

Forbidden induced bipartite graphs

Peter Allen*

Department of Mathematics
London School of Economics
Houghton St.
London WC2A 2AE
U.K.

September 19, 2006

CDAM Research Report LSE-CDAM-2006-10

Abstract

Given a fixed bipartite graph H , we study the asymptotic speed of growth of the number of bipartite graphs on n vertices which do not contain an induced copy of H . Whenever H contains either a cycle or the bipartite complement of a cycle, the speed of growth is $2^{\Omega(n^{\frac{6}{5}})}$. For every other bipartite graph except the path on seven vertices, we are able to find both upper and lower bounds of the form $n^{cn+o(n)}$. In many cases we are able to determine the correct value of c .

1 Introduction

It is well known (see Prömel and Steger [8]) that the number of simple graphs G on n vertices which do not contain an induced copy of H grows either as $n^{O(n)}$, when H is an induced subgraph of P_4 , or as $2^{\Theta(n^2)}$, when H is not an induced subgraph of P_4 . Brightwell, Grable and Prömel [4] have studied the equivalent problem for partial orders, where the situation is not so straightforward.

We consider the equivalent problem for bipartite graphs.

Let $G = G[X, Y]$ be a bipartite graph with bipartition (X, Y) . We say that X is the *lower part*, and Y the *upper part*, of G . We will draw diagrams accordingly. We say that the

*email: p.d.allen@lse.ac.uk

bipartite complement of G is the bipartite graph which has edges between X and Y exactly where G does not, together with the bipartition (X, Y) . If z is a vertex in $G[X, Y]$, then as usual we say that the degree of z , $d(z)$, is the number of vertices (in the part not containing z) adjacent to z . We say that the *co-degree* of z is the number of vertices in the part not containing z which are not adjacent to z .

Let $G = G[X, Y]$ and $H = H[W, Z]$. We say that G *contains a copy* of H if there exist $W' \subset X$, $Z' \subset Y$, such that the induced subgraph of G on the vertices $W' \cup Z'$, with bipartition (W', Z') , is isomorphic to $H[W, Z]$.

We consider three closely related problems.

First, let $H = H[W, Z]$. We wish to estimate the number $Forb_{m,n}(H)$ of graphs with bipartitions $G[X, Y]$ which do not contain a copy of H , in terms of the sizes m, n of the parts X, Y of G . We will restrict our attention to the case $n = \Theta(m)$.

Second, let $H = H[W, Z]$. We wish to estimate the number $Forb_n(H)$ of bipartite graphs G on n vertices such that no bipartition of G contains a copy of H .

Third, let H be a fixed bipartite graph. We wish to estimate the number $Forb_n^*(H)$ of bipartite graphs G on n vertices such that no bipartition of G contains a copy of any $H[W, Z]$, (W, Z) a bipartition of H .

As an illustration of the differences between these three problems, consider the bipartite graph on four vertices $SI(2, 1)$, as shown in Figure 1, with the bipartition as shown there.



Figure 1: $SI(2, 1)$ and allowed graphs for the first and second problems

A bipartite graph $G[X, Y]$ containing no copy of $SI(2, 1)$ with the given bipartition has the property that for each $x \in X$, either x is adjacent to no vertex in Y , to exactly one vertex in Y , or to every vertex in Y , for a total of $n + 2$ possibilities for each of the m vertices in X . Since every graph with this property contains no copy of $SI(2, 1)$ with the given bipartition, $Forb_{m,n}(SI(2, 1)) = (n + 2)^m$. The second graph in Figure 1 contains no copy of $SI(2, 1)$ with the given bipartitions – even though it is simply $SI(2, 1)$ the other way up.

By contrast, suppose that G is a bipartite graph on n vertices such that no bipartition of G contains $SI(2, 1)$ with the given bipartition. If G contains a vertex x of degree two or greater, then G must be connected and every vertex in the part not containing x must be adjacent to x . Thus G has three possible structures. First, G has only vertices of degree less than two. Second, G is a complete bipartite graph. Third, G is not a complete bipartite graph, but there are two adjacent vertices x and y in G such that every vertex in G is adjacent to either x or y , and every edge of G meets either x or y . The third graph in Figure 1 is an example of this third structure. It is clear that this condition is more restrictive than the condition for the first problem.

Finally, suppose that G is a bipartite graph on n vertices such that no bipartition of G contains a copy of $SI(2, 1)$ with any bipartition. Then certainly G does not contain $SI(2, 1)$ with the bipartition shown in Figure 1, so that G must be one of the three structures mentioned in the previous paragraph. But G also does not contain $SI(2, 1)$ with the bipartition having two vertices in each part. If n is at least five, the third structure in the previous paragraph must contain a copy of $SI(2, 1)$ with this alternative bipartition, so that (for $n \geq 5$) G is either a complete bipartite graph or contains only vertices of degree less than two.

We observe that $Forb_n(H[U, V])$ and $Forb_n^*(H)$ coincide when H is connected.

As is well known (see e.g. Bollobás [2]), when H is any bipartite graph there are $2^{o(n^2)}$ bipartite graphs on n vertices which do not contain H as a subgraph; a similar easy application of the Szemerédi Regularity Lemma shows that there are $2^{o(n^2)}$ bipartite graphs on n vertices which do not contain H as an induced subgraph. We will be interested in finding lower bounds and better upper bounds; we will be particularly interested in finding bounds of the form n^{cn} for constant c .

We will see that the bipartite graphs fall into the following classes: graphs containing cycles or the bipartite complements of cycles, five infinite families of graphs, and six exceptional graphs on six and seven vertices. Spinrad [9] observes that there is a similarity between partial orders of height two and bipartite graphs, so that we could use the results of Brightwell, Grable and Prömel to show that upper bounds of the form n^{cn} exist for some of these graphs. He also points out that there are graphs, such as P_5 , for which we can find tight bounds on $Forb_n(P_5)$, but which correspond to partial orders that Brightwell, Grable and Prömel were unable to classify.

We will find that our three problems are in fact very similar. Although the second and third problems seem more obvious and interesting, the methods we use to obtain upper bounds for each of the five infinite families naturally apply to the first problem. We spend most of the paper dealing with this problem.

We obtain the bounds given in the Tables 1 and 2 for each of the three problems. We observe that the results for the second and third problems differ only in that forbidding certain graphs ($SI(0, l)$, $DS(k, 0)$ and $DS^*(k, 0)$) makes sense in the context of the second problem where their bipartition is fixed, but in the context of the third problem they are examples of simpler graphs (the empty graph on $l + 1$ vertices, $SI(k, 1)$ and $SI(k, 2)$ respectively). Note that in a few cases we can find better bounds than those given in the tables; in particular we can show that the upper bound is correct for $Forb_n(JS(1, 0))$ and that the lower bounds are correct for $Forb_n(DS(k, 0))$.

A special case that might be of interest is that of the bipartite graphs on n vertices which do not contain the path on k vertices as an induced subgraph. Trivially when $k = 1, 2$ we have respectively zero and one bipartite graphs which are P_k -free. The P_3 -free bipartite graphs are the sub-matchings (disjoint unions of copies of K_1 and K_2), of which there are $n^{\frac{n}{2} + o(n)}$. The P_4 -free bipartite graphs are easily seen to be disjoint unions of complete bipartite graphs, and there are $n^{n + o(n)}$ such (we note that $P_4 = JS(1, 0)$; in this case the general lower bound in Tables 1 and 2 can be improved). The P_5 -free bipartite graphs are disjoint unions of difference graphs ($2K_2$ -free bipartite graphs), and the P_6 -free bipartite graphs are a subclass

of the bi-cographs introduced by Giakoumakis and Vanherpe [6]; in both cases there are $n^{n+o(n)}$ such bipartite graphs. We have neither good bounds on the growth rate of, nor useful structural information about, the P_7 -free bipartite graphs. For $k \geq 8$, P_k contains the bipartite complement of C_4 ; and there are $2^{\Omega(n^{\frac{6}{5}})}$ graphs whose bipartite complements have girth at least six and so do not contain P_k .

Throughout this paper we will use the names in Table 1 for the various graphs we study.

Following Balogh, Bollobás and Weinreich [1], we say that the *speed* of $Forb(H)$ is the rate of growth of $Forb(H)$. Balogh, Bollobás and Weinreich showed that while hereditary properties of graphs have highly constrained and well-behaved speeds when their speeds are bounded above by $n^{n+o(n)}$, this is no longer true for hereditary properties whose speeds are faster than $n^{n+o(n)}$ but slower than $2^{\epsilon n^2}$ for all $\epsilon > 0$. For example, some such properties have speeds which oscillate between n^{cn} and $2^{n^{2-\epsilon}}$.

As can be seen in the Tables 1 and 2, most of the interesting cases of our problems are hereditary properties whose speeds are in this penultimate range, but which nevertheless are reasonably well-behaved.

In Section 2 we show that, since there are many (more than n^{cn} for any c) graphs on n vertices with large girth, the speed of $Forb_{m,n}(H)$ is large for all H which contain either a cycle or the bipartite complement of a cycle. This leaves only five infinite families of graphs and a few exceptional graphs on 6 or 7 vertices; we will also find bounds for the simplest of these infinite families in this section.

It is obvious that any graph G with maximum degree (or co-degree) less than the maximum degree (or co-degree) of H cannot contain a copy of H . It is easy to show that there are $n^{\frac{kn}{2}}$ bipartite graphs on n vertices with maximum degree k . One might perhaps guess that, when H does not contain a cycle or the complement of a cycle, the speed of $Forb_{m,n}(H)$ should depend principally upon the maximum degree or co-degree of H ; and it is not too hard to show that for each of the infinite families this is true. This would lead us to expect that the lower bounds on $Forb_n(H)$ should be given by families of graphs with small maximum degree or co-degree. Interestingly, this is not always the case. We find large families of graphs giving substantially better lower bounds than the obvious ones for four of the five infinite families: $DS(k, l)$, $DS^*(k, l)$, $JS(k, l)$ and $JS^*(k, l)$. We are able to show that these large families of graphs actually give the correct speed for the first three infinite families when $k = l$; much of the work in this paper is involved in proving the upper and lower bounds on $Forb_{m,n}(H)$ for the four infinite families, which we do in Section 3. In Section 4 we use the bounds from the previous sections to obtain similarly good bounds on $Forb_n(H)$ and $Forb_n^*(H)$ for all H but the exceptional graphs.

In Section 5 we use a structural result of Lozin [7] to obtain good upper bounds on $Forb_n(H)$ for all of the exceptional graphs except the path on seven vertices, P_7 . This leaves finding good bounds for $Forb_n(P_7)$ as the most significant open problem. We observe that this structural result does not suffice to bound $Forb_{m,n}(H[U, V])$ above for three more of the exceptional graphs (see Table 1).


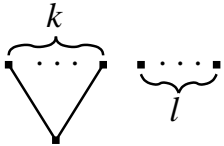
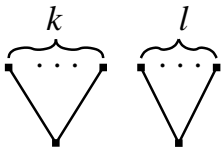
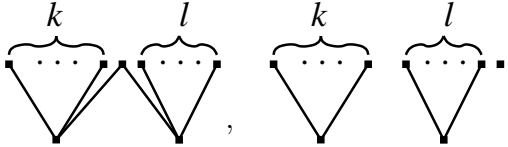
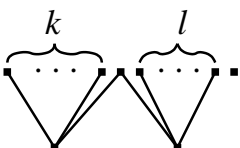


H	$Forb_{m,n}(H)$	
	Lower	Upper
 k $k \geq 1$	0 (for sufficiently large m, n)	
$SI(k, l)$  $k + l \geq 1$	$m^{\max(k-1, l-1)m + o(m)}$	
$DS(k, l)$  $k \geq l \geq 1$ or $k \geq 2, l = 0$	$m^{\max((k-1)m, lm+n) + o(m)}$	$m^{km+n+o(m)}$
$JS(k, l)$ $DS^*(k, l)$  $k \geq l \geq 1$ or $k \geq 1, l = 0$	$m^{\max(km, lm+n) + o(m)}$	$m^{km+n+o(m)}$
$JS^*(k, l)$  $k \geq l \geq 1$ or $k \geq 1, l = 0$	$m^{\max(km, lm+n) + o(m)}$	$m^{km+2n+o(m)}$
	$m^{m+n+o(m)}$	
	$m^{m+n+o(m)}$	$2^{o(m^2)}$
All other bipartite graphs	$2^{\Omega(m^{\frac{6}{5}})}$	$2^{o(m^2)}$

Table 1: Summary of the bounds obtained for the first problem


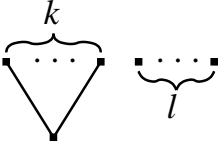
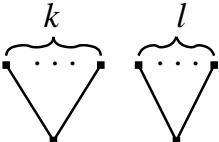
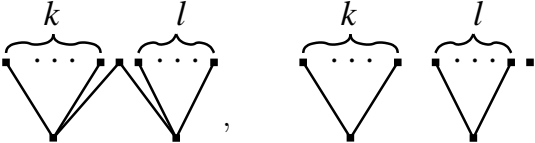
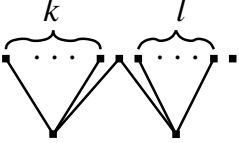


H	$Forb_n(H), Forb_n^*(H)$	
	Lower	Upper
 $k \geq 1$	0 (for sufficiently large n)	
$SI(k, l)$  $k + l \geq 1^{(*)}$	$n^{\max(\frac{(k-1)n}{2}, \frac{(l-1)n}{2}) + o(n)}$	
$DS(k, l)$  $k \geq l \geq 1^{(*)}$	$n^{\max(\frac{(k-1)n}{2}, \frac{(l+1)n}{2}) + o(n)}$	$n^{\frac{(k+1)n}{2} + o(n)}$
$JS(k, l)$ $DS^*(k, l)$  $k \geq l \geq 1$ or $k \geq 1, l = 0^{(*)}$,	$n^{\max(\frac{kn}{2}, \frac{(l+1)n}{2}) + o(n)}$	$n^{\frac{(k+1)n}{2} + o(n)}$
$JS^*(k, l)$  $k \geq l \geq 1$ or $k \geq 1, l = 0$	$n^{\max(\frac{kn}{2}, \frac{(l+1)n}{2}) + o(n)}$	$n^{\frac{(k+2)n}{2} + o(n)}$
	$n^{n+o(n)}$	
	$n^{n+o(n)}$	$2^{o(n^2)}$
All other bipartite graphs	$2^{\Omega(n^{\frac{6}{5}})}$	$2^{o(n^2)}$

Table 2: Summary of the bounds obtained for the second and third problems
 $(^*)$ $SI(0, l)$, $DS(k, 0)$ and $DS^*(k, 0)$ apply only to the second problem.

2 Preliminaries

In this section we solve the easy cases of the first problem, and characterise the remaining cases.

First we show that there are many graphs which do not contain short cycles. We make use of a result of Benson [3] showing that there exists a bipartite graph with large girth and many edges.

Theorem 1. *For q an odd power of 3, there exists a bipartite graph B with $q^5 + q^4 + q^3 + q^2 + q + 1$ vertices in each part, regular of degree $q + 1$, which has girth 12.*

We can now easily deduce the following corollary.

Corollary 2. *There are at least $2^{\Omega(m^{\frac{6}{5}})}$ bipartite graphs with bipartitions whose parts are of sizes m, n , which are connected, whose bipartite complements are connected, and which have girth at least 12.*

Proof. Let q be the greatest power of 3 such that $q^5 + q^4 + q^3 + q^2 + q + 1$ is not larger than either m or n . Then let $G[X, Y]$ be a graph obtained by adding sufficient vertices to the graph B given by Theorem 1 to ensure that the parts are of sizes m and n respectively, and sufficient edges to ensure that $G[X, Y]$ is connected, while creating no new cycles. This graph has at least $q^6 = \Omega(m^{\frac{6}{5}})$ edges, and girth 12. It is trivial to check that $G[X, Y]$ must have connected bipartite complement. Let T be a spanning tree of $G[X, Y]$. Then every spanning subgraph of $G[X, Y]$ which preserves the edges of T has girth at least 12, is connected, and has connected bipartite complement. There are at least $q^6 - m - n + 1 = \Omega(m^{\frac{6}{5}})$ edges of $G[X, Y]$ which are not edges of T , and hence there are $2^{\Omega(m^{\frac{6}{5}})}$ such graphs. \square

Although we do not need the connectedness part of the above corollary at this stage, it will be useful in a later section.

Corollary 2 provides a lower bound on $Forb_{m,n}(H)$ for all H which contain a cycle of length less than 12, or whose bipartite complement contains such a cycle. The following corollary allows us to list all the H which do not fall into that category.

Corollary 3. *If $H = H[U, V]$ is a bipartite graph on at least eight vertices, both of whose parts contain at least three vertices, then*

$$Forb_{m,n}(H) = 2^{\Omega(m^{\frac{6}{5}})} .$$

Proof. If H contains a cycle, then either the shortest cycle in H is of length at most 8, or the bipartite complement of H contains a 4-cycle.

But if H is acyclic, then it has at most $|H| - 1$ edges, so its bipartite complement has at least $3(|H| - 3) - |H| + 1 = 2|H| - 8 > |H| - 1$ edges and must have a smallest subgraph which is a cycle; since H is acyclic this cycle is of length at most 8.

Therefore either H or its bipartite complement contains a cycle of length at most 8, and either the graphs given by Theorem 2 all do not contain a copy of H , or their bipartite complements all do not contain a copy of H . In either case, we obtain the given bound. \square

We now have to deal only with those H whose smaller part has zero, one or two vertices, together with a small number of exceptional cases on six and seven vertices. The various possibilities are set out in Table 1.

Trivially if one part of H is empty, then for sufficiently large m, n , $\text{Forb}_{m,n}(H) = 0$.

Theorem 4. $\text{Forb}_{m,n}(SI(k, l)) = m^{\max(k-1, l-1)m+o(m)}$.

Proof. A graph G with bipartition (X, Y) which does not contain a copy of $SI(k, l)$ is precisely one in which every vertex in X is either adjacent to at most $k-1$ vertices in Y , or to all but at most $l-1$ vertices in Y . There are

$$\left(\binom{n}{0} + \binom{n}{1} + \dots + \binom{n}{k-1} + \binom{n}{l-1} + \binom{n}{l-2} + \dots + \binom{n}{0} \right)^m = m^{\max(k-1, l-1)m+o(m)}$$

such graphs (note that $n = \Theta(m)$, so that $n^m = m^{m+o(m)}$). \square

3 Four infinite families

We now consider bipartite graphs $H = H[W, Z]$ with two vertices in the lower part W .

Observe that if the two vertices in the lower part have more than one common neighbour, or there are two isolated vertices in the upper part, then either H or its bipartite complement contains a cycle and so Theorem 2 gives us a lower bound on $\text{Forb}_{m,n}(H)$.

Therefore we need to find bounds for the four infinite families of bipartite graphs $DS(k, l)$, $DS^*(k, l)$, $JS(k, l)$ and $JS^*(k, l)$ (see Table 1). Note that the bipartite complement of $JS(k, l)$ is $DS^*(k, l)$, so that the bounds which we find for the former give immediately bounds for the latter.

Observe that if $G[X, Y]$ does not contain a copy of $DS(k, l)$, $l < k$, then it certainly contains no copy of $DS(k, k)$, so that it suffices to bound above $\text{Forb}_{m,n}(DS(k, k))$.

Theorem 5. $\text{Forb}_{m,n}(DS(k, k)) \leq m^{km+n+o(m)}$.

Proof. We describe a process for recording information sufficient to reconstruct a bipartite graph $G[X, Y]$ containing no copy of $DS(k, k)$.

Choose any order x_1, x_2, \dots, x_m on X such that $d(x_i) \leq d(x_j)$ for every $1 \leq i < j \leq m$.

It is obvious that G contains no copy of $DS(k, k)$ if and only if $|\Gamma(x_i) - \Gamma(x_j)| \leq k-1$ for each $i \leq j$.

For each $2 \leq i \leq m$, let $U_{x_i} = \Gamma(x_{i-1}) - \Gamma(x_i)$, and let $V_{x_i} = \Gamma(x_i) - \Gamma(x_{i-1})$. Let $U_{x_1} = \emptyset$, and $V_{x_1} = \Gamma(x_1)$.

We call the sets U_{x_i} and V_{x_i} the *removed set* and *added set* at x_i .

It is clear that the following information, the *basic recording* of G , is sufficient to reconstruct G :

$$(X, Y) \\ [V_{x_1}, x_1, V_{x_2}, x_2, \dots, V_{x_m}, x_m] \\ [U_{x_1}, U_{x_2}, \dots, U_{x_m}]$$

where we write out the elements of each of the sets in the standard order. We call the first list $[V_{x_1}, x_1, \dots]$ the *list of vertices*, and the second list $[U_{x_1}, \dots]$ the *list of removals*.

Observe that the list of vertices is of length at most $m+n+(k-1)m$, since $n \geq |\Gamma(x_m)| = \sum_i (|V_{x_i}| - |U_{x_i}|) \geq \sum_i |V_{x_i}| - (k-1)m$.

This is already sufficient to give $\text{Forb}_{m,n}(DS(k, k)) \leq 2^{m+n}(m+n)^{km+n} \binom{n}{k-1}^{m-1} = m^{(2k-1)m+n+o(m)}$, despite only using the fact that consecutive members x_i, x_{i+1} of X may not be the lower part of a copy of $DS(k, l)$.

In fact, no two members of X are the lower part of a copy of $DS(k, k)$. We can use this to show that, given the list of vertices, there are not $m^{(k-1)m+o(m)}$ choices for the list of removals, but only $m^{o(m)}$. Suppose that y appears in a removed set at some vertex between x_{i+1} and x_j , $i < j$, in the degree sequence order, but not in any added set at those vertices. Then y is adjacent to x_i but not to x_j . Since x_i and x_j are not the lower part of a copy of $DS(k, k)$, $|\Gamma(x_i) - \Gamma(x_j)| \leq k-1$. So we expect to find that most members of removed sets must also be members of added sets at nearby vertices in the degree sequence order.

We *compress* the information given in the removed sets U_{x_i} . Suppose that y is the j th member of the removed set at the vertex x_i . We define a *reference tag* $R_{x_i, j}$ as follows.

If there is a p , $-\log m \leq p \leq \log m$, such that the entry p after x_i in the list of vertices is y , then let $R_{x_i, j} = V : p$. We say that the reference tag is a *good reference tag*.

If there is no such p , then let $R_{x_i, j} = P : y$. We say that this is a *bad reference tag*.

We now write out the *compressed recording* of G :

$$(X, Y) \\ [V_{x_1}, x_1, V_{x_2}, x_2, \dots, V_{x_m}, x_m] \\ [(R_{x_1, 1}, R_{x_1, 2}, \dots), (R_{x_2, 1}, R_{x_2, 2}, \dots), \dots]$$

It is clear that this recording gives enough information to reconstruct the basic recording, and hence G .

We will now show that for any $G[X, Y]$ with no copy of $DS(k, k)$, there are few bad reference tags.

We divide X into *blocks* A_1, \dots as follows. Let $A_1 = \{x_1, x_2, \dots, x_a\}$ where x_a is within distance $\log m$ of x_1 in the list of vertices, but x_{a+1} is not. Let $A_2 = \{x_{a+1}, \dots, x_b\}$, where x_b is within distance $\log m$ of x_{a+1} in the list of vertices, but x_{b+1} is not, and so on. Since the list of vertices is of length at most $km+n$, there are at most $\lceil \frac{km+n}{\log m} \rceil$ blocks.

Suppose that $R_{x_i, j} = P : y$ is a bad reference tag: so y is in the removed set at x_i , but it does not appear in the list of vertices within $\log m$ of x_i . If $x_i \in A_r = \{x_c, \dots, x_d\}$, then

y does not appear in an added set at any of x_{c+1}, \dots, x_d . If $x_i \neq x_c$, then y is adjacent to x_c , but not to x_d . If there were k bad reference tags among those at vertices x_{c+1}, \dots, x_d then there would be k vertices in Y adjacent to x_c and not to x_d . This would mean that $|\Gamma(x_c) - \Gamma(x_d)| \geq k$, so $\{x_c, x_d\}$ would be the lower part of a copy of $DS(k, k)$. Therefore there can be at most $2(k-1)$ bad reference tags in a block (at most $k-1$ at the first vertex in the block, and at most $k-1$ among those at the remaining vertices). Therefore there are at most $\frac{2k(km+n)}{\log m}$ bad reference tags.

There are $(1 + 2 \log m)$ possible good reference tags, and n possible bad ones. Therefore we can bound above the number of possible compressed recordings by

$$2^{m+n} (m+n)^{km+n} 2^{(k-1)m} (1 + 2 \log m)^{(k-1)m} n^{\frac{2k(km+n)}{\log m}}$$

so $Forb_{m,n}(DS(k, k)) \leq m^{km+n+o(m)}$ as required. \square

The upper bound in Theorem 5 gives the correct speed.

Theorem 6. $Forb_{m,n}(DS(k, k)) = m^{km+n+o(m)}$.

Proof. We have the upper bound already; we construct a family of graphs which is of sufficient size.

Let $X = \{1, \dots, m\}$, $Y = \{m+1, \dots, m+n\}$. Let $X_0 = \{1, \dots, \lfloor \frac{n}{\log m} \rfloor\}$. Let $Y_0 = \{m+1, \dots, m + \lfloor \frac{m}{\log m} \rfloor\}$.

Partition $X - X_0$ into sets X_1, X_2, \dots , each (except possibly the last) of size $\lfloor \log m \rfloor$. We can obtain such a partition by taking any order on $X - X_0$, which has size $m - \lfloor \frac{m}{\log m} \rfloor$, and letting X_1 be the first $\lfloor \log m \rfloor$ vertices in that order, X_2 the next $\lfloor \log m \rfloor$, and so on. There are $(m - \lfloor \frac{m}{\log m} \rfloor)! = m^{m-o(m)}$ ways to order $X - X_0$. The number of distinct orders which generate each partition is $|X_1|! |X_2|! \dots \leq \lfloor \log m \rfloor!^{\lfloor \frac{m}{\log m} \rfloor + 1} = m^{o(m)}$. Therefore there are $m^{m-o(m)}$ such distinct partitions.

Partition $Y - Y_0$ into sets Y_1, Y_2, \dots , each (except possibly the last) of size $\lfloor \log m \rfloor$. Similarly, there are $n^{n-o(n)}$ ways to do this.

Choose, for each vertex x_i in $X - X_0$, a set N_i of $k-1$ vertices in $Y - Y_0$. There are $n^{(k-1)(m - \lfloor \frac{m}{\log m} \rfloor) - o(m)} = m^{(k-1)m+o(m)}$ ways to do this.

Construct a bipartite graph $G[X, Y]$ as follows. Put an edge from each $i \in X_0$ to each vertex in $Y_0 \cup Y_1 \cup \dots \cup Y_{i-1}$. Put an edge from each $m+i \in Y_0$ to each vertex in $X_0 \cup X_1 \cup \dots \cup X_{i-1}$. Put an edge from each $i \in X - X_0$ to each vertex in N_i .

Observe that whatever choices were made, G does not contain a copy of $DS(k, k)$. Furthermore, different choices imply different G . Therefore $Forb_{m,n}(DS(k, k)) = m^{km+n+o(m)}$ as required. \square

Observe that if the recording method described in Theorem 5 were applied to a typical graph $G[X, Y]$ constructed as in Theorem 6, then given any $\epsilon > 0$ we would find the following.

There are no sets V_x of size greater than ϵn .

The list of vertices is of length at least $(k - \epsilon)m + n$.

There are at most ϵm vertices in X with any given degree.

There are at least $m^{1-\epsilon}$ different vertex degrees in X .

It is easy to check, by considering the recording method, that given $\epsilon > 0$, the speed of graphs $G[X, Y]$ which do not contain a copy of $DS(k, k)$ and which fail to satisfy any of the above conditions is at most $m^{km+n-\epsilon'm+o(m)}$, slower than the speed of $Forb_{m,n}(DS(k, k))$. In the first two cases, this is because there are not enough possibilities for the list of vertices, and in the last two, because there are $m^{\epsilon m}$ distinct orderings of X by increasing degree, so that each graph can be recorded in $m^{\epsilon m}$ different ways. So the graphs constructed in Theorem 6 are in some sense typical.

Since $K_{1,k} = SI(k, 0)$ is an induced subgraph of $DS(k, l)$, any $G[X, Y]$ which does not contain $SI(k, 0)$ does not contain $DS(k, l)$, so we have the lower bound $Forb_{m,n}(DS(k, l)) \geq m^{(k-1)m+o(m)}$. It is trivial to check that in the case $k \geq 2, l = 0$, this lower bound gives the correct speed.

Note that, if $1 \leq l \leq k - 1$, since $DS(l, l)$ is an induced subgraph of $DS(k, l)$, the construction in Theorem 6 gives a lower bound $Forb_{m,n}(DS(k, l)) \geq m^{\max(lm+n, (k-1)m)+o(m)}$. When $l = k - 1$ this bound is certainly better than the above, and it seems reasonable to conjecture that it is correct.

We now examine $JS(k, l)$. We will obtain an upper bound by modifying the argument used in Theorem 5; again we will find an upper bound on $Forb_{m,n}(JS(k, k))$ and observe that as $JS(k, l)$ is an induced subgraph of $JS(k, k)$ when $k \geq l$, this gives an upper bound for $Forb_{m,n}(JS(k, l))$.

Theorem 7. $Forb_{m,n}(JS(k, k)) \leq m^{km+n+o(m)}$.

Proof. Again we will describe a process for recording bipartite graphs $G[X, Y]$ which contain no copy of $JS(k, k)$. Observe that if we have some guarantee that some vertices in X share a common neighbour in Y , then we can apply the same recording procedure as in Theorem 5 to these vertices.

Observe that $G[X, Y]$ contains no copy of $JS(k, k)$ if and only if whenever $x, x' \in X$ share a common neighbour, with $d(x) \leq d(x')$, so $|\Gamma(x) - \Gamma(x')| \leq k - 1$.

It is convenient to record the graph G in several steps. First we find a way to record the neighbours of the set Q of vertices in X which have at most $\log m$ neighbours.

We do this as follows. First we construct a set $P \subset Q$ by reading through the vertices in Q in order of decreasing degree, and choosing for P every vertex whose neighbourhood is disjoint from all those previously chosen. Now any two vertices in P have disjoint neighbourhoods, and if $q \in Q - P$, then there is a $p \in P$ whose neighbourhood intersects that of q and which has $d(p) \geq d(q)$.

Let $\Gamma(P)$ be the set of vertices in Y which are neighbours of at least one vertex in P . Then we can record the neighbours of each vertex in P by writing down $\Gamma(P)$ and the partition

of $\Gamma(P)$ into the sets $\Gamma(p)$ for $p \in P$.

Now let q be in $Q - P$. There is $p \in P$ with $d(p) \geq d(q)$ and such that p and q share at least one neighbour. Then $|\Gamma(q) - \Gamma(p)| \leq k - 1$, since $\{p, q\}$ is not the lower part of a copy of $JS(k, l)$. So we can record the neighbours of q by writing down the vertex p , the neighbours of p which are also neighbours of q , and the at most $k - 1$ vertices in $\Gamma(q) - \Gamma(p)$. This does not require us to have the vertices in $Q - P$ in any particular order, so we can record the set $Q - P$ by simply choosing them from X .

So we can record the neighbours of all the vertices in Q in at most

$$2^n m^{|\Gamma(P)|} 2^m (m 2^{\log \log m} n^{k-1})^{|Q-P|} = m^{k|Q|+|\Gamma(P)|-k|P|+o(m)}$$

ways.

Now we record the neighbours of the remaining vertices $X' = X - Q$, each of which has degree at least $\log \log m > 2k - 1$.

We choose a set of vertices $S_1 \subset X'$ by reading through X' in order of increasing degree, and choosing for S_1 every vertex whose neighbours are disjoint from all those previously chosen. Let $X_1 = X' - S_1$. Now S_1 satisfies three properties. First, no two vertices in S_1 share a common neighbour. Second, every vertex in X_1 shares at least one common neighbour with some vertex in S_1 . Third, for every $x \in X_1$, there is an $s \in S_1$ which shares a common neighbour with x and satisfies $d(s) \leq d(x)$.

Observe that since $G[X, Y]$ contains no copy of $JS(k, k)$ and all vertices in X' have degree at least $\log \log m > 2k - 1$, these three properties imply that for every $x \in X_1$, every $s \in S_1$ which shares a common neighbour with x satisfies $d(s) \leq d(x)$. For if not, then let $s \in S_1$ be a vertex sharing a common neighbour with x and with $d(x) < d(s)$. Since x shares a neighbour with, and has degree not smaller than, some $s' \in S_1$, we must have $|\Gamma(s') - \Gamma(x)| \leq k - 1$ or $\{x, s'\}$ would be the bottom part of a copy of $JS(k, k)$. Then x has at least k neighbours in common with s' , none of which are neighbours of s . So $|\Gamma(x) - \Gamma(s)| \geq k$, but then $\{x, s\}$ are the bottom part of a copy of $JS(k, k)$.

We assign to the vertices in X' *removed sets* and *added sets* U_x , and V_x by following the process below.

For each $s \in S_1$, let $U_s = \emptyset$ and let $V_s = \Gamma(s)$.

Let x_1 be a vertex in X_1 with minimum degree. We distinguish two possibilities.

If x_1 shares a common neighbour with only one $s_1 \in S_1$, then $d(x_1) \geq d(s_1)$ and we can write $\Gamma(x_1) = (\Gamma(s_1) - U_{x_1}) \cup V_{x_1}$, where as before $|V_{x_1}| \geq |U_{x_1}| \leq k - 1$. We let $S_2 = (S_1 - \{s_1\}) \cup \{x_1\}$, and $X_2 = X_1 - \{x_1\}$. We say that x_1 is part of the *degree sequence process* starting at s_1 .

If x_1 shares a common neighbour with more than one member of S , then let these members be s_1, \dots, s_a . Let $U_{x_1} = (\Gamma(s_1) \cup \Gamma(s_2) \cup \dots \cup \Gamma(s_a)) - \Gamma(x_1)$, and let $V_{x_1} = \Gamma(x_1) - (\Gamma(s_1) \cup \dots \cup \Gamma(s_a))$. Observe that $|U_{x_1}| \leq a(k - 1)$, since none of the sets $\Gamma(s_i) - \Gamma(x)$ have more than $k - 1$ members. We let $S_2 = (S_1 - \{s_1, \dots, s_a\}) \cup \{x_1\}$, and $X_2 = X_1 - \{x_1\}$. We say that the vertex x_1 *joins the neighbourhoods* of the vertices s_1, \dots, s_a .

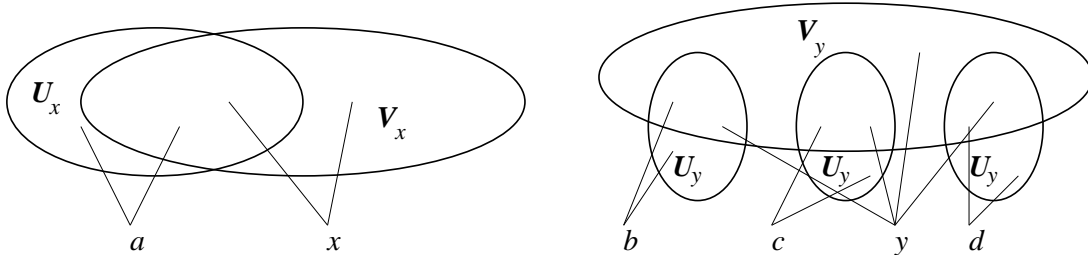


Figure 2: x follows a in a degree sequence process; y joins the neighbourhoods of b , c and d .

By construction, no two vertices in S_2 share a common neighbour. If $x \in X_2$ shares a common neighbour with $s \in S_2$, then either $s \in S_1$, in which case $d(x) \geq d(s)$, or $s = x_1$, in which case $d(x) \geq d(s)$ by choice of x_1 . If $x \in X_2$, then x shares a common neighbour with $s \in S_1$. Either $s \in S_2$, or s shares a common neighbour with x_1 . In the latter case, both x and x_1 have degree at least $d(s) > \log \log m > 2k - 1$, so that x and x_1 are each adjacent to all but at most $k - 1$ neighbours of s , and so must share a common neighbour. Therefore S_2 and X_2 satisfy the same conditions as S_1 and X_1 , so we can continue this process with x_2 , a vertex in X_2 with minimum degree, and the set S_2 , and so on.

If we know that x follows a in a degree sequence process, then we can recover the neighbours of x given $\Gamma(a)$, U_x and V_x .

If we know that y joins the neighbourhoods of b, \dots, d , then we can recover the neighbours of y given $\Gamma(b), \dots, \Gamma(d)$, U_y and V_y .

Then we can write down a recording of $G[X, Y]$ as in the following example.

(X, Y)
Recording of the low degree vertices and their neighbours
 $[V_{s_1}, s_1, V_{x_1}, x_1, \dots]$
 $[U_{s_1}, U_{x_1}, \dots]$
 \dots
 $[V_{s_{|S_1|}}, s_{|S_1|}, \dots]$
 $[U_{s_{|S_1|}}, \dots]$
JOIN : b, \dots, d
 $[V_y, y, \dots]$
 $[U_y, \dots]$
JOIN : \dots
 \dots

Each of the pairs of lines $[V_{s_1}, s_1, V_{x_1}, x_1, \dots]$, $[U_{s_1}, U_{x_1}, \dots]$ et cetera represents a degree sequence process as in Theorem 5; so the neighbourhood of s_1 is V_{s_1} , the neighbourhood of x_1 is $\Gamma(s_1) \cup V_{x_1} - U_{x_1}$, and so on. The ordering of the degree sequence processes is immaterial.

Each triple of lines **JOIN** : b, \dots, d , $[V_y, y, \dots]$, $[U_y, \dots]$ et cetera represents a new degree sequence process; in the example, the first vertex in this degree sequence process is y , whose neighbourhood is $(\Gamma(b) \cup \dots \cup \Gamma(d)) - U_y \cup V_y$. Again the ordering of these triples is

immaterial.

As in Theorem 5 we call the lists $[V_{s_1}, s_1, V_{x_1}, x_1, \dots]$ et cetera the *lists of vertices* and the lists $[U_{s_1}, U_{x_1}, \dots]$ et cetera the *lists of removals*.

It is clear that we can reconstruct G from such a recording; we call this the *basic recording* of G .

Observe that $|S_1| \leq \frac{n}{\log \log m}$, since every member of S_1 has at least $\log \log m$ neighbours. If S_{i+1} is obtained from S_i by joining the neighbourhoods of j vertices, then $|S_{i+1}| = |S_i| + 1 - j$. Since $|S_1| \leq \frac{n}{\log \log m}$, the total number of neighbourhoods joined is at most $\frac{2n}{\log \log m}$.

Let $\Gamma(X')$ be the set of vertices in Y which are adjacent to at least one vertex in X' . The neighbourhoods of the vertices in S_i are disjoint for each i ; so the sum of their sizes is at most $|\Gamma(X')| \leq n$. Observe that whether S_{i+1} is obtained from S_i by letting x_i continue a degree sequence process or by letting it join some neighbourhoods,

$$\sum_{s \in S_{i+1}} |\Gamma(s)| = |V_{x_i}| - |U_{x_i}| + \sum_{s \in S_i} |\Gamma(s)| .$$

Now $|U_{x_i}| \leq k - 1$ if x_i continues a degree sequence process; if x_i joins some r neighbourhoods then $|U_{x_i}| \leq r(k - 1)$. Since at most $\frac{2n}{\log \log m}$ neighbourhoods are joined in total,

$$\sum_{x \in X'} |V_x| \leq |\Gamma(X')| + (k - 1)|X'| + \frac{2kn}{\log \log m} .$$

Therefore the total length of the lists of vertices is at most $|\Gamma(X')| + k|X'| + \frac{2kn}{\log \log m}$, and the total length of the lists of removals is at most $(k - 1)|X'| + \frac{2kn}{\log \log n}$. The total number of vertices whose neighbourhoods are joined (and which are therefore listed on some JOIN : line in the recording) is at most $\frac{2n}{\log \log m}$.

This is already sufficient to give

$$\begin{aligned} \text{Forb}_{m,n}(JS(k, l)) &\leq 2^{m+n} m^{k|Q|+|\Gamma(P)|-k|P|+o(m)} (m+n)^{|\Gamma(X')|+k|X'|+\frac{2kn}{\log \log m}} n^{(k-1)|X'|+\frac{2kn}{\log \log n}} m^{\frac{2n}{\log \log m}+o(m)} \\ &= m^{k|Q|+|\Gamma(P)|-k|P|+|\Gamma(X')|+(2k-1)|X'|+o(m)} \leq m^{(2k-1)m+2n+o(m)} . \end{aligned}$$

As in Theorem 5, we expect to find that vertices appearing in U_{x_i} are likely to appear in V_{x_j} for some x_j close to x_i in the same degree sequence process. We can make this precise by applying a virtually identical compression argument. We define the reference tag $R_{x_i, j}$ in the same way as in that theorem, with reference to the list of vertices which contains x_i .

We can again divide X' into blocks, with each block containing vertices in just one degree sequence process. If a block starts at a vertex x which joins the neighbourhoods of r vertices, then it may contain at most $k - 1 + r(k - 1)$ bad reference tags; otherwise a block may contain at most $2(k - 1)$ bad reference tags.

The total length of the lists of vertices is less than $2(km + n)$, so that there are at most $\frac{2(km+n)}{\log m} + \frac{2n}{\log \log m}$ blocks, the extra $\frac{2n}{\log \log m}$ coming from possible ‘short’ blocks at the ends of degree sequence processes. Therefore there are at most $\frac{3n}{\log \log m}$ bad reference tags in total.

As in Theorem 5, we can now write the *compressed recording* of G , where instead of writing the lists of removals $[U_x, \dots]$ et cetera, we write lists of reference tags $[(R_{x,1}, \dots), \dots]$ et cetera.

This allows us to improve our bound for $\text{Forb}_{m,n}(JS(k, l))$; instead of bounding above the choices for the lists of removals by $m^{(k-1)|X'|+o(m)}$, we can now bound above the choices for the lists of removals by $m^{o(m)}$. We find that

$$\text{Forb}_{m,n}(JS(k, l)) \leq m^{k|Q|+|\Gamma(P)|-k|P|+|\Gamma(X')|+k|X'|+o(m)} \leq m^{km+2n+o(m)} .$$

Finally, we wish to obtain the claimed bound. We use our knowledge of the neighbours of vertices in P to produce an *extra-compression* of the lists of vertices.

For each $p \in P$, either we can find an $x_p \in X'$ which is the first vertex in the lists of vertices to share a common neighbour with p , or $\Gamma(p) \cap \Gamma(X') = \emptyset$. Let P_1 be the set of vertices $p \in P$ for which x_p exists, and $P_2 = P - P_1$ be the vertices whose neighbourhoods are disjoint from $\Gamma(X')$.

For each $p \in P_1$, let $I_{p,x_p} = \Gamma(p) \cap \Gamma(x_p)$. Since $d(x_p) > \log \log m \geq d(p)$, $|I_{p,x_p}| \geq |\Gamma(p)| - (k-1)$.

For each $x \in X'$, if $x \neq x_p$ for every $p \in P_1$, let $V'_x = V_x$. If $x = x_p$ for at least one p , let

$$V'_x = V_x - \bigcup_{p:x=x_p} I_{p,x_p} .$$

We write down the *extra-compressed recording* of G as in the following example.

```

{X, Y}
Recording of the low degree vertices and their neighbours
[I_{p_1, x_{p_1}}, x_{p_1}, I_{p_2, x_{p_2}}, x_{p_2}, \dots]
[V'_{s_1}, s_1, V'_{x_1}, x_1, \dots]
[(R_{s_1,1}, \dots), \dots]
...
[V'_{s_{|S_1|}}, s_{|S_1|}, \dots]
[(R_{s_{|S_1|},1}, \dots), \dots]
JOIN :b, \dots, d
[V'_y, y, \dots]
[(R_{y,1}, \dots), \dots]
JOIN : \dots
...

```

where $P_1 = \{p_1, p_2, \dots\}$ with $p_1 < p_2 < \dots$ in the standard order. We can clearly recover the compressed recording of G from this; we have only to insert each of the sets $I_{p_i, x_{p_i}}$ into the identified $V'_{x_{p_i}}$. Therefore $\text{Forb}_{m,n}(JS(k, l))$ is bounded above by the number of possible extra-compressed recordings.

We now wish to find the total length of the lists of vertices in the extra-compressed

recording of $G[X, Y]$. Recall that

$$\sum_{x \in X'} |V_x| \leq |\Gamma(X')| + (k-1)|X'| + \frac{2kn}{\log \log m}.$$

Observe that

$$\sum_{x \in X'} |V'_x| = \sum_{x \in X'} |V_x| - \sum_{p \in P_1} |I_p| \leq \sum_{x \in X'} |V_x| + (k-1)|P_1| - \sum_{p \in P_1} |\Gamma(p)|,$$

and

$$|\Gamma(X')| \leq n - \sum_{p \in P_2} |\Gamma(p)|.$$

Then the total length of the lists of vertices in the extra-compressed recording is at most $|X'| + n + (k-1)|P| - \sum_{p \in P} |\Gamma(p)| + (k-1)|X'| + \frac{2kn}{\log \log m}$.

The *list of insertions* $[I_{p_1, x_{p_1}}, x_{p_1}, \dots]$ can be chosen in at most $(2 \binom{\log \log n}{k-1} m)^{|P|} = m^{|P|+o(m)}$ ways.

Finally, we can obtain the claimed bound:

$$\begin{aligned} \text{Forb}_{m,n}(JS(k, l)) &\leq 2^{m+n} m^{k|Q|+|\Gamma(P)|-k|P|+o(m)} m^{|P|+o(m)} m^{k|X'|+n+(k-1)|P|-|\Gamma(P)|+o(m)} \\ &\leq m^{km+n+o(m)}. \end{aligned}$$

□

As $DS(k, k)$ is an induced subgraph of $JS(k, k)$, the family of graphs given in Theorem 6 provides a lower bound for $JS(k, k)$ which matches the upper bound, so $\text{Forb}_{m,n}(JS(k, k)) = m^{km+n+o(m)}$.

Corollary 8. $\text{Forb}_{m,n}(DS^*(k, l)) \leq m^{km+n+o(m)}$.

Proof. The bipartite complement of $DS^*(k, l)$ is $JS(k, l)$, so $\text{Forb}_{n,m}(DS^*(k, l)) \leq m^{km+n+o(m)}$. □

Again we observe that $\text{Forb}_{m,n}(DS^*(k, k)) = m^{km+n+o(m)}$.

Corollary 9. $\text{Forb}_{m,n}(JS^*(k, l)) \leq m^{km+2n+o(m)}$.

Proof. Let $G = G[X, Y]$ be a bipartite graph not containing $JS^*(k, l)$.

Let Y' be Y if $|Y|$ is odd, and $Y - \{y\}$, some $y \in Y$, if $|Y|$ is even.

Let X' be the vertices in X with less than $\frac{|Y'|}{2}$ neighbours in Y' , and X'' those with more than $\frac{|Y'|}{2}$ neighbours in Y' . Let $m' = |X'|$, and $m'' = |X''|$.

Observe that the neighbourhoods of any two vertices in X' cover at most $|Y'| - 1$ vertices. Therefore the subgraph of $G[X, Y]$ induced by $X' \cup Y'$ contains no copy of $JS(k, l)$. Similarly, the subgraph of $G[X, Y]$ induced by $X'' \cup Y'$ contains no copy of $DS^*(k, l)$.

Therefore $\text{Forb}_{m,n}(H) \leq 2^m (m')^{km'+n+o(m')} (m'')^{km''+n+o(m'')} \leq m^{km+2n+o(m)}$. □

4 Second and third problems

We have now established good bounds on $Forb_{m,n}(H[U, V])$ for every bipartite graph $H[U, V]$ except for the six exceptional graphs. It is convenient to use these bounds to find good bounds on $Forb_n(H[U, V])$ and $Forb_n^*(H)$ at this point.

Note that if G is a bipartite graph which has a bipartition (X, Y) , then the statement that no bipartition of G contains a copy of $H[U, V]$ is certainly at least as strong as the statement that both $G[X, Y]$ contains no copy of $H[U, V]$ and $G[Y, X]$ contains no copy of $H[U, V]$. Then it is trivial that

$$Forb_n^*(H) \leq Forb_n(H[U, V]) \leq 2^n \max_{r < n} \min(Forb_{r, n-r}(H[U, V]), Forb_{n-r, r}(H[U, V])) , \quad (1)$$

so that we can obtain an upper bound for $Forb_n(H[U, V])$ by finding the worst case of $\min(Forb_{r, n-r}(H[U, V]), Forb_{n-r, r}(H[U, V]))$, and an upper bound for $Forb_n^*(H)$ by finding the worst case of $Forb_n(H[U, V])$. We will see that these worst cases are, respectively, the case $r = \frac{n}{2}$ and H as drawn in Table 1.

Observe that the condition ‘ G with any bipartition does not contain a copy of H with any bipartition’ is in general significantly stronger than ‘ $G[X, Y]$ does not contain a copy of $H[U, V]$ ’, so we might expect the upper bounds obtained from the above inequality to be poor. This is not the case.

Theorem 10. *If H contains a cycle, or all of its bipartite complements contain a cycle, then $Forb_n^*(H) = 2^{\Omega(n^{\frac{6}{5}})}$. If $H[U, V]$ or its bipartite complement contains a cycle then $Forb_n(H[U, V]) = 2^{\Omega(n^{\frac{6}{5}})}$.*

Proof. If H contains a cycle or all of its bipartite complements contain a cycle, then either it has girth at most eight, or all of its bipartite complements have girth at most eight. By Corollary 2 there are $2^{\Omega(n^{\frac{6}{5}})}$ bipartite graphs on n vertices which are connected, have connected bipartite complement and girth at least 12. In the first case, all of these graphs contain no copy of H ; in the second case, the unique connected bipartite complement of each of these graphs contains no copy of H . \square

We now have only to establish appropriate lower bounds on $Forb_n^*(H)$ for the five infinite families $SI(k, l)$, $DS(k, l)$, $DS^*(k, l)$, $JS(k, l)$ and $JS^*(k, l)$ to match those we have for $Forb_{m,n}(H[U, V])$. Again, we observe that both $SI(k, 0)$ and $DS(l, l)$ are induced subgraphs of each of $DS(k, l)$, $DS^*(k, l)$, $JS(k, l)$ and $JS^*(k, l)$ for $l \leq k$; so it suffices to find lower bounds on $Forb_n^*(H)$ for $SI(k, l)$ and $DS(k, k)$.

Theorem 11. *For any fixed r , there are at least $n^{\frac{rn}{2} + o(n)}$ bipartite graphs whose maximum degree is r and in which no two vertices have three or more common neighbours.*

Proof. Let G be a bipartite graph on n vertices obtained by choosing uniformly at random r matchings from $\{1, \dots, \lfloor \frac{n}{2} \rfloor\}$ to $\{\lfloor \frac{n}{2} \rfloor + 1, 2\lfloor \frac{n}{2} \rfloor\}$ and putting an edge in G whenever that edge is present in any of the r matchings.

We call x, y a *problem pair* if x and y have at least three common neighbours.

The probability that a given pair of vertices $x, y \leq \frac{n}{2}$ is a problem pair is at most

$$\sum_{i=3}^r \binom{r}{i} \left(\frac{2r}{n}\right)^i,$$

so the expected number of problem pairs in G is at most $2^{\lfloor \frac{n}{2} \rfloor}$ times that. For sufficiently large n , this quantity is smaller than $\frac{1}{2}$, so there are at least

$$\frac{\lfloor \frac{n}{2} \rfloor!^r}{2} = n^{\frac{rn}{2} + o(n)}$$

choices of r matchings which give rise to graphs G with no problem pairs. Now given such a graph G , each of the at most $\frac{rn}{2}$ edges is present in some subset of the r matchings, so there are at most $(2^r)^{rn} = n^{o(n)}$ distinct sets of r matchings giving rise to G . Thus we have at least $n^{\frac{rn}{2} + o(n)}$ distinct graphs G with maximum degree r and no two vertices having three or more common neighbours, as required. \square

Corollary 12.

$$Forb_n^*(SI(k, l)) = n^{\max(\frac{(k-1)n}{2}, \frac{(l-1)n}{2}) + o(n)},$$

and

$$Forb_n(SI(k, l)) = n^{\max(\frac{(k-1)n}{2}, \frac{(l-1)n}{2}) + o(n)}.$$

Proof. The upper bound follows from Theorem 4 and the inequality (1), since

$$Forb_{\frac{n}{2}, \frac{n}{2}}(SI(k, l)) = n^{\max(\frac{(k-1)n}{2}, \frac{(l-1)n}{2}) + o(n)}$$

is the worst case.

For the lower bound, if $k \geq l$, by Theorem 11 we can find $n^{\frac{(k-1)n}{2} + o(n)}$ bipartite graphs which have maximum degree $k - 1$ and which therefore do not contain a copy of $SI(k, l)$ with any bipartition. If on the other hand $k < l$, then again by Theorem 11 we can find $n^{\frac{(l-1)n}{2} + o(n)}$ bipartite graphs which have maximum degree $l - 1$ and in which no two vertices have three or more common neighbours. Now observe that although $SI(k, l)$ has several bipartitions, and hence several bipartite complements, all of the bipartite complements of $SI(k, l)$ have either a vertex of degree l or two vertices sharing three common neighbours. So there are $n^{\frac{(l-1)n}{2} + o(n)}$ bipartite graphs which do not contain a copy of any of the bipartite complements of $SI(k, l)$, and there must be $n^{\frac{(l-1)n}{2} + o(n)}$ bipartite graphs which do not contain a copy of $SI(k, l)$. This gives us the required inequality

$$n^{\max(\frac{(k-1)n}{2}, \frac{(l-1)n}{2}) + o(n)} \leq Forb_n^*(SI(k, l)) \leq Forb_n(SI(k, l)) \leq n^{\max(\frac{(k-1)n}{2}, \frac{(l-1)n}{2}) + o(n)}.$$

\square

We now have only to bound $Forb_n^*(DS(k, k))$. We use a similar construction to that in Theorem 6.

Theorem 13. $Forb_n^*(DS(k, k)) = n^{\frac{(k+1)n}{2} + o(n)}$.

Proof. The upper bound follows from the inequality (1) and Theorem 5.

For the lower bound, let $X = \{1, \dots, \lfloor \frac{n}{2} \rfloor\}$, $Y = \{\lfloor \frac{n}{2} \rfloor + 1, \dots, n\}$. Let $X_0 = \{1, \dots, \lfloor \frac{n}{2 \log n} \rfloor\}$. Let $Y_0 = \{\lfloor \frac{n}{2} \rfloor + 1, \dots, \lfloor \frac{n}{2} \rfloor + \lfloor \frac{n}{2 \log n} \rfloor\}$.

Partition $X - X_0$ into sets X_1, X_2, \dots , each (except possibly the last) of size $\lfloor \log n \rfloor$. As in Theorem 6, there are $n^{\frac{n}{2} - o(n)}$ such distinct partitions.

Partition $Y - Y_0$ into sets Y_1, Y_2, \dots , each (except possibly the last) of size $\lfloor \log n \rfloor$. Similarly, there are $n^{\frac{n}{2} - o(n)}$ ways to do this.

By Theorem 11 there are $n^{\frac{(k-1)n}{2} + o(n)}$ bipartite graphs with bipartition $\{X - X_0, Y - Y_0\}$ whose maximum degree is $k - 1$.

Construct a bipartite graph G as follows. Put an edge from each $i \in X_0$ to each vertex in $Y_0 \cup Y_1 \cup \dots \cup Y_{i-1}$. Put an edge from each $m + i \in Y_0$ to each vertex in $X_0 \cup X_1 \cup \dots \cup X_{i-1}$. Put edges between $X - X_0$ and $Y - Y_0$ in any way such that the maximum degree of the subgraph induced by $(X - X_0) \cup (Y - Y_0)$ is at most $k - 1$.

Observe that, whatever choices were made, G does not contain a copy of $DS(k, k)$ with any bipartition. Furthermore, different choices imply different G . Therefore $Forb_n^*(DS(k, k)) = n^{\frac{(k+1)n}{2} + o(n)}$ as required. \square

5 Exceptional graphs

The only bipartite graphs which we have not yet covered are those with three vertices in the smaller part which are acyclic and whose bipartite complements are acyclic. These are graphs on either six or seven vertices, shown in Table 3.

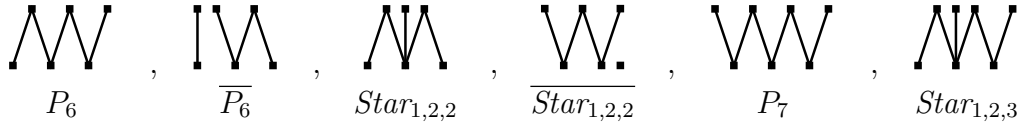


Table 3: Exceptional bipartite graphs

Note that the first pair of these are bipartite complements of each other, as are the second pair; the last two are both self-complementary.

Each of these exceptional bipartite graphs contains the graph P_4 , so that we have trivial lower bounds for each exceptional H .

For the first problem, we have $Forb_{m,n}(H) \geq Forb_{m,n}(P_4) = m^{m+n+o(m)}$. We will show that these lower bounds are correct for the graphs P_6 and $\overline{P_6}$.

For the second and third problems, again we have $Forb_n(H) \geq Forb_n(P_4) = n^{n+o(n)}$ and similarly $Forb_n^*(H) \geq n^{n+o(n)}$, and we will show that these lower bounds are correct for each of the exceptional graphs except P_7 .

Thus far, we have examined ways to record a bipartite graph one vertex at a time. An alternative method is to consider breaking a graph down into smaller pieces by specified operations. Results along these lines are called decomposition results, and there exist several relating to bipartite graphs.

Giakoumakis and Vanherpe [6] considered the two operations of bipartite complement and disjoint union. They defined the class of *bi-cographs* to be the class of bipartite graphs which can be fully decomposed using only these two operations: a single vertex in either part is a bi-cograph, the bipartite complement of a bi-cograph is a bi-cograph, and the disjoint union of two bi-cographs is a bi-cograph. They were able to prove that the class of bi-cographs is exactly the class of bipartite graphs which contain no induced P_7 , $Star_{1,2,3}$ or Sun_4 , where the graph Sun_4 is the bipartite graph on eight vertices given by taking a copy of C_4 and adding a matching from the vertices of the C_4 to the other four vertices.

Fouquet, Giakoumakis and Vanherpe [5] then introduced a further decomposition operation. Suppose that the bipartite graph $G[X, Y]$ is such that there exist non-trivial induced subgraphs $G_1[X_1, Y_1]$ and $G_2[X_2, Y_2]$, with $X = X_1 \sqcup X_2$ and $Y = Y_1 \sqcup Y_2$, and the subgraph of $G[X, Y]$ induced by (X_1, Y_2) is a complete bipartite graph, while that induced by (X_2, Y_1) is an empty graph. Then we say that $G_1[X_1, Y_1]$, $G_2[X_2, Y_2]$ is a $K+S$ -decomposition of $G[X, Y]$. They called decomposition by using the three operations of taking bipartite complement, disjoint union and $K+S$ -decomposition the *canonical decomposition* of a bipartite graph, and defined the class of *weak-bisplit graphs* to be those graphs which can be fully decomposed using the canonical decomposition. They proved that the weak-bisplit graphs are exactly those bipartite graphs containing no induced P_7 or $Star_{1,2,3}$.

A *prime* bipartite graph is one in which $\Gamma(x) = \Gamma(y)$ if and only if $x = y$.

Lozin [7] was able to prove that the class of prime bipartite graphs which can be decomposed to $K_{1,3}$ -free graphs using the canonical decomposition is exactly the class of $Star_{1,2,3}$ -free prime bipartite graphs. We will use this result to obtain our remaining upper bounds, so we state it explicitly and give a short proof, based on that of Lozin.

Theorem 14. *Any prime bipartite graph $G[X, Y]$ that does not contain an induced copy of $Star_{1,2,3}$ can be decomposed, using the three canonical decomposition operations, to the class of $K_{1,3}$ -free graphs (equivalently, to the class of paths and cycles).*

Proof. We observe that if $G[X, Y]$ is a bipartite graph, then $G[X, Y]$ has a $K+S$ -decomposition if and only if the bipartite complement of $G[X, Y]$ has one.

Suppose the theorem is false. Then there is a prime bipartite graph $G[X, Y]$ of minimal order which does not contain an induced $Star_{1,2,3}$ and which does not satisfy the conditions of the theorem.

Since, by the result of Fouquet, Giakoumakis and Vanherpe, every bipartite graph that is both $Star_{1,2,3}$ -free and P_7 -free can be fully decomposed by the canonical decomposition, $G[X, Y]$ cannot be P_7 -free. In particular, we may assume w.l.o.g. that the vertices $U = (1, 2, 3, 4, 5, 6, 7)$ induce a P_7 , in that order, with $1, 3, 5, 7 \in Y$ and $2, 4, 6 \in X$. We show, following the method of Lozin, that since $G[X, Y]$ is prime and $Star_{1,2,3}$ -free, the possibilities for edges from a vertex z to any induced P_7 are very limited.

Let V be any ordered set of seven vertices of $G[X, Y]$ inducing a P_7 in that order. Let T be a subset of $1, 2, 3, 4, 5, 6, 7$. Then let $S_V(T)$ be the set of vertices $z \in V(G[X, Y]) - V$ which are adjacent to the vertices in V at positions T , and not adjacent to those in V at any other position. We will write $S(T)$ for $S_U(T)$.

Now we can easily check that $S(3) = S(4) = S(5) = S(1, 5) = S(3, 7) = S(2, 6) = S(1, 3, 5) = S(3, 5, 7) = \emptyset$, for if there were a vertex z in any of these sets we would find an induced $Star_{1,2,3}$ within $\{1, 2, 3, 4, 5, 6, 7, z\}$.

Suppose $x \in S(3, 5)$. Since $G[X, Y]$ is prime, x and 4 may not have the same neighbourhoods, and we can assume w.l.o.g. that there is y , not in U , which is adjacent to 4 but not to x . Since $S(4)$ is empty y is adjacent also to at least one of $\{2, 6\}$, and we can quickly check that in any case $\{1, 2, 3, 4, 5, 6, 7, x, y\}$ contains an induced $Star_{1,2,3}$. Then $S(3, 5) = \emptyset$, and by similar methods we can check that $S(1, 3) = S(5, 7) = \emptyset$.

Now if x is adjacent to 3 or 5 or both, it must be adjacent also to both 1 and 7.

Suppose $y \in S(2)$. Since $G[X, Y]$ is prime, w.l.o.g. we can find x adjacent to 1 but not to y . Observe that $(y, 2, 3, 4, 5, 6, 7)$ also induces a P_7 , and so x may not be adjacent to 3 or 5 by the previous results applied to this new P_7 . But then $\{1, 2, 3, 4, 5, x, y\}$ induces a copy of $Star_{1,2,3}$. So $S(2) = \emptyset$, and similarly $S(4) = S(2, 4) = S(4, 6) = \emptyset$.

This leaves only $S(\emptyset), S(2, 4, 6), S(1), S(1, 7), S(7), S(1, 3, 7), S(1, 5, 7), S(1, 3, 5, 7)$ as possible non-empty sets. These adjacency results hold for every induced P_7 in $G[X, Y]$.

Observe that if $x \in S(1) \cup S(1, 7)$ then $V = (x, 1, 2, 3, 4, 5, 6)$ induces a P_7 , so $S(1, 3, 7) = S_V(2, 4) = \emptyset = S_V(2, 6) = S(1, 5, 7)$; similarly if $x \in S(7)$ then $S(1, 3, 7) = S(1, 5, 7) = \emptyset$.

We consider two cases.

Case 1: $S(1, 3, 7) = S(1, 5, 7) = \emptyset$.

We define a set $A \subset V(G[X, Y])$ recursively as follows. First, $U \subset A$. Second, if $V \subset A$ induces a P_7 , and $z \in S_V(1) \cup S_V(1, 7) \cup S_V(7)$, then $z \in A$. We can think of the set A as being the vertices of $G[X, Y]$ which are covered by starting with the P_7 induced by U and moving it along one vertex at a time.

If $z \in A$ has two or more neighbours in A , by construction there is an induced P_7 with vertices in A which includes z and two of its neighbours. Then z has exactly two neighbours in A , or the adjacency results would be violated with respect to this P_7 . Thus A induces a subgraph of $G[X, Y]$ which is either a path or a cycle on at least seven vertices.

Now if $x \in X - A$ is adjacent to some $a \in A$, then there is a set $V \subset A$ which contains a and induces a P_7 . Since $x \notin A$, $x \notin S_V(1) \cup S_V(1, 7) \cup S_V(7)$. Therefore $x \in S_V(1, 3, 5, 7) \cup S_V(2, 4, 6)$. By applying the adjacency results to each P_7 contained in A , we see that x is adjacent to every vertex in $A \cap Y$.

Similarly, if $y \in Y - A$, then y is adjacent either to every vertex, or to no vertex, in $A \cap X$. We let the set of vertices in $V(G[X, Y]) - A$ which are not adjacent to any vertices in A be E , and let $K = S(1, 3, 5, 7) \cup S(2, 4, 6)$. Then $V(G[X, Y]) = A \cup E \cup K$.

Since $G[X, Y]$ is not $K_{1,3}$ -free, there must be a vertex of $G[X, Y]$ not in A .

Since $G[X, Y]$ is not disconnected, if there is a vertex in E there must also be a vertex in K .

Since $G[X, Y]$ does not have disconnected bipartite complement, if there is a vertex in K there must also be a vertex in E . Thus there is at least one vertex in both E and K .

Recall that $G[X, Y]$ is a counterexample of minimal order. Then the subgraph G' of $G[X, Y]$ induced by removing the vertex 1 is not a counterexample. Since vertices of $G[X, Y]$ not in A which are adjacent to 1 are also adjacent to 3, G' must be prime, and G' is certainly $Star_{1,2,3}$ -free.

Observe that G' is not $K_{1,3}$ -free, for there is a vertex of degree at least three in K . Its bipartite complement is not $K_{1,3}$ -free, for there is a vertex in E of co-degree at least three. It is connected, since there is a vertex in K , and its bipartite complement is connected, since there is a vertex in E . Thus G' must have a $K+S$ -decomposition G'_1, G'_2 . But it is not hard to check that the bipartite graph P_6 does not have a $K+S$ -decomposition, and the vertices 2, 3, 4, 5, 6, 7 in G' induce a P_6 . So either G'_1 contains all of these vertices, or G'_2 does; in either case, we can find G_1, G_2 containing G'_1, G'_2 respectively which are a $K+S$ -decomposition of $G[X, Y]$. This is a contradiction.

Case 2: there is a vertex x in $S(1, 3, 7) \cup S(1, 5, 7)$.

Consider the bipartite complement $J[X, Y]$ of $G[X, Y]$. This contains an induced P_7 on the vertices (3, 6, 1, 4, 7, 2, 5) in that order, and the vertex x in $S(1, 3, 7) \cup S(1, 5, 7)$ is adjacent to either 5 or to 3, and to no other vertex in U . Thus $J[X, Y]$ is a counterexample of minimal order which fulfills the conditions of case 1, so it does not exist. \square

This now allows us to count the number of $Star_{1,2,3}$ -free bipartite graphs on n vertices.

Corollary 15. $Forb_n(Star_{1,2,3}) = n^{n+o(n)}$.

Proof. Since $JS(1, 1) = P_5$ is an induced subgraph of $Star_{1,2,3}$, we have the claimed lower bound.

Now suppose we have a bipartite graph G on n vertices which does not contain a copy of $Star_{1,2,3}$.

We can find a bipartite $Star_{1,2,3}$ -free graph G' which is prime by identifying sets of vertices with identical neighbourhoods in G .

By Theorem 14, either G' is disconnected, or has a $K+S$ -decomposition, or is a path (on at least two vertices) or cycle, or one of these is true of its bipartite complement.

We can record G in the following way. First, we record a bipartition X, Y of G . Then we find sets of vertices with identical neighbourhoods and replace each set with single vertices with that neighbourhood, labelled d_1, \dots . This gives a bipartite graph G' which is prime. If G' is disconnected, we write UNION(, followed by the recordings of each of the components of G' , then the closed bracket. If G' has a $K+S$ -decomposition, we write $K + S$ (, followed by the recordings of the decomposition graphs G_1 and G_2 (where G_1 is the subgraph of G' induced by $X_1 \cup Y_1$ and G_2 that induced by $X_2 \cup Y_2$, $X_1 \cup Y_2$ induces a bipartite clique and

$X_2 \cup Y_1$ an independent set), followed by the closed bracket. We write the recording of G_1 before that of G_2 . If G' is a path, we write **PATH**(, followed by the vertices of the path, in the path order, then the closed bracket. If G' is a cycle, we write **CYCLE**(, followed by the vertices of the cycle, in an order of the cycle, then the closed bracket. If we cannot do any of the previous we write **COMPLEMENT**(, then the recording of the bipartite complement of G' , then the closed bracket. Finally, we replace the vertices d_i in the recording of G' by **IDENTIFY**(, followed by the set of vertices with identical neighbourhoods which were identified to give d_i , then the closed bracket.

Now the total number of appearances in the recording of **UNION**(, **K + S**(, **PATH**(, **CYCLE**(and **IDENTIFY**(is at most $n - 1$, and the total number of appearances of **COMPLEMENT**(is also at most this number. Thus the whole recording consists of the bipartition, a linear order on the n vertices of $G[X, Y]$, and at most $4n$ insertions of seven different strings (including the closed bracket). Thus there are at most $2^n n^n 8^{5n} = n^{n+o(n)}$ possible recordings of bipartite graphs with n vertices not containing $Star_{1,2,3}$, and $Forb_n(Star_{1,2,3}) = n^{n+o(n)}$. \square

Since each of the exceptional graphs except for P_7 is an induced subgraph of $Star_{1,2,3}$, we see that $Forb_n(H) = n^{n+o(n)}$ and $Forb_n^*(H) = n^{n+o(n)}$ for each of P_6 , $\overline{P_6}$, $Star_{1,2,2}$, $\overline{Star_{1,2,2}}$ and $Star_{1,2,3}$.

Observe that if $G[X, Y]$ does not contain a copy of $P_6[U, V]$, then certainly $G[X, Y]$ does not contain an induced $Star_{1,2,3}$ with any bipartition, so by Corollary 15 $Forb_{m,n}(P_6) = m^{m+n+o(m)}$. Similarly $Forb_{m,n}(\overline{P_6}) = m^{m+n+o(m)}$. But we cannot use Corollary 15 to bound $Forb_{m,n}(H[U, V])$ for H any of $Star_{1,2,2}$, $\overline{Star_{1,2,2}}$ or $Star_{1,2,3}$: there are graphs $G[X, Y]$ which do not contain a copy of $H[U, V]$ but which do have an induced $Star_{1,2,3}$. For example, $Star_{1,2,3}[V, U]$ does not contain a copy of $Star_{1,2,3}[U, V]$.

6 Remaining problems

There are still some unresolved problems. Most importantly, we do not have good bounds on $Forb_n^*(P_7) = Forb_n(P_7)$. A possible approach to finding such bounds would be to find a decomposition result for P_7 -free bipartite graphs, perhaps in a manner similar to Lozin's result for $Star_{1,2,3}$ -free graphs.

For completeness, it would be nice to find more accurate bounds for $Forb_{m,n}(H[U, V])$ for each of the four infinite families $DS(k, l)$, $DS^*(k, l)$, $JS(k, l)$ and $JS^*(k, l)$. We know that the upper bound for $JS(1, 0) = P_4$ is correct, but we conjecture that in every other case the lower bound is accurate (and so also for $Forb_n(H[U, V])$ and $Forb_n^*(H)$).

It would be of some interest to find good bounds on $Forb_{m,n}(H[U, V])$ for the three exceptional graphs $Star_{1,2,2}$, $\overline{Star_{1,2,2}}$ and $Star_{1,2,3}$. It seems likely that the lower bounds should be correct.

Finally, we recall that Brightwell, Grable and Prömel left unclassified the speed of partial orders without certain induced sub-orders: those corresponding to the four infinite families $DS(k, l)$, $DS^*(k, l)$, $JS(k, l)$ and $JS^*(k, l)$, and the six exceptional graphs. They conjectured

that in each case the correct speed should be $n^{O(n)}$. Our results certainly support this conjecture.

References

- [1] József Balogh, Béla Bollobás and David Weinreich, The penultimate rate of growth for graph properties, *European J. Comb* **22** (2001), 277–289
- [2] Béla Bollobás, *Modern Graph Theory*, Springer-Verlag, 1998, xiii+394pp
- [3] C. Benson, Minimal regular graphs of girth eight and twelve, *Canad. J. Math.* **26** (1966), 1091–1094
- [4] Graham Brightwell, David Grable and Hans Jürgen Prömel, Forbidden Induced Partial Orders, *Discrete Mathematics* **201** (1999), 53–80
- [5] Jean-Luc Fouquet, Vassilis Giakoumakis and Jean-Marie Vanherpe, Bipartite graphs totally decomposable by canonical decomposition, *Internat. J. Foundations Comput. Sci.* **10(4)** (1999) 513–533
- [6] Vassilis Giakoumakis and Jean-Marie Vanherpe, Bi-complement Reducible Graphs, *Advances in Applied Mathematics* **18** (1997), 389–402
- [7] Vadim Lozin, Bipartite graphs without a skew star, *Discrete Mathematics* **257** (2002), 83–100
- [8] Hans Jürgen Prömel and Angelika Steger, Excluding induced subgraphs III: a general asymptotic, *Random Structures and Algorithms* **3** (1992), 19–31
- [9] Jeremy P. Spinrad, *Efficient Graph Representations*, American Mathematical Society, 2003, 342pp