - 1) = (\Delta_i - 2)^{\ell_{i+1} - \ell_i}$*, is maximal.*
- *Let* D *be a set of representatives of spans of the trees from* $\mathcal{Q}_n$*, i.e.,* $|D \cap \mathrm{span}(Q)| \geq 1$ *for each* $Q \in \mathcal{Q}_n$*, and* $|D| \leq |\mathcal{Q}_n| = q$*.*
- *Let* $\mathbf{b}$ *be a binary sequence avoiding* D *with frequency at least* $4^{-|D|}$ *in* I_ι*, guaranteed by Lemma 4.*

Finally, let our desired tree be

$$T = T(\mathcal{Q}_s \cup \mathcal{Q}_n) := B(\mathbf{b} \circ \mathbf{u}).$$

Theorem 1 (Part 3)**.** *The constructed tree T avoids all trees of the family $\mathcal{T}$ and has size at least $2^{\frac{k}{4}\log^{p+1}k(1+o(1))4^{-q}}$.*

Proof. T avoids all trees of the family:

By Lemma 3, $\text{span}(T) \cap D = \emptyset$. With that, we get that for any non-spider $Q \in \mathcal{Q}_n$ there is a $d \in \text{span}(Q)$ with $d \notin \text{span}(T)$, and therefore T avoids all $Q \in \mathcal{Q}_n$. Since T is a subtree of T_s, and T_s avoids the spiders $\mathcal{Q}_s$, we have that T avoids $\mathcal{Q}_s$ too.

Analysis of the size of T:

Let $\mathbf{u}' := \mathbf{b} \circ \mathbf{u}$. Split up I_ι into I' and I'', where $I' = \{i \in I_\iota : b_i = \mathtt{1}\}$ and $I'' = \{i \in I_\iota \ : b_i = \mathtt{0}\}$. From Lemma 4 we have that $\text{freq}(\mathbf{b}, I) \geq 4^{-|D|} \geq 4^{-q}$, so $|I'| \geq |I_\iota| 4^{-q}$. It follows from (3) that $(\Delta_\iota - 2)^{|I_\iota|} \geq 2^{\frac{k}{4}\log^{p+1}k(1+o(1))}$. Using this information, we get the following lower bound (note that $0 \notin I_\iota$ ensuring that the second product is not 0):

$$\begin{aligned}
|V(B(\mathbf{u}'))| &\overset{(1)}{\geq} u'_0 \prod_{i=1}^{m}(u'_i - 1) \\
&\geq \prod_{i \in I_\iota}(u'_i - 1) \\
&= \prod_{i \in I'}(\Delta_\iota - 2) \prod_{i \in I''} 1 \\
&= (\Delta_\iota - 2)^{|I'|} \\
&\geq \left((\Delta_\iota - 2)^{|I_\iota|}\right)^{4^{-q}} \\
&\geq 2^{\frac{k}{4}\log^{p+1}k(1+o(1))4^{-q}}.
\end{aligned}$$

□

4 Fork-forests in bi-colored complete bipartite graphs

This section is concerned with global unavoidable substructures in balanced 2-colorings of $K_{n,n}$. Before proving the main theorem concerning fork forests, we need to introduce some required terminology.

After that a classic theorem about bipartite graphs, Kőnig's theorem, is presented. The proofs in this part are based on the ideas presented in [48]. Kőnig's theorem is very useful in proving the results in this section.

We then proceed to consider a simplified version of the main problem, presenting techniques that are used later in the proof of the main theorem of this section.

The proof of the main theorem concludes this section. This part is based on [5]. We first prove how many forks it is always possible to find, and then reduce the problem of finding largest fork forests to a problem of finding perfect matchings of minimum weight in edge-weighted graphs. With a known algorithm for the latter problem, our main theorem, Theorem 2, follows.

4.1 Definitions

Recall that a matching is a vertex disjoint forest whose components are K_2, i.e., two vertices linked a by a single edge.

A *maximum cardinality matching* M in a graph G, or *maximum matching* for short, is a matching in G such that no other matching in G has more edges than M. A *perfect matching* is a matching that contains all vertices of G.

A *vertex cover* S of G is a set of vertices such that every edge of G is incident to some vertex in S.

For a graph G and a matching M, an *M-augmenting path* P is a path in G alternatingly using edges from $G - M$ and M, beginning and ending in an edge not in M.

For any sketches, edges colored in "white" are drawn in light gray.

Again, for a bipartite graph the partite sets are implicitely X and Y unless otherwise stated, and in all calculations we omit floors and ceilings when their usage is clear from the context.

4.2 Kőnig's theorem

Before starting with the proof to Kőnig's theorem, we show a lemma giving a necessary and sufficient condition for a matching to be of maximum cardinality.

Lemma 5. *Let G be a graph and M a matching. M is a maximum matching in G if and only if there are no M-augmenting paths in G.*

Proof. If there is an M-augmenting path P in G, then substituting the edges along P that belong to M with the edges that do not belong to M gives a matching with one more edge.

Now assume there is a matching M' with more edges than M. Consider the graph $F \subseteq G$ whose edge set contains exactly the edges of G that are in M or in M', but not in both. As each vertex of G can only have one matching edge of M and M' respectively incident to it, the maximum degree of F is at most 2. Therefore the components of F consist of cycles and paths. The cycles are of even length, as they must alternate between edges of M and M' in G. Because $|M'| > |M|$, one of the paths must start and end with edges from M' in G. Therefore this path is an M-augmenting path. □

Theorem 22 (Kőnig's theorem)**.** *Given a bipartite graph, the size of a maximum matching is equal to the size of a minimum vertex cover.*

Proof. Obviously, a vertex cover has to have to least the size of the matching, as it has to cover all of the matching edges, which have no incident vertices in common.

It remains to show that we can always find a vertex cover of the size of the maximum matching.

Let M be a matching. We show that we can either find a matching of larger size, or a vertex cover of the same size as M. Let S_0 be those vertices in X that are not incident to a matching edge.

Iterate the following loop:

Let T_i be the set of vertices that can be reached from S_i via an edge, but that are not contained in any T_j for $j < i$ yet. If some vertex in $t \in T_i$ has no matching edge incident to it, then we have found an augmenting path. t was reached by a non-matching edge from a vertex $s \in S_i$, which was reached by a matching edge from a vertex in T_{i-1}, and so on. The augmenting path ends in a non-matching edge leading into a vertex from S_0. The matching edges along the path can be substituted as in Lemma 5 to get a larger matching.

If every vertex in T_i has a matching edge incident to it, let S_{i+1} be the set of vertices that can be reached from T_i via edges of the matching. Continue the loop increasing i 1 in each step, until we have found an augmenting path or T_i is empty, i.e., no new vertices in Y can be reached from S_i.

If at some point T_i is empty, we shall show that $T \cup (X - S)$ is a vertex cover of size $||M||$, where S is the union of the S_i and T is the union of the T_i.

To show that $T \cup (X - S)$ is a vertex cover, we need to show that there are no edges between $Y - T$ and S. Assume there is an edge st with $s \in S_i$ for some i and $t \in Y - T$. If it were a non-matching edge, then t would have been put into T_i. If it were a matching edge, then $s \notin S_0$, and t is the vertex in S_{i-1} from which s was reached via some matching edge.

Each $t \in T$ is matched via some edge of M to a vertex in S. Now take some matching edge that is not incident to a vertex in T. It must be incident to some vertex in $X - S$, as S_0 contains the vertices not incident to a matching edge, and the rest of S those vertices that are linked to a vertex in T by a matching edge. On the other hand, every vertex of $X - S$ is incident to a matching edge, as otherwise it would be in S_0, but not matched to a vertex in T. This shows that each matching edge is incident to exactly one vertex in $T \cup (X - S)$, and vice versa, showing that $|T \cup (X - S)| = ||M||$.

□

Note that the proof from above also provides us with an algorithm to find the maximum matching and minimum vertex cover, by starting with an empty matching and repeatedly applying the algorithm to find a larger matching. As the size of the maximum matching is at most $n/2$, and as the algorithm above considers each vertex only once, and therefore each edge only twice, we can find the maximum matching in running time $O(nm)$, where n is the number of vertices and m the number of edges.

Faster algorithms for finding the maximum matching are known, the Hopcroft-Karp algorithm runs in $O(\sqrt{n}m)$ [34], another algoritm was found by Alt et al. running in $O(n^{1.5}\sqrt{m/\log n})$ [4].

4.3 A related problem

Theorem 16 computed the minimum number i needed such that any c-coloring of $K_{i,i}$ forces a copy of m_jK_2, i.e., a matching of size m_j, in color j for some $1 \leq j \leq c$. In particular, for $c = 2$ matchings of size k, it showed that this number is $2k - 1$. This implies that in any 2-coloring of the complete bipartite graph with partite sets of size n, we find a monochromatic matching of size $n/2$.

If we require in addition that the coloring has to be balanced, we now show that we can find a matching of the same size $n/2$ for both colors.

Recall that a balanced coloring is a coloring where each color is used almost the same number of times, that is, the number of edges of one color differs from the number of edges of another color by at most one.

Proposition 23. *Given a balanced black-white coloring of the complete bipartite graph $G = K_{n,n}$, there is always a black matching of size $n/2$. There is some balanced coloring for which no larger matching exists.*

Proof. To find a coloring with no matching on more than $n/2$ edges in either color, take half of the vertices in X, and color all edges incident to them in black. For the remaining $n/2$ vertices of X, color all incident edges in white. As the first half of the vertices in X are a vertex cover for the black edges, the maximum matching has size at most $n/2$.

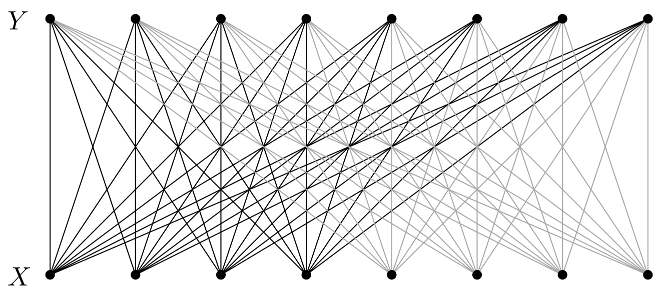


Figure 7: A coloring with no monochromatic matching of size greater than $n/2$.

For the lower bound consider a balanced coloring of the edges of $K_{n,n}$ with partite sets X and Y in black and white. Let G_1 be the graph formed by the black edges, and let G_2 be such a graph formed by the white edges. Let M be a maximum matching of G_1. By Kőnig's theorem applied to G_1, there is a vertex cover S of G_1 such that $|S| = |M|$. We have that $S \subseteq V(M)$. Let $A = V(M) \cap X$, $B = V(M) \cap Y$, $A' = A \cap S$, $B' = B \cap S$, $A'' = A - A'$ and $B'' = B - B'$. Note that $|A'| = |B''|$ and $|A''| = |B'|$. Then we see that the vertex set $(X - A') \cup (Y - B')$ induces no edges in G_1, as otherwise S would not be a vertex cover.

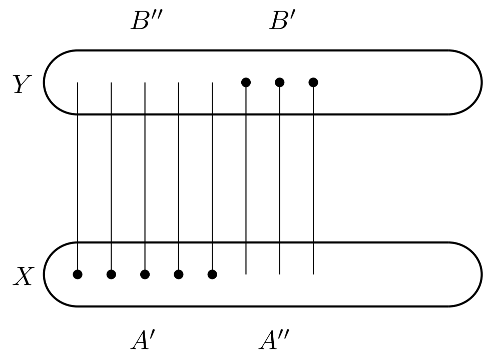


Figure 8: The matching edges M and vertex cover S of G_1, and the labelling of vertex sets introduced above. $A = A' \cup A''$, $B = A' \cup A''$.

As none of the edges between $X - A$ and $Y - B$ are incident to any of the vertex cover edges, they all must be white. Assume for contradiction that the matching has fewer than $n/2$ edges, and therefore $|X - B| > n/2$. Counting the white edges, we get

$$\begin{aligned} |X - A'| \cdot |Y - B'| &= (|A''| + |X - A|)(|X - B'|) \\ &> (|B'| + n/2)(n - |B'|) \\ &= n^2/2 - |B'|(n/2 - |B'|) \\ &\geq n^2/2 \end{aligned}$$

in contradiction to the balanced coloring, which requires that exactly $n^2/2$ edges in G are white.

□

4.4 Main theorem

While we have seen that we can find a both a black and a white matching of size $n/2$ each in a balanced black-white coloring of $K_{n,n}$, if we require the black and white edges to be paired up into structures called forks, we now show that we cannot guarantee as many such forks.

Let $G = K_{n,n}$ with partite sets X and Y be edge colored with two colors.
For a two-coloring c of $E(G)$ we call a path on 3 vertices whose central vertex is in X (or Y) and which has a edges of two different colors a fork centered in X (or Y).

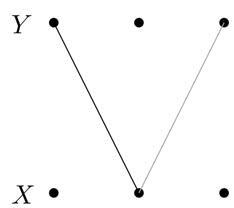


Figure 9: A fork centered in X. (The other edges of $G = K_{n,n}$ have been left out.)

A set of vertex-disjoint forks all centered in X (or Y) is called a *fork forest* centered at X (or Y). The number of forks in a fork forest F is the *size of a forest*, denoted $|F|$. For a coloring c of G let $f(G, c)$ be the largest size of a fork forest centered either at X or at Y. Finally, let $f(n)$ be the minimum $f(G, c)$ taken over all balanced colorings c using two colors. The first part of our main result of this section is

Theorem 2 (Part 1)**.** *For any $n > 1$, $f(n) = (1 - \frac{1}{\sqrt{2}})n$.*

Proof. For the upper bound, take G to be a two-colored $K_{n,n}$ with edges of one color forming a graph isomorphic to $K_{\frac{n}{\sqrt{2}},\frac{n}{\sqrt{2}}}$. This coloring is balanced because $\frac{n}{\sqrt{2}} \cdot \frac{n}{\sqrt{2}} = \frac{n^2}{2}$ edges are black.

Only $n/\sqrt{2}$ of the vertices in X (or Y respectively) have both white and black edges incident to them, so only these are possible centers for biforks. All the white edges incident to them have the same $(1-1/\sqrt{2})n$ common endpoints, so after picking $(1-1/\sqrt{2})n$ biforks, no more vertices remain that are linked to the possible bifork centers by a white edge.

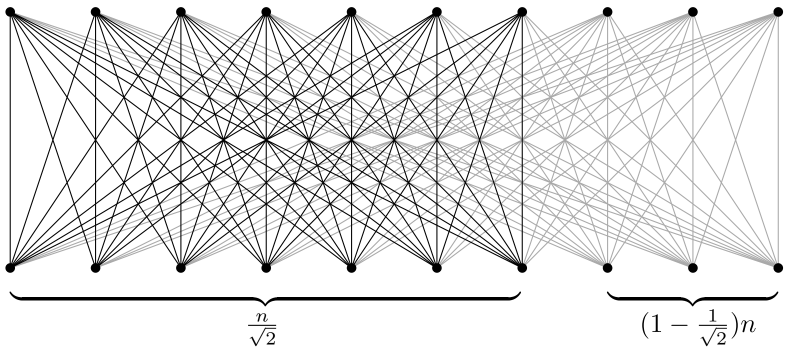


Figure 10: Schematic sketch of a coloring admitting no more than $(1 - 1/\sqrt{2})n$ biforks. Note that for most n we have that $\lfloor \frac{n}{\sqrt{2}} \rfloor \cdot \lfloor \frac{n}{\sqrt{2}} \rfloor$ is slightly less than $\lfloor \frac{n^2}{2} \rfloor$, so a few of the white edges have to be black for the coloring to be exactly balanced, allowing for one additional bifork.

For the lower bound consider a balanced coloring of edges of $K_{n,n}$ with partite sets X and Y in black and white. Let G_1 be the graph formed by the black edges, and let G_2 be such a graph formed by the white edges. Let M be a maximum matching of G_1, and S the vertex cover given by Kőnig's theorem. Define A, A', A'', B, B', B'' as in Proposition 23 and Figure 8.

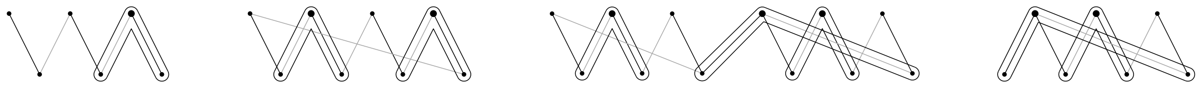
Assume, without loss of generality, that $|A'| \geq |B'|$.

Case 1: $|A'| \leq \frac{n}{\sqrt{2}}$.
We have that $\frac{n^2}{2} = |E(G_1)| \leq n|A'|+(n-|A'|)|B'| \leq n|A'|+(n-|A'|)|A'| = 2n|A'|-|A'|^2$. So, from this we have that $|A'| \geq (1 - \frac{1}{\sqrt{2}})n$. Since $|X - A'| \geq (1 - \frac{1}{\sqrt{2}})n$, there is a fork forest centered at B'', using edges of M and edges of $G_2[B'', X - A']$ with

$\min\{|B''|, |X - A'|\} \geq (1 - \frac{1}{\sqrt{2}})n$ forks.

Case 2: $|A'| > \frac{n}{\sqrt{2}}$.
Let $|A'| = \frac{n}{\sqrt{2}} + c$ for some positive c. We can assume that there is a matching M' in G_2 of size at least $\frac{n}{\sqrt{2}}$, as otherwise Case 1 applies for G_2. By counting, we can observe that at least $x := |M'| - |Y - B''| - |X - A'| \geq \frac{n}{\sqrt{2}} - 2(n - \frac{n}{\sqrt{2}} - c) = n(\frac{3}{\sqrt{2}} - 2) + 2c$ edges of M' have both endpoints in $A' \cup B''$. In the next two paragraphs we show that there are at least $\frac{x}{2}$ forks between B'' and A' centered at B'':

Consider the union $G' = (M \cup M')[A' \cup B'']$, i.e., the black and white matching edges with one endpoint in A' and the other in B''. There are x edges on M' in this graph and each component is either an iterated even cycle or a path ending with edges of M. It is easy to see that one could choose at least $\frac{k}{2}$ forks centered at B'' from a component of G' containing k edges of M' that is either a path or a cycle of length divisible by 4. We also observe that one can choose $\frac{1}{2}(k_1 + k_2)$ forks centered at B'' from two cycles of G' with k_1 and k_2 edges of M', where k_1 and k_2 are odd, by using a single edge between these cycles and additional edges from the cycles.



(a) P_5 (b) C_8 (c) pair of C_6 (d) remaining C_6

Figure 11: Finding forks in the components of G', with one example for each type. White edges are drawn in light gray, the vertices belonging to B'' are positioned at the top. The outlined edges have been chosen to be used in forks. In the latter two cases this choice depends on the color of the single edge not contained in the cycles.

So, we can pair up all but at most one of the components of G' that are cycles of length 2 modulo 4. In the remaining such component with k edges of M' we can choose $\frac{k+1}{2}$ forks centered at B'' by using one additional edge going from the component into a previously unused vertex in A' if available. If not, then all vertices in A' have been used up from the previously chosen forks, so we already have got $\lfloor \frac{1}{2}|A'| \rfloor \geq \lfloor \frac{n}{2\sqrt{2}} \rfloor$ forks. By combining the selected forks, we see that there are at least $\frac{x}{2}$ forks centered in B'' and having leaves in A'.

We observe that with each chosen fork, at most two matching edges of M have become unavailable for later use, so there are at least $|B''| - x$ black matching edges with both endpoints in $A' \cup B''$ remaining. These can be combined into forks centered at B'' with non-edges leading into $X - A'$. This results in a total of $\frac{x}{2} + \min\{|B''| - x, |X - A'|\}$ forks. Since

$$\begin{aligned}\frac{x}{2} + \min\{|B''| - x, |X - A'|\} &\geq n(\frac{3}{2\sqrt{2}} - 1) + c + \min\{n(2 - \frac{2}{\sqrt{2}}) - c, n(1 - \frac{1}{\sqrt{2}}) - c\} \\ &= \min\{n(2 - \frac{2}{\sqrt{2}} + \frac{3}{2\sqrt{2}} - 1), n(1 - \frac{1}{\sqrt{2}} + \frac{3}{2\sqrt{2}} - 1)\} \\ &= \min\{n(1 - \frac{1}{2\sqrt{2}}), n(\frac{1}{2\sqrt{2}})\} = \frac{n}{2\sqrt{2}} \geq n(1 - \frac{1}{\sqrt{2}}),\end{aligned}$$

it follows that $f(G, c) \geq n(1 - \frac{1}{\sqrt{2}})$.

□

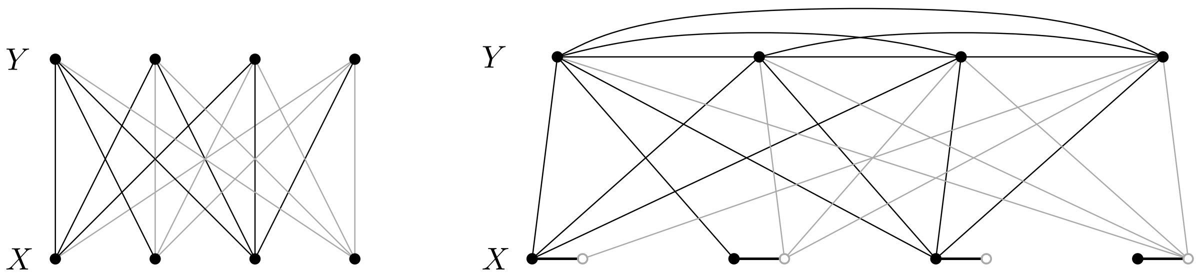
4.5 Algorithm for the main theorem

The second part of the theorem is

Theorem 2 (Part 2)**.** *There is an algorithm finding a largest fork forest centered in X in any two-colored complete bipartite graph with partite sets X and Y and running in time $O(n^2 \log n \sqrt{n\alpha(n^2, n) \log n})$.*

We show that there is an efficient algorithm for finding the largest fork forest centered at X in G by reducing this problem to the problem of finding a perfect matching of minimum weight in an edge-weighted graph G'. The case of a fork forest centered at Y is symmetric.

Informally, G' is obtained from G by first splitting each vertex of X into two adjacent vertices, with one of them being assigned the black edges incident to the original vertex, and the other taking the white edges. Then all edges in Y are added, and if n is odd, one additional vertex is added adjacent to all vertices of Y.



(a) original graph G (b) transformed graph G'

Figure 12: A coloring of $G = K_{4,4}$ with white edges drawn in light gray, and its transformed version G' on the right. For $x \in X$, vertices $x_b \in G'$ with the black edges incident to them are drawn in black, while $x_w \in G'$ with white incident edges are drawn in light gray.

Construction
For a $\{b, w\}$-coloring, c, of $G = K_{n,n}$ with partite sets X and Y, let $V(G')$ be a disjoint union $Y' \cup \{x_b : x \in X\} \cup \{x_w : x \in X\}$, where $Y' = Y$ if n is even and $Y' = Y \cup \{y\}$ if n is odd. Let $E(G')$ be the union of $\{x_b x_w : x \in X\}$, $\{yx_b : c(yx) = b, x \in X, y \in Y\}$, $\{yx_w : c(yx) = w, x \in X, y \in Y\}$, and all possible edges with endpoints in Y'. Let $\tau : E(G') \to \{0, 1\}$ be such that $\tau(x_b x_w) = 1$ for all $x \in X$, and $\tau(e) = 0$, for all other edges.

Further, if M is a perfect matching in G', denote by $\text{fork}(M)$ a fork forest in G containing all forks on vertices x, y, y' if $x_b y \in M, x_w y' \in M$. Recall that $|\text{fork}(M)|$ is the number of forks in $\text{fork}(M)$.

Lemma 6. *If M is a minimum weight perfect matching of (G', τ) then $\text{fork}(M)$ is a maximum fork forest of (G, c) centered at X.*

Proof. Let M be a minimum weight perfect matching of (G', τ). Note that the weight of M is equal to the number of edges $x_b x_w \in E(M)$. We see that $x \notin V(\text{fork}(M))$ if and only if $x_b x_w \in E(M)$, so the weight of M is $n - |\text{fork}(M)|$.

Assume that $\text{fork}(M)$ is not a largest fork forest of (G, c) centered at X. Then, for a larger fork forest F' of (G, c) centered at X, let M' be a perfect matching of G' that contains edges $x_b y$ and $x_w y'$ if x, y, y' induces a fork of F', and edge $x_b x_w$, otherwise. Note that one can always match vertices of Y that are not in F' with remaining vertices of Y'. This matching M' has weight $n - |F'| < n - |\text{fork}(M)|$, a contradiction. □

In [25] it is shown that the time complexity of finding the minimum weight matching in a graph with n vertices, m edges, and edge-weights 0 or 1 is $O(\sqrt{n\alpha(m, n) \log n}m \log n)$, where α denotes the slowly growing inverse of the Ackermann function. Since G' contains at most $3n + 1$ vertices and $\frac{3}{2}(n^2 + n)$ edges, the minimum weight perfect matching problem for (G', w) can be solved in $O(n^2 \log n \sqrt{n\alpha(n^2, n) \log n})$ time. Thus, the second part of the main theorem follows as well. □

5 Rainbow paths

Unlike the classic Ramsey problem for paths, as seen in Section 2, there is an equivalent rainbow Ramsey-like problem using proper colorings that is far from solved. That is, for what minimal $f(P_n) = k$ does any proper coloring of K_k contain a rainbow copy of P_n? Recall that a proper coloring is a coloring where no two incident edges have the same color.

The best known bounds are $n \le f(P_n) \le \lfloor (3n + 1)/2 \rfloor$ with a non-trivial bound $n + 1 \le f(P_n)$ for certain n. Both are not difficult to prove. While these bounds have been known for a while, and an algorithm from [31] can find such paths as a side-effect, no concise proof specifically aimed at this problem has been published.

Proposition 24.

$$f(P_n) \leq \lfloor (3n-1)/2 \rfloor.$$

That is, in each properly colored $K_{\lfloor (3n-1)/2 \rfloor}$ there is a rainbow path on n vertices.

Proof. Let c be a proper coloring of the complete graph $K_{\lfloor (3n-1)/2 \rfloor}$. Let P be the longest rainbow path in this colored graph. Assume it has fewer than n vertices, and define $x > 0$ so that the number of vertices it contains is $|P| = n - x$. Let s be the starting vertex and t the ending vertex of the path. Let S be the set of edges incident to s that do not have an endpoint in $V(P) - s$, and T the set of edges incident to t that do not have an endpoint in $V(P) - t$.

Due to the maximality of P, the sets S and T only contain colors in $c(P)$, i.e., colors that are also used in P. Now consider the set D of edges in P that have a color in $c(S)$. We have that $|T| = |S| = \lfloor (n-1)/2 \rfloor + x$, and thus $|D| = \lfloor (n-1)/2 \rfloor + x$ as well.

Now consider the edges incident to t. Out of the $||P||$ edges from t into P, at most $||P|| - |T|$ have a color that is already used along P, as all edges in T have a color in $c(P)$, and no color incident to t appears twice. So at least $||P|| - (||P|| - |T|) = |T| = \lfloor (n-1)/2 \rfloor + x$ of these have a color that is not used in P. Define this set of edges as F. $|F| \geq |T| = \lfloor (n-1)/2 \rfloor + x$.

Recall that the size of D, which contained the edges along P using colors from $c(S)$, was $\lfloor (n-1)/2 \rfloor + x$. We get that if $x > 0$, then one edge d from D shares its endpoint that is further from t along P with some edge f in F. Replacing d with f, we get a new rainbow path of the same length as P and one endpoint s. Note that f has a different color from d as the coloring is proper. As d used a color from $c(S)$, the color is now available and the path can be extended by the edge incident to S that used that color.

This is a contradiction to the maximality of P. So x must be 0 or less, meaning that a longest rainbow path P contains at least n vertices.

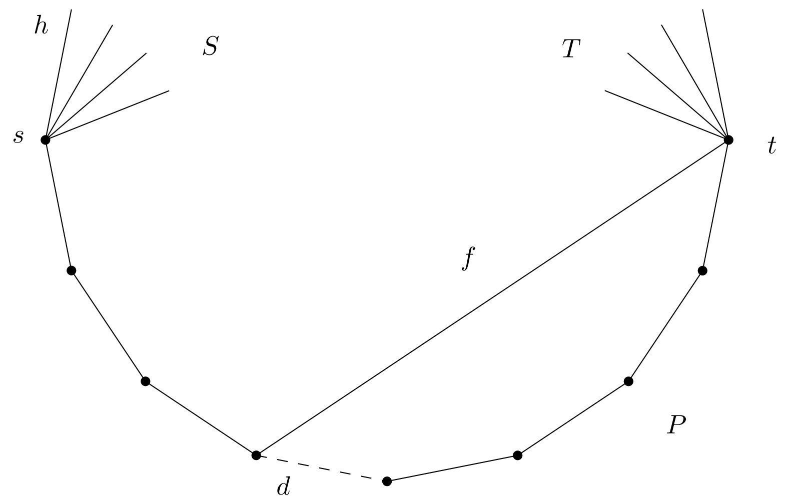


Figure 13: The edge d gets replaced by the edge f. As d and some edge $h \in S$ have the same color, this new path can be extended by h.

□

Note that the proof above gives rise to an algorithm to find a path of at least such length, by iteratively applying the method from the proof to extend the path. As in each step, the edges of the path, and in addition the edges incident to its endpoints need to be considered, the algorithms runs in time $O(n^2)$. While not having impact on the asymptotic runtime, until half of the vertices of the complete graph have been used up, the path can simply be extended using a greedy strategy: As fewer than half of the colors are used, there is an edge incident to one endpoint of a color not yet used.

The lower bound is briefly noted in [31].

Proposition 25. *There are infinitely many n such that*

$$f(P_n) > n,$$

that is, there is a proper coloring of K_n such that there is no rainbow path on n vertices.

Proof. Let $n = 2^k$. Let the vertex set be $V = \{0, \ldots, n-1\}$ and E the set of edges. Let the coloring be $c\colon E \to \{1, \ldots, n-1\}$, $c(uv) = u \oplus v$ where $\oplus$ denotes the binary "xor", the exclusive or. This coloring is called *geometric factorization* and it is a proper coloring.

Assume there is a rainbow path of length n. It uses all available colors. The binary "xor" of all colors is 0. However, as the binary "xor" of a value with itself is 0, the binary "xor" of the colors along the path is the same as the binary "xor" of its endpoints. This value being 0 however means that the endpoints are equal, a contradiction.

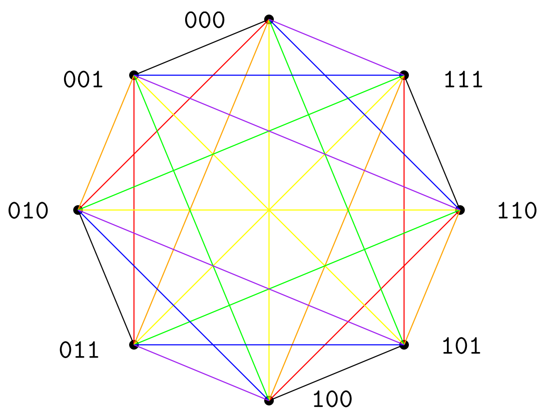


Figure 14: The geometric factorization of K_n for $n = 2^3$.

□

6 Conclusion

With many problem having been covered in the past, and two new problems solved, what remains to be done?

The solution to the problem about unavoidable subtrees gives us the insight that spiders are the determining factor for the size of a tree avoiding subtrees, and that generally we can construct a tree of exponential size to avoid the subtrees. However, what happens if we are more concerned about the general structure of the trees we want to avoid, i.e., if we also want to avoid all subdivisions of the trees from a family? Obviously, this does not have any effect on spiders, but it makes impossible the technique to avoid non-spiders while compromising the size of the constructed tree only marginally. Could non-spiders be the determining factor of the maximum size of a tree to avoid subdivisions of trees from a family now, and would the size of these trees still be exponential?

The problem on fork forests opens up a whole range of possible problems to be examined. Generalizations could ask for other unavoidable patterns to be examined, or more than two colors to be used. Other restriction on the coloring than "balanced" could be imposed, for instance on the number of edges of a color incident to a vertex, or other graphs than complete bipartite graphs as host graphs could be considered.

The problem concerning rainbow paths in properly colored complete graphs still remains unsolved. An interesting question is: Can we find longer paths if we limit the colorings to factorizations, i.e., consider colorings where the color classes are (almost) perfect matchings, meaning that for each color there is at most one vertex not having an edge of that color incident to it? Might the additional structure of factorizations make a difference here, or might it be possible to reduce colorings that are not factorizations to such? Could examining this problem for other trees than paths give us more insight into the problem?

There are also other Ramsey generalizations for which the rainbow path problem could be examined. Or what happens if we restrict the number of edges of one color that may be incident to a vertex not to 1, but some other number?

Evidently, the topic of unavoidable trees and forests in graphs still offers plenty of research opportunities, and there are still many problems left to be solved.

References

[1] A. Abouelaoualim, K. Ch. Das, W. Fernandez de la Vega, M. Karpinski, Y. Manoussakis, C. A. Martinhon, and R. Saad. Cycles and paths in edge-colored graphs with given degrees. *Journal of Graph Theory*, 64(1):63–86, 2010.

[2] M. Ajtai, J. Komlós, M. Simonovits, and E. Szemerédi. Solution of the Erdős-Sós conjecture. in preparation.

[3] N. Alon and G. Gutin. Properly colored hamilton cycles in edge colored complete graphs. *Random Structures and Algorithms*, 11(2):179–186, 1997.

[4] H. Alt, N. Blum, K. Mehlhorn, and M. Paul. Computing a maximum cardinality matching in a bipartite graph in time $o(n^{1.5}\sqrt{m/\log n})$. *Inf. Process. Lett.*, 37(4):237–240, February 1991.

[5] M. Axenovich, M. Krug, G. Osang, and I. Rutter. Fork-forests in bi-colored complete bipartite graphs. submitted, 2012.

[6] M. Axenovich and G. Osang. Unavoidable subtrees. submitted, 2012.

[7] H. Bielak. Multicolor Ramsey numbers for some paths and cycles. *Discuss. Math. Graph Theory*, 29(2):209–218, 2009.

[8] S. A. Burr and J. A. Roberts. On Ramsey numbers for stars. *Utilitas Math.*, 4:217–220, 1973.

[9] G. Chen and R. H. Schelp. Graphs with linearly bounded Ramsey numbers. *J. Combin. Theory Ser. B*, 57(1):138–149, 1993.

[10] F. R. K. Chung and R. L. Graham. Universal caterpillars. *J. Combin. Theory Ser. B*, 31(3):348–355, 1981.

[11] F. R. K. Chung, R. L. Graham, and D. Coppersmith. On trees containing all small trees. In *The theory and applications of graphs (Kalamazoo, Mich., 1980)*, pages 265–272. Wiley, New York, 1981.

[12] E. J. Cockayne and P. J. Lorimer. The Ramsey number for stripes. *J. Austral. Math. Soc.*, 19:252–256, 1975.

[13] D. Conlon. A new upper bound for diagonal Ramsey numbers. *Ann. of Math. (2)*, 170(2):941–960, 2009.

[14] R. Diestel. *Graph Theory.* Springer-Verlag, 2006.

[15] P. Erdős, R. J. Faudree, R. J. Gould, M. S. Jacobson, and J. Lehel. Edge disjoint monochromatic triangles in 2-colored graphs. *Discrete Mathematics*, 231(1–3):135–141, 2001.

[16] P. Erdős and R. Rado. A combinatorial theorem. *J. London Math. Soc.*, 25:249–255, 1950.

[17] P. Erdős and T. Gallai. On maximal paths and circuits of graphs. *Acta Math. Acad. Sci. Hungar*, 10:337–356 (unbound insert), 1959.

[18] P. Erdős, M. Simonovits, and V. T. Sós. Anti-Ramsey theorems. In *Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Vol. II*, pages 633–643. Colloq. Math. Soc. János Bolyai, Vol. 10. North-Holland, Amsterdam, 1975.

[19] P. Erdös and A. H. Stone. On the structure of linear graphs. *Bull. Amer. Math. Soc.*, 52:1087–1091, 1946.

[20] G. Fan and L. Sun. The Erdős-Sós conjecture for spiders. *Discrete Math.*, 307(23):3055–3062, 2007.

[21] R. J. Faudree and R. H. Schelp. Path-path Ramsey-type numbers for the complete bipartite graph. *J. Combinatorial Theory Ser. B*, 19(2):161–173, 1975.

[22] R. J. Faudree and R. H. Schelp. Path Ramsey numbers in multicolorings. *J. Combinatorial Theory Ser. B*, 19(2):150–160, 1975.

[23] R. J. Faudree and R. H. Schelp. Ramsey numbers for all linear forests. *Discrete Math.*, 16(2):149–155, 1976.

[24] S. Fujita, C. Magnant, and K. Ozeki. Rainbow generalizations of Ramsey theory: a survey. *Graphs Combin.*, 26(1):1–30, 2010.

[25] H. N. Gabow and R. E. Tarjan. Faster scaling algorithms for general graph matching problems. *Journal of the ACM*, 38(4):815–853, 1991.

[26] L. Gerencsér and A. Gyárfás. On Ramsey-type problems. *Ann. Univ. Sci. Budapest. Eötvös Sect. Math.*, 10:167–170, 1967.

[27] M. K. Goldberg and È. M. Livšic. Minimal universal trees. *Mat. Zametki*, 4:371–379, 1968.

[28] W. T. Gowers. Lower bounds of tower type for Szemerédi's uniformity lemma. *Geom. Funct. Anal.*, 7(2):322–337, 1997.

[29] J. W. Grossman, F. Harary, and M. Klawe. Generalized Ramsey theory for graphs. X. Double stars. *Discrete Math.*, 28(3):247–254, 1979.

[30] A. Gyárfás and J. Lehel. A Ramsey-type problem in directed and bipartite graphs. *Period. Math. Hungar.*, 3(3-4):299–304, 1973.

[31] A. Gyárfás and M. Mhalla. Rainbow and orthogonal paths in factorizations of K_n. *J. Combin. Des.*, 18(3):167–176, 2010.

[32] A. Gyárfás, M. Ruszinkó, G. N. Sárközy, and E. Szemerédi. Three-color Ramsey numbers for paths. *Combinatorica*, 27(1):35–69, 2007.

[33] P. E. Haxell, T. Łuczak, and P. W. Tingley. Ramsey numbers for trees of small maximum degree. *Combinatorica*, 22(2):287–320, 2002. Special issue: Paul Erdős and his mathematics.

[34] J. E. Hopcroft and R. M. Karp. An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. *SIAM Journal on Computing*, 2(4):225–231, 1973.

[35] R. E. Jamison, T. Jiang, and A. C. H. Ling. Constrained Ramsey numbers of graphs. *J. Graph Theory*, 42(1):1–16, 2003.

[36] R. M. Karp. Reducibility among combinatorial problems. In *Complexity of computer computations (Proc. Sympos., IBM Thomas J. Watson Res. Center, Yorktown Heights, N.Y., 1972)*, pages 85–103. Plenum, New York, 1972.

[37] A. E. Kézdy and C. Wang. Alternating walks in partially 2-edge-colored graphs and optimal strength of graph labeling. *Discrete Mathematics*, 194(1–3):261–265, 1999.

[38] A. McLennan. The Erdős-Sós conjecture for trees of diameter four. *J. Graph Theory*, 49(4):291–301, 2005.

[39] J. Pach, J. Solymosi, and G. Tóth. Unavoidable configurations in complete topological graphs. *Discrete Comput. Geom.*, 30(2):311–320, 2003. U.S.-Hungarian Workshops on Discrete Geometry and Convexity (Budapest, 1999/Auburn, AL, 2000).

[40] R. G. Powers. The ith Ramsey number for matchings. *Discrete Math.*, 61(1):115–118, 1986.

[41] F. P. Ramsey. On a Problem of Formal Logic. *Proc. London Math. Soc.*, S2-30(1):264, 1930.

[42] K. B. Reid and N. C. Wormald. Embedding oriented n-trees in tournaments. *Studia Sci. Math. Hungar.*, 18(2-4):377–387, 1983.

[43] R. Shamir and D. Tsur. Faster subtree isomorphism. *J. Algorithms*, 33(2):267–280, 1999.

[44] M. Simonovits and V. T. Sós. On restricted colourings of K_n. *Combinatorica*, 4(1):101–110, 1984.

[45] J. Spencer. Ramsey's theorem—a new lower bound. *J. Combinatorial Theory Ser. A*, 18:108–115, 1975.

[46] P. Turán. Eine Extremalaufgabe aus der Graphentheorie. *Mat. Fiz. Lapok*, 48:436–452, 1941.

[47] O. Tveretina, C. Sinz, and H. Zantema. An exponential lower bound on OBDD refutations for pigeonhole formulas. In E. Markakis and I. Milis, editors, *Proceedings of the 4th Athens Colloquium on Algorithms and Complexity (ACAC'09)*, pages 13–21, 2009.

[48] D. West. *Introduction to Graph Theory*. Prentice Hall, 2001.

Statutory Declaration

I hereby declare that this thesis is a result of my own work, that I used no other than the indicated references and resources, that all information that has been taken directly or indirectly from other sources is indicated as such, and that I regarded the statute of the Karlsruhe Institute of Technology on securing good scientific practice in its currently applicable version.

Erklärung

Hiermit versichere ich, dass ich diese Arbeit selbständig verfasst und keine anderen, als die angegebenen Quellen und Hilfsmittel benutzt, die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht und die Satzung des Karlsruher Instituts für Technologie zur Sicherung guter wissenschaftlicher Praxis in der jeweils gültigen Fassung beachtet habe.

Karlsruhe, July 19, 2012