Large homogeneous subgraphs in bipartite graphs with forbidden induced subgraphs

Maria Axenovich* Casey Tompkins† Lea Weber‡

Abstract

For a bipartite graph G, let $\tilde{h}(G)$ be the largest t such that either G contains $K_{t,t}$, a complete bipartite subgraph with parts of size t, or the bipartite complement of G contains $K_{t,t}$ as a subgraph. For a class of graphs $\mathcal{F}$, let $\tilde{h}(\mathcal{F}) = \min\{\tilde{h}(G) : G \in \mathcal{F}\}$. We say that a bipartite graph H is strongly acyclic if neither H nor its bipartite complement contains a cycle. By $\mathrm{Forb}(n, H)$ we denote the set of bipartite graphs with parts of size n, which do not contain H as an induced bipartite subgraph respecting the sides. One can easily show that $\tilde{h}(\mathrm{Forb}(n, H)) = O(n^{1-\epsilon})$ for a positive ϵ if H is not strongly acyclic. Here we ask whether $\tilde{h}(\mathrm{Forb}(n, H))$ is linear in n for any strongly acyclic graph H. We answer this question in the positive for all but four strongly acyclic graphs. We do not address this question for the remaining four graphs in this paper.

Introduction

A conjecture of Erdős and Hajnal [5] states that for any graph H, there is a constant $\epsilon > 0$ such that any n-vertex graph that does not contain H as an induced subgraph has either a clique or a coclique on at least n^ϵ vertices. While this conjecture remains open, see for example a survey by Chudnovsky [4], we address a bipartite variation of the problem.

Let G be a bipartite graph with parts U and V of size n each, we write $G = ((U, V), E)$, $E \subseteq U \times V$. We further write $E = E(G)$, and for an edge $(u, v) \in E$, we simply write uv. We shall often depict the sets U and V as sets of points on two horizontal lines in the plane and call U the top part and V the bottom part. We say that a graph is the *bipartite complement* of G if it has the same vertex set as G and its edge set is $(U \times V) \setminus E$. We denote the bipartite complement of a graph G by G'. By $\tilde{\omega}(G)$ we denote the largest integer t such that there are $A \subseteq U$, $B \subseteq V$ with $|A| = |B| = t$ and $ab \in E$ for all $a \in A$, $b \in B$, i.e., A and B form a *biclique*. By $\tilde{\alpha}(G)$ we denote the largest integer t such that there are $A \subseteq U$, $B \subseteq V$ with $|A| = |B| = t$ and $ab \notin E$ for all $a \in A$, $b \in B$, i.e., A and B form a *co-biclique*. Let $\tilde{h}(G) = \max\{\tilde{\alpha}(G), \tilde{\omega}(G)\}$.

For bipartite graphs $H = ((U, V), E)$ and $G = ((A, B), E')$, we say that H is an *induced bipartite subgraph of G respecting sides* if $U \subseteq A$, $V \subseteq B$, and for any $u \in U$, $v \in V$, we have $uv \in E(H)$ if and only if $uv \in E(G)$. We say that a bipartite graph $H = ((U, V), E)$ is a *copy* of a bipartite graph $H^* = ((U^*, V^*), E^*)$ if H^* is isomorphic to H with isomorphism $\varphi : U^* \cup V^* \to U \cup V$ such that $\varphi(U^*) = U$ and $\varphi(V^*) = V$. Let $\mathrm{Forb}(n, H)$ denote the set of all bipartite graphs with parts of size n which do not contain a copy of H as an induced bipartite

*Karlsruhe Institute of Technology, Karlsruhe, Germany, maria.aksenovich@kit.edu.
†Karlsruhe Institute of Technology, Karlsruhe, Germany
and Institute for Basic Science, Daejeon, South Korea, ctompkins496@gmail.com.
‡Karlsruhe Institute of Technology, Karlsruhe, Germany, 0815lea@gmx.net.

subgraph respecting sides. We call a bipartite graph H-free if it does not contain a copy of H. Let

$$\tilde{h}(n,H) = \tilde{h}(\text{Forb}(n,H)) = \min\{\tilde{h}(G) : \; G \in \text{Forb}(n,H)\}.$$

It is implicit from a result of Erdős, Hajnal, and Pach [6] that for any bipartite H with the smaller part of size k, $\tilde{h}(n,H) = \Omega(n^{1/k})$. A standard probabilistic argument shows that if H or its bipartite complement contains a cycle, then $\tilde{h}(n,H) = O(n^{1-\epsilon})$ for a positive ϵ. Here, we address the question of when $\tilde{h}(n,H)$ is linear in n. We say that a bipartite graph H is *strongly acyclic* if neither H nor its bipartite complement contain a cycle. We show that for all but at most four strongly acyclic graphs H, $\tilde{h}(n,H)$ is linear in n. Moreover, for several graphs H we determine $\tilde{h}(n,H)$ exactly.

Theorem 1. *There is a set $\mathcal{H}$ of four graphs with the property that for any strongly acyclic bipartite graph H, such that neither H nor H' is in $\mathcal{H}$, there is a positive constant $c = c(H)$ such that $\tilde{h}(n,H) \geq cn$.*

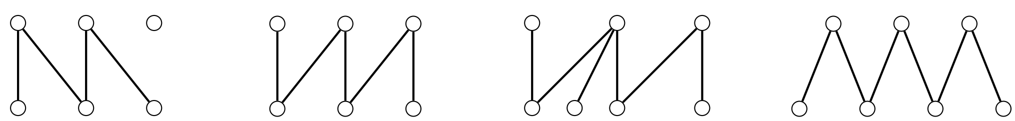
The set $\mathcal{H} = \{\tilde{P}_5, P_6, \tilde{H}_{3,4}, P_7\}$ is given in Figure 1. It is sufficient to take $c(H) = \frac{1}{30|V(H)|}$.

Note that the notion of large bicliques and co-bicliques in ordered bipartite graphs with forbidden induced subgraphs corresponds to the notion of submatrices of all 0's or of all 1's in binary matrices with forbidden submatrices. A paper of Korándi, Pach, and Tomon [9] addresses a similar question for matrices. In addition, one could interpret bipartite graphs as set systems consisting of all the neighborhoods of vertices from one part. Structural properties of these graphs in terms of VC-dimension of the respective set system in connection to the Erdős-Hajnal conjecture are addressed for example by Fox, Pach, and Suk [8].

The paper is structured as follows. In Section 1 we characterize all strongly acyclic bipartite graphs. In Section 2 we find linear lower bounds on $\tilde{h}(n,H)$ for each of the strongly acyclic graphs with few exceptions, thus, proving Theorem 1. In Section 3 we determine the optimal constant c in Theorem 1 exactly for forbidden bipartite graphs with two vertices in each part. In Section 4 we prove some general bounds for completeness. Finally, in Section 5 we discuss some progress which has occurred since the submission of this manuscript.

1 Characterization of all strongly acyclic graphs

In this section we determine all strongly acyclic bipartite graphs up to a bipartite complement.



$\tilde{P}_5$ $\quad$ P_6 $\quad$ $\tilde{H}_{3,4}$ $\quad$ P_7

Figure 1: Strongly acyclic subgraphs with parts of size at least 3.

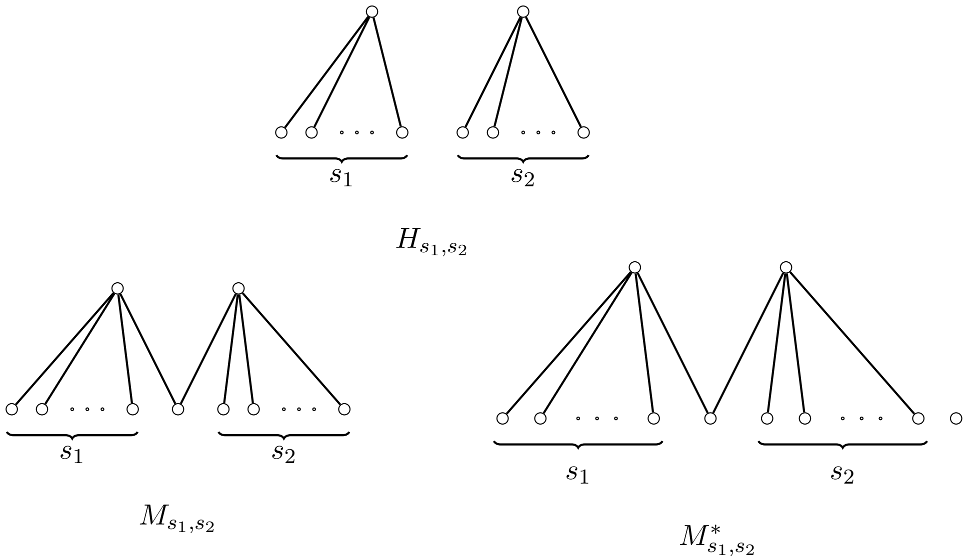


Figure 2: Strongly acyclic bipartite graphs with one part of size 2.

We denote a cycle of length i by C_i, a path on i vertices by P_i, and a complete bipartite graph with parts of size s and t by $K_{s,t}$.

Theorem 2. *Let H be a strongly acyclic bipartite graph. Then one of its parts has size at most* 3 *and either H or its bipartite complement is an induced subgraph of a graph from $\{\tilde{H}_{3,4}, P_7, M^*_{s,s}\}$, for some positive integer s. If H is not a strongly acyclic bipartite graph, then H or its bipartite complement contains C_4, C_6, or C_8.*

Remark: The complete list of all strongly acyclic graphs (up to bipartite complementation) is given in Figures 1 and 2. Note that $\tilde{P}_5$, P_6, H_{s_1,s_2}, and M_{s_1,s_2} are induced subgraphs of $\tilde{H}_{3,4}$, P_7, $M^*_{s,s}$, and $M^*_{s,s}$, respectively, for $s \geq s_1, s_2$.

Proof of Theorem 2. Let H be a bipartite graph with top part $U = \{u_1, \ldots, u_k\}$ and bottom part $V = \{v_1, \ldots, v_l\}$, where $2 \leq k \leq l$. Denote by H' the bipartite complement of H and $d(u)$ the degree of a vertex u in H.

Assume that $k = 2$. Consider u_1, u_2 and their neighborhoods. We see that these neighborhoods share at most one vertex, otherwise we have a cycle of length four. The same holds for the bipartite complement of H. Thus, H is an induced subgraph of $M^*_{s_1,s_2}$, for some s_1, s_2, as shown in Figure 2.

Now let $k \geq 3$. Since H and H' are acyclic, the number of edges in each of H and H' is at most $|U|+|V|-1$, i.e., the total number of edges in these two graphs is at most $2(|U|+|V|-1)$. On the other hand, this number is $|U||V|$. We see however, that if $|U|, |V| \geq 4$, then $|U||V| > 2(|U|+|V|-1)$. Similarly, if $|U| = 3$ and $|V| \geq 5$, we have that $|U||V| > 2(|U|+|V|-1)$. Thus, $k = |U| = 3$ and $|V| \leq 4$.

Let $|U| = 3$ and $|V| = 3$. Assume there is a vertex from U of degree 0, say $d(u_2) = 0$. Then we have $d(u_1), d(u_3) \geq 2$, otherwise there is a C_4 in H'. Moreover we must have $|N(u_1) \cap N(u_3)| \leq 1$. Thus, $H = \tilde{P}_5$. By considering H', we can assume that no vertex in U has degree 3. Thus, all vertices of H have degrees 1 or 2. Since H is strongly acyclic, there are at most 5 and at least 4 edges. So, up to bipartite complementation, we can assume that there are 4 edges in H with respective degrees 1, 1, and 2 in both parts. This is only possible when H is a disjoint union of

K_2 and P_4, whose bipartite complement is P_6.

Let $|U| = 3$ and $|V| = 4$. Assume there is a vertex $u \in U$ with $d(u) = 4$. Then any two other vertices in U each have 3 non-neighbors in $N(u)$, and thus, have at least two common non-neighbors, resulting in C_4 in H', a contradiction. Thus, $d(u) \leq 3$ for any $u \in U$. By considering H', we see that there are no vertices of U of degree 0. I.e., the degrees of vertices from U could be $1, 2$, or 3. Since H is strongly acyclic, the number of edges is at most 6 and at least $12 - 6 = 6$. So, H has 6 edges, and the degrees of vertices in U are $1, 2, 3$ or $2, 2, 2$. In the case where the degrees are $1, 2, 3$, we see that the neighborhoods of the degree 2 and 3 vertices intersect in exactly one vertex. The vertex of degree 1 must be adjacent to a neighbor of a degree 3 vertex which is not adjacent to a degree 2 vertex, otherwise there is a C_4 in the bipartite complement of H. Thus, we have that $H = \tilde{H}_{3,4}$. If the degrees of vertices in U are $2, 2, 2$, then the only option is P_7.

We only need to show that for any bipartite graph H which is not strongly acyclic, either H or its bipartite complement contains C_4, C_6, or C_8. If H has one part of size at most 4, we are done, since any cycle in H or H' has length at most 8. If H has both parts of size at least 5, one can easily verify that either H or H' contains a C_4. $\square$

2 Proof of Theorem 1

Let $H_s = H_{s,s}$, $M_s = M_{s,s}$, and $M_s^* = M_{s,s}^*$. In the following results we use the fact that if H is an induced bipartite subgraph of K respecting sides, then $\tilde{h}(n, H) \geq \tilde{h}(n, K)$. We omit ceilings and floors where they are not crucial. The essential part of the proof of Theorem 1 is Lemma 4, which is in turn based on Lemma 3. However, we include the short Lemmas 1 and 2 that provide better bounds on $\tilde{h}(n, H)$ when H is either a star or a vertex-disjoint union of two stars.

Lemma 1. *Let G be a bipartite graph with parts U and V of size n each. Assume that the degrees of the vertices from U are less than s. Then $\widetilde{\alpha}(G) \geq n/s$.*

Proof. Let U' be a subset of n/s vertices of U. Then $|N(U')| \leq (s-1)n/s = n - n/s$. Let $V' = V - N(U')$, we have $|V'| \geq n/s$ and (U', V') forms a co-biclique. $\square$

Lemma 2. *Let $s_1 \geq s_2 > 0$. Then $\tilde{h}(n, H_{s_1,s_2}) \geq \tilde{h}(n, H_{s_1}) \geq \frac{n}{2s_1}$.*

Proof of Lemma 2. Let $H = H_s$, for $s = s_1$. Let G be an H-free bipartite graph with top part U and bottom part V, each of size n. We will show that $\tilde{h}(G) \geq \frac{n}{2s}$. Let $\{u_1, \ldots, u_n\}$ be an ordering of the vertices of U, such that $d(u_i) \leq d(u_j)$ if $i < j$. Since H is isomorphic to its bipartite complement, we can assume that $d(u_{n/2}) \leq n/2$. Assume first that there is an $i < n/2$ with $|N(u_i) \setminus N(u_{n/2})| \geq s$. Then we have a set V' of s vertices, $V' \subseteq N(u_i) \setminus N(u_{n/2})$, and since $|N(u_i)| \leq |N(u_{n/2})|$, we also have a set V'' of s vertices, $V'' \subseteq N(u_{n/2}) \setminus N(u_i)$. But then $\{u_i, u_{n/2}\}$ and $V' \cup V''$ induce H. Let $Y = V \setminus N(u_{n/2})$. We have $|Y| \geq n/2$ and by the above argument, we have $|Y \cap N(u_i)| \leq s - 1$, for all $i \leq n/2$. Applying Lemma 1 to a subgraph of G induced by Y and $\{u_1, \ldots, u_{n/2}\}$, we get $\widetilde{\alpha}(G) \geq \frac{n}{2s}$. $\square$

Note that in the case where $s_2 = 0$, one can show that $\tilde{h}(n, H_{s,0}) \geq \frac{n}{2s-1}$.

We need an auxiliary lemma about rooted trees. We call two vertex disjoint subforests of a rooted tree *independent* if no vertex in one forest is an ancestor of a vertex in the other forest.

Lemma 3. *Any rooted tree on n vertices has either height at least $n/4$, or it contains two independent subforests on at least $n/4$ vertices each.*

Proof. Let T be a rooted tree on n vertices with root r. We say that a path in T is a *root-leaf* path if its endpoints are r and a leaf of T. For a root-leaf path P in T, let $m(P)$ be the order of a largest component in $T - V(P)$. Among all root-leaf paths in T, let P have the smallest value of $m(P)$. If $m(P) > n/2$, consider a leaf u of T which belongs to the largest component of $T - V(P)$. Then $m(P') < m(P)$ for an r-u path P', a contradiction. Thus, we have that $m(P) \leq n/2$. If P has at least $n/4$ vertices, the height of T is at least $n/4$, and we are done. Thus, the number of vertices in $T - V(P)$ is at least $3n/4$. Consider a set S of components of $T - V(P)$ with the total number of vertices s as small as possible, such that $s \geq n/4$. If the number of remaining vertices of $T - V(P)$ is less than $n/4$, then $s > n/2$, so there are at least two components in S. Removing a smallest one results in a set of components of size less than s and at least $n/4$ vertices, a contradiction. Thus, S and the remaining components of $T - V(P)$ form two independent subforests on at least $n/4$ vertices each. $\square$

Recall that $M_s = M_{s,s}$ and $M_s^* = M_{s,s}^*$.

Lemma 4. *Let $s_1 \geq s_2 > 0$. Then $\tilde{h}(n, M_{s_1,s_2}) \geq \tilde{h}(n, M_{s_1}) \geq n\left(\frac{1}{4}\frac{1}{2s_1+1}\right)\left(1 - \frac{1}{8s_1}\right)$ and $\tilde{h}(n, M_{s_1,s_2}^*) \geq \tilde{h}(n, M_{s_1}^*) \geq n\left(\frac{1}{8}\frac{1}{2s_1+1}\right)\left(1 - \frac{1}{8s_1}\right)$.*

Proof. Let $s = s_1$, $s \geq 1$. Let $G' \subseteq K_{n,n}$ have partite sets U' and V and assume that G' has no induced copy of M_s with the smaller partite set in U'. We shall show that $\tilde{h}(G') \geq n\left(\frac{1}{4}\frac{1}{2s+1}\right)(1 - \frac{1}{8s})$. Assume that $\tilde{h}(G') < \frac{n}{8s}$, otherwise we are done.

Let S be the set of vertices from U' of degree at most $2s$ in G'. Assume $|S| \geq \frac{n}{8s}$ and let $S' \subseteq S$, $|S'| = \frac{n}{8s}$. If $V' = N(S')$, then we have $|V'| \leq \frac{n}{4}$. Thus, $(V - V', S')$ forms a co-biclique with parts of size at least $\frac{n}{8s}$. This contradicts our assumption that $\tilde{h}(G') < \frac{n}{8s}$. Thus, $|S| \leq \frac{n}{8s}$.

Let $G = G' - S$ and let $U = U' - S$. We see that $|U| \geq n - \frac{n}{8s}$. We shall show that $\tilde{h}(G) \geq n\left(\frac{1}{4}\frac{1}{2s+1}\right)(1 - \frac{1}{8s})$, that would imply the desired result for G' since $\tilde{h}(G') \geq \tilde{h}(G)$.

Introduce an *auxiliary graph* I with vertex set U and two vertices adjacent if and only if their neighborhoods in G intersect.

Claim 0. Let $x, y, z \in U$, $xy, yz \in E(I)$, $xz \notin E(I)$. Then $d(x) + d(z) < d(y) + 2s$.
Assume otherwise, then without loss of generality $|N(x) \setminus N(y)| \geq s$. Hence, $|N(y) \setminus N(x)| < s$, implying $|N(y) \setminus N(z)| > s$, which in turn implies that $|N(z) \setminus N(y)| < s$. Since $N(z) \cap N(y) \subseteq N(y) \setminus N(x)$, we have $|N(z) \cap N(y)| < s$. Thus, $|N(z)| < 2s$, a contradiction.

Claim 1. In each nontrivial component F of any induced subgraph of I, a vertex v of F that has the highest degree in G is adjacent to all other vertices in F.
We shall prove this statement by induction on the order of F with a trivial basis. Consider F, v, and a component F' of $F - v$ with at least two vertices. By induction, a vertex v' of F' with highest degree in G is adjacent to all other vertices in F'. We know that v is adjacent to some vertex of F'. Assume that v is not adjacent to some vertex in F'. Then there are two vertices $w, w' \in V(F')$, such that $ww', wv \in E(I)$, $w'v \notin E(I)$, where $v' \in \{w, w'\}$. Then $d(v) + d(w') < d(w) + 2s$ by Claim 0. However, $d(w) \leq d(v)$, thus, $d(w') < 2s$, a contradiction, since we deleted all small degree vertices from the respective part of our graph.

Thus, each component of I is a *tree closure*, i.e., a graph obtained from a rooted tree by adding, for each vertex v, all edges between v and each vertex on a path from v to the root.

Consider a component of I on a vertex set Q. By Lemma 3, either there is a root-leaf path on $|Q|/4$ vertices in the underlying tree, which in turn corresponds to a clique in the tree-closure, or there are two disjoint subsets of vertices of size at least $|Q|/4$ each so that there are no edges from I between them. I.e., each component of I has either a clique on a quarter of its vertices, or a co-biclique with at least a quarter of the component's vertices in each part. We say a component is of *Type 1* in the former case and we say that it is of *Type 2* in the latter case.

Case 1. There is a subset X of U, $|X| \geq \frac{2s}{2s+1}|U|$, so that X is spanned by the components of I of Type 1, i.e., with large cliques.

We have that X contains pairwise disjoint, pairwise nonadjacent cliques in I on a vertex set X', $|X'| \geq |X|/4$. If each clique has size at most $|X'|/3$, then split the cliques of $I[X']$ into two groups of total size at least $|X'|/3$, let the vertex sets of these groups be X'' and X'''. Assume, without loss of generality, that $|N(X'')| \leq |N(X''')|$ in G. Then $(V - N(X'), X''')$ induces a co-biclique with parts of size at least $|X'|/3$. If there is a clique in $I(X')$ on a set of vertices X'', $|X''| \geq |X'|/3$. Consider the set $Y = N_G(X'')$. If $|Y| < 2n/3$, then $(X'', V - Y)$ induces a co-biclique with parts of size at least $\min\{|X''|, n/3\}$. If $|Y| \geq 2n/3$, then we see that (X'', Y) does not induce H_s, otherwise these stars and a common neighbor of their centers induce M_s. Thus, by Lemma 2, $\tilde{h}(G) \geq \min\{|X''|, 2n/3\}/(2s)$. Since $|X''| \geq \frac{1}{4}|X| \geq \frac{1}{4}\frac{2s}{2s+1}|U| \geq \frac{1}{4}\frac{2s}{2s+1}(n - \frac{n}{8s})$, $\tilde{h}(G) \geq \frac{1}{4(2s+1)}(n - \frac{n}{8s})$.

Case 2. There is a subset X of U, $|X| \geq \frac{1}{2s+1}|U|$, spanned by the components of I of Type 2, i.e., with large co-bicliques. Consider one part of such co-bicliques and form their union, X'. Similarly, consider the other parts of the co-bicliques and let their union be X''. We have $|X'|$, $|X''| \geq |X|/4$, and there are no edges between X' and X''. In particular, $N_G(X') \cap N_G(X'') = \emptyset$. Assume, without loss of generality, that $|N_G(X')| < n/2$. Then $(X', V - N_G(X'))$ induces a co-biclique with parts of sizes at least $\min\{|X'|, n/2\} \geq |X|/4 \geq \frac{1}{4}\frac{1}{2s+1}(n - \frac{n}{8s})$. Thus, $\tilde{h}(G) \geq \frac{1}{4(2s+1)}(n - \frac{n}{8s})$.

This concludes the proof of the lemma for M_s. Now, consider an M_s^*-free bipartite graph G with top part U and bottom part V, each of size n. If there is a vertex $v \in V$ of degree $d(v) \leq n/2$, then the graph $G[U - N(v), V \setminus \{v\}]$ is M_s-free. Thus, by the previous result on M_s, we have $\tilde{h}(G) \geq \frac{1}{8(2s+1)}(n - \frac{n}{8s})$. If V does not contain a vertex of degree at most $n/2$, consider the bipartite complement G' of G. Since M_s^* is (bipartite) self-complementary, G' does not contain M_s^* either, but we have a vertex $v \in V$ with $d_{G'}(v) \leq \frac{n}{2}$ and thus, we can apply the same argument. □

Proof of Theorem 1. Consider a strongly acyclic graph H such that neither H nor H' is in the family $\mathcal{H}$. By Theorem 2, H is an induced subgraph of $M_{s,s}^*$ for some s. Thus $\tilde{h}(n, H) \geq \tilde{h}(n, M_{s,s}^*) \geq n/(30s) \geq n/(30|V(H)|)$ as follows from Lemma 4 and the fact that we can choose $s \leq |V(H)|$. □

3 Tight bounds for all strongly acyclic graphs with two vertices in each part

We consider strongly acyclic bipartite graphs with each part of size 2. These are exactly $2K_2$, P_4, and H_4, where H_4 is such a graph with exactly two adjacent edges and $2K_2$ is a graph with exactly two disjoint edges. We shall give bounds for $\tilde{h}(H)$ for each of these graphs. Recall that $\tilde{h}(H) = \tilde{h}(H')$, where H' is a bipartite complement of H.

Proposition 1. *Let G be a bipartite P_4-free graph with n vertices in each part. Then $\tilde{h}(G) \geq \lceil n/3 \rceil$. This bound is tight.*

Proof. Let G have partite sets U and V. It is easy to see that G is a pairwise vertex disjoint union of bicliques. Let, for some index set I, these bicliques have partite sets U_i and V_i of sizes a_i, b_i, respectively, $U_i \subseteq U$, $i \in I$. Observe first that $\min\{a_i, b_i\} < n/3$, for each $i \in I$, otherwise the i^{th} biclique gives us $\tilde{\omega}(G) \geq n/3$. Moreover, $a_i < n/3$ and $b_i < n/3$, for each $i \in I$ since otherwise a co-biclique with parts $U_i, V - V_i$ or $V_i, U - U_i$ has parts of size at least $n/3$.

Let I' be the set of indices, so that $a_i \leq b_i$, $i \in I'$. Let $I'' = I \setminus I'$. Let $U' = \cup_{i \in I'} U_i$, $V' = \cup_{i \in I'} V_i$, $U'' = U - U'$, $V'' = V - V'$, $a' = |U'|$, $a'' = |U''|$, $b' = |V'|$, $b'' = |V''|$. Consider a co-biclique with parts V', U''. We can assume that either b' or a'' is less than $n/3$, say $b' < n/3$. Then $a' < n/3$ since for each $i \in I'$, $a_i \leq b_i$. Thus, $a'' > 2n/3$.

Consider a minimal subset $I''' \subseteq I''$ such that $U''' = \cup_{i \in I'''} U_i$ has size $a''' > n/3$. Then $a''' < 2n/3$ otherwise for any $i \in I'''$, $|U''' - U_i| > 2n/3 - n/3 = n/3$. In particular, we could have taken $I''' - \{i\}$ instead of I''', contradicting its minimality. Thus, $V''' = \cup_{i \in I'''} V_i$ has size less than $2n/3$. This implies that U''' and $V - V'''$ form a co-biclique with each part of size at least $n/3$.

We have shown $\tilde{h}(G) \geq n/3$ (and thus, $\tilde{h}(G) \geq \lceil n/3 \rceil$). To see that this bound is sharp when 3 divides n, take 3 vertex disjoint copies of $K_{n/3,n/3}$. For $n = 3m+1$, take one copy $K_{m+1,m+1}$ and two of $K_{m,m}$, and for $n = 3m+2$, take two copies of $K_{m+1,m+1}$ and one of $K_{m,m}$. □

Proposition 2. *Let G be a $2K_2$-free bipartite graph with n vertices in each part. Then $\tilde{h}(G) \geq \lceil n/2 \rceil$. This bound is tight.*

Proof. Let G be bipartite $2K_2$-free with parts U, V of size n each. Then we have for any vertices $u, u' \in U$ that $N(u) \subseteq N(u')$ or $N(u') \subseteq N(u)$. Thus, there is a total ordering $\{u_1, \ldots, u_n\}$ of the vertices in U where $i < j$ if and only if $N(u_j) \subseteq N(u_i)$. Consider the two subgraphs $G_1 = G[U_1 \cup V_1]$, $G_2 = G[U_2 \cup V_2]$, with $U_1 = \{u_1, \ldots, u_{\lceil \frac{n}{2} \rceil}\}$, $V_1 = N(u_{\lceil \frac{n}{2} \rceil})$, $U_2 = \{u_{\lceil \frac{n}{2} \rceil}, \ldots, u_n\}$, $V_2 = V \setminus N(u_{\lceil \frac{n}{2} \rceil})$.

By our vertex ordering, we have that $N(u_{\lceil \frac{n}{2} \rceil}) \subseteq N(u_i)$, $1 \leq i \leq \lceil \frac{n}{2} \rceil$, and thus, G_1 is a biclique. On the other hand, $V \setminus N(u_{\lceil \frac{n}{2} \rceil}) \subseteq V \setminus N(u_i)$, $\lceil \frac{n}{2} \rceil \leq i \leq n$, and thus, G_2 is a co-biclique. We know that $|U_1| = \lceil \frac{n}{2} \rceil$ and $|U_2| \geq \lceil \frac{n}{2} \rceil$. Since $|V_1| + |V_2| = n$, one of V_1 and V_2 has to have size at least $\lceil \frac{n}{2} \rceil$, which gives us $\max\{\widetilde{\omega}(G), \widetilde{\alpha}(G)\} \geq \lceil \frac{n}{2} \rceil$.

Note that $\lceil \frac{n}{2} \rceil$ is sharp. Consider a bipartite graph G that is a union of a complete bipartite graph $K_{\lceil n/2 \rceil, n}$ and $\lfloor \frac{n}{2} \rfloor$ isolated vertices added to the smaller part. Both parts have size n, we have $\tilde{h}(G) = \lceil \frac{n}{2} \rceil$ and G is $2K_2$-free. □

Recall that H_4 is a bipartite graph with two vertices in each part and two edges, such that the edge are adjacent.

Proposition 3. *Let G be an H_4-free bipartite graph with n vertices in each part. Then $\tilde{h}(G) \geq \lfloor 2n/5 \rfloor$. This bound is tight for $n \equiv 0 \pmod 5$.*

Proof. Let G have parts U and V of size n each, assume n is divisible by 5. Denote by G' the bipartite complement of G. Let U correspond to the part of H_4 containing the vertex of degree 2. First observe, that $|N(u) \setminus N(u')| \leq 1$, for any $u, u' \in U$.

Claim: There is a set $X \subseteq V$ such that for a graph Q, where $Q = G$ or $Q = G'$ the following holds. For any $u \in U$ we have $X \subseteq N_Q(u)$ and $|N_Q(u)| \leq |X| + 1$. I.e., the neighborhoods of the vertices from U form a sunflower set system with petals of size at most one.

To prove the claim, consider a vertex $u' \in U$ of largest degree. If $N_G(u) \subseteq N_G(u')$ for each $u \in U$, then $X = V \setminus N_G(u')$ satisfies the conditions of the claim with $Q = G'$. So, we assume that there is a vertex $u'' \in U$ such that $N(u'') \not\subseteq N(u')$. In particular, we must have that $N_G(u') = V' \cup \{v'\}$ and $N_G(u'') = V' \cup \{v''\}$, for $V' \subseteq V$, and $v', v'' \in V \setminus V'$, $v' \neq v''$. Let $V'' = V' \cup \{v', v''\}$.

If there is a vertex $u \in U$, such that $N_G(u) \not\subseteq V''$ and $V' \not\subseteq N_G(u)$, then there are vertices $v_1 \in V'$, $v_2 \in V''$ such that $v_1, v_2 \notin N_G(u)$. Then either $\{u', u, v_1, v_2\}$ or $\{u'', u, v_1, v_2\}$ induces H_4, a contradiction. Thus, for each $u \in U$, either $V' \subseteq N(u)$ or $N(u) \subseteq V''$.

If for any vertex $u \in U$, $N(u) \subseteq V''$, then $X = V - V''$ satisfies the conditions of the claim with $Q = G'$. If for any $u \in U$, $V' \subseteq N(u)$, then the claim is satisfied with $X = V'$ and $Q = G$. If there are u_1, u_2 such that $u_1, u_2 \notin \{u', u''\}$, $V' \not\subseteq N(u_1) \subseteq V''$ and $V' \subseteq N(u_2) \not\subseteq V''$, we see that u_1, u_2, v', v'' form a copy of H_4. This proves the claim.

Now that we proved the claim, it remains to find a large biclique or co-biclique. Assume without loss of generality that $Q = G$ in the claim. If $|X| \geq 2n/5$, then (U, X) induces a biclique with parts of sizes at least $2n/5$. Assume that $|X| < 2n/5$. Let $Y = V \setminus X$. We see that (U, Y) induces a pairwise disjoint union of stars in G with centers in Y. Let $s = |Y|$, note that $s \geq 3n/5$. Let $Y = \{y_1, \dots, y_s\}$ such that $d(y_i) \leq d(y_{i+1})$, $i = 1, \dots, s-1$. Let $Y_1 = \{y_1, \dots, y_{2n/5}\}$. If $|N(Y_1)| \leq 3n/5$, then $(Y_1, U \setminus N(Y_1))$ induces a co-biclique with parts of sizes at least $2n/5$. If $|N(Y_1)| > 3n/5$, then $d(y_i) \geq 2$ for all $i > 2n/5$. Thus, $|N(Y)| \geq |N(Y_1)| + 2|Y - Y_1| > 3n/5 + 2 \cdot (3n/5 - 2n/5) = n$, a contradiction since $N(Y) \subseteq U$.

To show that the bound is tight, construct the following graph G with parts U and V of sizes n, $n \equiv 0 \pmod 5$. Let $U = U_1 \cup U_2$ where $|U_1| = 2n/5$ and $|U_2| = 3n/5$. Let $V = V_1 \cup V_2 \cup V_3$, where $|V_1| = |V_2| = 2n/5$ and $|V_3| = n/5$. Let G have all edges between V_1 and U, form a perfect matching between U_1 and V_2, and form a perfect matching between U_2 and $V_2 \cup V_3$. Note that if G has a copy of H_4, this copy has a vertex u of degree 2 in U. Thus, this copy must have a neighbor of u in V_1, that in turn is adjacent to all of U and thus could not have degree 1 in a copy of H. Thus, G is H_4-free. In addition, we see that $\tilde{h}(G) = 2n/5$. □

4 General bounds

In this section we work out known arguments for completeness.

Theorem 3. *Let H be a bipartite graph that is not strongly acyclic. Then there is an $\epsilon > 0$ such that for each sufficiently large n, $\tilde{h}(n, H) \leq n^{1-\epsilon}$. Moreover, if H or its bipartite complement contains C_4, C_6, or C_8, then ϵ could be taken any positive real strictly less than $1/3, 1/6$, or $1/16$, respectively.*

Proof. First, recall from Theorem 2, that if H is not strongly acyclic, then H or its bipartite complement contains C_4, C_6, or C_8. In case of C_4, we know by a result of Caro and Rousseau [2] that there is a bipartite graph G with parts of size n each that does not contain C_4 and with $\tilde{\alpha}(G) = O(n^{2/3})$. This result was shown using Lovász Local Lemma that we abbreviate as LLL. The LLL tells us that if there are bad events $A_i, \dots$ and positive numbers $x_i, \dots$ associated with these events, such that $\text{Prob}(A_i) \leq (1 - x_i) \prod_{j \sim i} x_j$, where $i \sim j$ if and only if A_i is adjacent to A_j in the dependency graph, then with positive probability none of the bad events happen.

We use the same approach to randomly create respective graphs with no C_6 and with no C_8 and not having large co-cliques. Consider $K_{n,n}$ and color each edge red with probability p and blue with probability $(1 - p)$. For a specific C_6, we say that there is a red bad event if this C_6 is red. Similarly, for a specific $K_{t,t}$, we say that there is a blue bad event if this $K_{t,t}$ is blue. We shall use the LLL to prove that with positive probability there are no bad events. Let P_r and P_b be the probabilities of the the red bad and blue events, respectively. Consider the dependency graph. Let d_{rr} be the number of red bad events adjacent to a red bad event, d_{rb} be the number of blue bad events adjacent to a red bad event, d_{bb} be the number of blue bad events adjacent to a blue bad event, and finally d_{br} be the number of red bad events adjacent to a blue bad event.

Then we have that $P_r = p^6$ and $P_b = (1-p)^{t^2}$. Counting the number of C_6's sharing an edge with a given C_6, we have $d_{rr} \le 6\binom{n}{2}^2 4 \le 6n^4$; counting the number of $K_{t,t}$'s sharing an edge with a given C_6 we have $d_{rb} \le 6\binom{n}{t-1}^2 \le n^{2t}$; counting the number of C_6's sharing an edge with a given $K_{t,t}$ we have $d_{br} \le t^2\binom{n}{2}^2 4 \le t^2 n^4$; and finally, counting the number of $K_{t,t}$'s sharing an edge with a given $K_{t,t}$ we have $d_{bb} \le t^2\binom{n}{t-1}^2 4 \le n^{2t}$.
Since here we have bad events of two types, let $x_i = x$ for red bad events and $x_i = y$ for blue bad events. We shall assign the values to p, t, x, and y such that

$$p^6 \le (1-x)x^{6n^4}y^{n^{2t}} \quad \text{and} \quad (1-p)^{t^2} \le (1-y)x^{n^2t^2}y^{n^{2t}}.$$

Let ϵ, ϵ' be small positive constants, say $\epsilon < 1/6$, $\epsilon' < 1/(6\epsilon) - 1$. Let

$$t = n^{1-\epsilon},\ x = 1 - n^{-5},\ y = 1 - n^{-2n^{1-\epsilon}},\ p = n^{-1+\epsilon(1+\epsilon')}.$$

We shall be using the fact that $(1-s) \approx e^{-s}$ for small s. Then, for large n we have

$$\begin{aligned}
p^6 &\approx n^{-6+6\epsilon(1+\epsilon')}, \\
(1-p)^{t^2} &\approx e^{-n^{-1+\epsilon(1+\epsilon')}\cdot n^{2-2\epsilon}} = e^{-n^{1-\epsilon(1-\epsilon')}}, \\
(1-x)x^{6n^4}y^{n^{2t}} &\approx n^{-5}e^{-n^{-5}6n^4}e^{-1} \gtrsim e^{-1}n^{-5}, \\
(1-y)x^{n^4t^2}y^{n^{2t}} &\approx n^{-2n^{1-\epsilon}}e^{-n^{-5}n^4n^{2-2\epsilon}}e^{-1}.
\end{aligned}$$

Thus, $p^6 \le (1-x)x^{6n^4}y^{n^{2t}}$ and $(1-p)^{t^2} \le (1-y)x^{n^2t^2}y^{n^{2t}}$. Therefore, by the LLL there is an edge-coloring of $K_{n,n}$ with no red C_6's and no blue $K_{n^{1-\epsilon},n^{1-\epsilon}}$.

To see the result for C_8, we closely follow the above argument, choose the parameters and notation similarly and define red bad event corresponding to a red C_8 and blue bad event as before.

Then we have that $P_r = p^8$ and $P_b = (1-p)^{t^2}$. We have $d_{rr} \le 8\binom{n}{3}^2 c \le Cn^6$, $d_{rb} \le 8\binom{n}{t-1}^2 \le n^{2t}$, $d_{br} \le t^2\binom{n}{3}^2 c \le Ct^2n^6$, and $d_{bb} \le t^2\binom{n}{t-1}^2 4 \le n^{2t}$.
Let ϵ, ϵ' be a small positive constants, say $\epsilon < 1/16$, $\epsilon' < 1/(16\epsilon) - 1$. Let

$$t = n^{1-\epsilon},\ x = 1 - n^{-7.5},\ y = 1 - n^{-2n^{1-\epsilon}},\ p = n^{-1+\epsilon(1+\epsilon')}.$$

Then, for large n we have

$$\begin{aligned}
p^8 &\approx n^{-8+8\epsilon(1+\epsilon')}, \\
(1-p)^{t^2} &\approx e^{-n^{-1+\epsilon(1+\epsilon')}\cdot n^{2-2\epsilon}} = e^{-n^{1-\epsilon(1-\epsilon')}}, \\
(1-x)x^{Cn^6}y^{n^{2t}} &\approx n^{-7.5}e^{-n^{-7.5}Cn^6}e^{-1}, \\
(1-y)x^{Cn^6t^2}y^{n^{2t}} &\approx n^{-2n^{1-\epsilon}}e^{-Cn^{-7.5}n^6n^{2-2\epsilon}}e^{-1}.
\end{aligned}$$

Thus, $p^8 \le (1-x)x^{Cn^6}y^{n^{2t}}$ and $(1-p)^{t^2} \le (1-y)x^{n^6t^2}y^{n^{2t}}$. Therefore, by the LLL there is an edge-coloring of $K_{n,n}$ with no red C_8's and no blue $K_{n^{1-\epsilon},n^{1-\epsilon}}$.

Now, let H be a bipartite graph containing C_{2k}, $k \in \{2,3,4\}$. Consider an edge-coloring of $K_{n,n}$ with no red C_{2k} and no blue $K_{n^{1-\epsilon},n^{1-\epsilon}}$. Let G be a graph formed by the red edges. Then G does not contain C_{2k} and thus, does not contain H, which implies that it does not contain an induced copy of H. In particular G does not have $K_{4,4}$. On the other hand, the bipartite complement of G does not contain $K_{n^{1-\epsilon},n^{1-\epsilon}}$. Thus, for sufficiently large n, $\tilde{h}(G) \le n^{1-\epsilon}$. $\square$

Theorem 4. *[6] Let H be a bipartite graph with parts of sizes k and l, $2 \leq k \leq l$. Let G be a bipartite graph with parts of sizes n, $n \geq l^k$. Then either G is H-free or $\tilde{h}(G) \geq t$, where $t = \lfloor (\frac{n}{l})^{1/k} \rfloor$.*

Lemma 5. *Let G be an $(l+1)$-partite graph with vertex classes $U_1, \ldots, U_l, V$, $|U_i| \geq t^m$, $|V| \geq tl$, for some integers $l, t, m \geq 2$. Let $\widetilde{\alpha}(U_i, V)$ and $\widetilde{\omega}(U_i, V)$ denote $\widetilde{\alpha}$ and $\widetilde{\omega}$ of the bipartite subgraph of G induced by (U_i, V). Let $\widetilde{\alpha}(U_i, V) < t$, $\widetilde{\omega}(U_i, V) < t$ for all $i \in [l]$. Then for any map $f : [l] \to \{0, 1\}$, there exists a vertex $v \in V$, such that*

$$\begin{aligned} |N(v) \cap U_i| &\geq t^{m-1}, \qquad f(i) = 1, \\ |U_i \setminus N(v)| &\geq t^{m-1}, \qquad f(i) = 0. \end{aligned}$$

Proof. Let G be an $(l+1)$-partite graph as in the statement and fix a function $f : [l] \to \{0, 1\}$. Assume there is no such vertex $v \in V$. Then for every $v \in V$, there must be at least one index $i_v \in [l]$, such that U_{i_v} is *bad* for v, meaning that

$$\begin{aligned} & |N(v) \cap U_{i_v}| \leq t^{m-1} - 1, && \text{if } f(i) = 1, \\ \text{or} \quad & |U_{i_v} \setminus N(v)| \leq t^{m-1} - 1, && \text{if } f(i) = 0. \end{aligned}$$

Since there are only l sets, we have a set U_i that is bad for at least $\frac{|V|}{l} \geq t$ vertices in V. Choose $V' \subseteq V$ such that $|V'| = t$ and $i_v = j$ for all $v, w \in V'$.

Case 1: $f(j) = 0$. Consider the subset $U' \subseteq U_j$ of vertices, that are adjacent to all vertices in V'. Since every vertex in V' is non-adjacent to at most $t^{m-1} - 1$ vertices we obtain $|U'| \geq |U_j| - t(t^{m-1} - 1) \geq t^m - (t^m - t) \geq t$. Thus, the pair (U', V') contains a copy of $K_{t,t}$, which contradicts our assumption of $\widetilde{\omega}(U_j, V) < t$.

Case 2: $f(j) = 1$. Consider the subset $U' \subseteq U_j$ of vertices, that have no neighbor in V'. Since every vertex in V' is adjacent to at most $t^{m-1} - 1$ vertices we obtain $|U'| \geq |U_j| - t(t^{m-1} - 1) \geq t^m - (t^m - t) \geq t$. Thus, the pair (U', V') contains a co-biclique of size t, which contradicts our assumption of $\widetilde{\alpha}(U_j, V) < t$.
Hence, for every f we find a vertex v, which is good for all sets U_i. ☐

Proof of Theorem 4. Let $H = (X \cup Y, E_H)$ be a bipartite graph with parts $X = \{x_1, \ldots, x_l\}$, $Y = \{y_1, \ldots, y_k\}$ with $2 \leq k \leq l$, let $n_0 = l^{k-1}$. Assume that $\widetilde{\alpha}(U, V) < t$ and $\widetilde{\omega}(U, V) < t$. We show how to find an induced copy of H. Note that from the choice of t and n, we have that $n \geq t^k l$ and $t \geq l$.

Partition U into l subsets $U_1, \ldots, U_l$, each of size at least t^k. Partition V into k subsets $V_1, \ldots, V_k$ each of size at least t^k. Since $t \geq l$, we have that $|V_i| \geq tl^{k-1} \geq tl$, for all $i \in [k]$. We shall apply Lemma 5 a total of k times to obtain subsets $U_i \supseteq U_i^1 \supseteq \cdots \supseteq U_i^k$, such that we can embed $x_i \in U_i^k$ and $y_j \in V_j$ for $i = 1, \ldots, l$ and $j = 1, \ldots, k$.

In Step 1, we apply Lemma 5 to the sets $U_1, \ldots, U_l, V_1$ with $m = k$ and

$$f_1 : [l] \to \{0, 1\}, \ i \mapsto \begin{cases} 1, & x_i y_1 \in E_H \\ 0, & x_i y_1 \notin E_H \end{cases}$$

to find a vertex $v_1 \in V$ and subsets

$$U_i^1 = \begin{cases} N(v_1) \cap U_i, & x_i y_1 \in E_H \\ U_i \setminus N(v_1), & x_i y_1 \notin E_H \end{cases}$$

such that $|U_i^1| \geq t^{k-1}$ for $i \in [l]$.

Assume that after Step j we have subsets $U_1^j, \dots, U_l^j$, $|U_i^j| \geq t^{k-j}$.
In Step $j+1$ for $j < k$, we apply the lemma again, to the sets $U_1^j, \dots, U_l^j, V_{j+1}$ with $m = k-j-1 \geq 2$ and

$$f_j : [l] \to \{0,1\},\ i \mapsto \begin{cases} 1, & x_i y_{j+1} \in E_H \\ 0, & x_i y_{j+1} \notin E_H \end{cases}$$

to find a vertex $v_{j+1} \in V_{j+1}$ and subsets

$$U_i^{j+1} = \begin{cases} N(v_1) \cap U_i^j, & x_i y_{j+1} \in E_H \\ U_i^j \setminus N(v_1), & x_i y_{j+1} \notin E_H \end{cases}$$

such that $|U_i^{j+1}| \geq t^{(k-j)-1} = t^{k-(j+1)}$ for $i \in [l]$.

We finish after k steps, and by our choice of n, t, we still obtain $|U_i^k| \geq 1$, $i \in [l]$. Thus, we have found vertices $v_1, \dots, v_k$ where we can embed $\{x_1, \dots, x_k\}$ and nonempty sets of candidates U_i^k, in which we can embed Y. This concludes the proof. $\square$

5 Concluding Remarks

During the review process of this paper, more results and insights on this problem have been obtained. We could now say more about the remaining four open cases. Using an approach very similar to one of Bousquet, Lagoutte, and Thomassé [3], we can show that $\tilde{h}(\mathrm{Forb}(n, P_7))$ and $\tilde{h}(\mathrm{Forb}(n, P_6))$ are linear in n, thus, taking care of two out of the four missing cases. For the last two graphs $H = \tilde{P}_5$ and $H = \tilde{H}_{3,4}$, a more general result on induced trees from a very recent manuscript by Scott, Seymour, and Spirkl [7], implies that $\tilde{h}(\mathrm{Forb}(n, H))$ is also linear in n. In addition, in [1], we could determine the asymptotic behavior of $\tilde{h}(\mathrm{Forb}(n, K_{1,s}))$ exactly for fixed but large s, improving on the bound in Lemma 1.

Acknowledgments The authors thank Jacob Fox for bringing the connections between the considered problem and the VC-dimensions of sets systems to their attention. The first author thanks the Department of Mathematics at Stanford University for the hospitality. The second author was partially supported by the Institute for Basic Science (IBS-R029-C1). The authors thank both referees for careful reading of the manuscript.

References

[1] Maria Axenovich, Jean-Sébastien Sereni, Richard Snyder, and Lea Weber. *Bipartite independence number in graphs with bounded maximum degree.* arXiv:2002.10930 (2020).

[2] Yair Caro and Cecil Rousseau. *Asymptotic bounds for bipartite Ramsey numbers.* The Electronic Journal of Combinatorics, 8(1):17 (2001).

[3] Nicolas Bousquet, Aurélie Lagoutte, and Stéphan Thomassé. *The Erdős-Hajnal conjecture for paths and antipaths.* Journal of Combinatorial Theory, Series B 113 (2015) 261–264.

[4] Maria Chudnovsky. *The Erdős-Hajnal conjecture – a survey.* Journal of Graph Theory 75 (2014) 178–190.

[5] Paul Erdős and András Hajnal. *Ramsey-type theorems.* Discrete Applied Mathematics 25 (1989) 37–52.

[6] Paul Erdős, András Hajnal, and János Pach. *A Ramsey-type theorem for bipartite graphs.* Geombinatorics, 10(DCG-ARTICLE-2000-001) (2000) 64–68.

[7] Alex Scott, Paul Seymour, and Sophie Spirkl. *Pure pairs. IV. Trees in bipartite graphs.* arXiv:2009.09426 (2020).

[8] Jacob Fox, János Pach, and Andrew Suk. *Erdős-Hajnal conjecture for graphs with bounded VC-dimension.* 33rd International Symposium on Computational Geometry, Art. No. 43, 15 pp., LIPIcs. Leibniz Int. Proc. Inform., 77, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern (2017).

[9] Daniel Korándi, János Pach, and István Tomon. *Large homogeneous submatrices.* arXiv:1903.06608 (2019).