# Graph Parameters and the Speed of Hereditary Properties

**Andrew Collins**

Thesis submitted for the degree of

Doctor of Philosophy in Mathematics

The University of Warwick

Department of Mathematics

July 2017

# Contents

# Acknowledgements

I would like to acknowledge and thank the following important people, without whom this thesis would not have been possible.

Firstly, I would like to thank my PhD supervisor Dr. Vadim V. Lozin. I would also like to extend these thanks to all of my co-authors.

I would also like to thank all of the staff and students at the Mathematics Institute of the University of Warwick, especially Carole Fisher for her support.

I, of course, would like to thank all of my friends (and my cat, Molly) for helping me get to this point.

Finally, I would like to thank my family. My sister for her enthusiasm and motivation, my mother for her frustration at not being able to proofread this thesis and for providing a home throughout my studies, and my father who unfortunately is no longer here to read these words. I miss you.

# Declaration

The results in this thesis are obtained in collaboration with other researchers, as specified below. This work has not been submitted for a degree at another university.

- Chapter 4 is based on [9] which was joint work with A. Atminas, J. Foniok and V. Lozin.
- Chapter 5 is based on [21] which was joint work with J. Foniok, N. Korpelainen, V. Lozin and V. Zamaraev.
- Chapter 6 is based on [10] which was joint work with A. Atminas, V. Lozin and V. Zamaraev.
- Chapter 7 is based on [22] which was joint work with S. Kitaev and V. Lozin.

# Abstract

In this thesis we study the speed of hereditary properties of graphs and how this defines some of the structure of the properties. We start by characterizing several graph parameters by means of minimal hereditary classes. We then give a global characterization of properties of low speed, before looking at properties with higher speeds starting at the Bell number. We then introduce a new parameter, clique-width, and show that there are an infinite amount of minimal hereditary properties with unbounded clique-width. We then look at the factorial layer in more detail and focus on $P_7$-free bipartite graphs. Finally we discuss word-representable graphs.

# Chapter 1

# Introduction

A *graph property* (or a *class of graphs*[1]) is a set of graphs closed under isomorphism. A property is *hereditary* if it closed under taking induced subgraphs. It is well-known (and not difficult to see) that a property $X$ is hereditary if and only if it can be characterized in terms of forbidden induced subgraphs. In other words, $X$ is hereditary if and only if there exists a set $M$ of graphs such that $G$ belongs to $X$ if and only if $G$ contains no graph from $M$ as an induced subgraph, in which case we say that $G$ is $M$-free. The set of all $M$-free graphs will be denoted $Free(M)$.

In general, there may exist many sets $M$ such that $X = Free(M)$ for a hereditary class $X$. However, for every hereditary class there exists a unique set of *minimal* forbidden induced subgraphs, which is also well-known and not difficult to see. Similarly, some families of hereditary classes can be characterize by means of minimal "forbidden" elements, i.e. minimal hereditary classes that do not belong to these families. In the present dissertation, we provide such a characterization for some families defined in terms of their speed. To introduce this notion, let us denote by $X_n$ for the set of graphs in $X$ with vertex set $\{1, 2, \ldots, n\}$, i.e. $X_n$ is the set of *labelled* graphs in $X$. The *speed* of $X$ is the function $f(n) = |X_n|$.

Trivially, if $X$ is the class of all simple graphs, then $|X_n| = 2^{\binom{n}{2}}$, and if $X$ is the class of complete graphs, then $|X_n| = 1$ for all natural $n$. The last observation shows, in particular, that the class of complete graphs is infinite. Also, it is not difficult to see

[1]Throughout the text we use the two notions – graph property and graph class – interchangeably.

that this is a *minimal* hereditary class containing infinitely many graphs. Indeed, if $Y$ is any of its proper hereditary subclasses, then at least one complete graph $K_n$ is not in $Y$, i.e. $Y \subset Free(K_n)$. But then $Y$ contains only graphs with at most $n-1$ vertices, or equivalently, only $n-1$ graphs. Similarly, the class of empty (edgeless) graphs is a minimal hereditary class with infinitely many graphs. Moreover, complete graphs and empty graphs are the *only* two minimal infinite hereditary classes. This fact is highly non-trivial and follows, in particular, from Ramsey's Theorem.

In 1930, a 26 years old British mathematician Frank Ramsey proved the following theorem, known nowadays as Ramsey's Theorem.

**Theorem 1.** [59] *For any positive integers $k$, $r$, $p$, there exists a positive integer $R = R(k, r, p)$ with the following property. If the $k$-subsets of an $R$-set are colored with $r$ colors, then there is a monochromatic $p$-set, i.e., a $p$-set all of whose $k$-subsets have the same color.*

It is not difficult to see that with $k = 1$ the theorem coincides with Pigeonhole Principle: for any $r$ and $p$, there exists an $n = n(r, p)$ such that for any coloring of $n$ objects with $r$ different colors there exist $p$ objects of the same color.

For $k = 2$, the theorem admits a nice interpretation in the terminology of graph theory, since coloring 2-subsets can be viewed as coloring the edges of a complete graph: for any positive integers $r$ and $p$, there is a positive integer $n = n(r, p)$ such that if the edges of an $n$-vertex complete graph are colored with $r$ colors, then there is a monochromatic clique of size $p$, i.e., a clique all of whose edges have the same color.

In the case of $r = 2$ colors, the graph-theoretic interpretation of Ramsey's Theorem can be further rephrased as follows.

**Theorem 2.** *For any positive integer $p$, there is a positive integer $n = n(p)$ such that every graph with at least $n$ vertices has either a clique of size $p$ or an independent set of size $p$.*

**Definition 1.** The minimum $n$ such that every graph with $n$ vertices contains either a clique of size $p$ or an independent set of size $p$ is the *symmetric Ramsey number $R(p)$*.

Theorem 2 also admits a non-symmetric formulation as follows.

**Theorem 3.** *For any positive integer $p$ and $q$, there is a positive integer $n = n(p,q)$ such that every graph with at least $n$ vertices has either a clique of size $p$ or an independent set of size $q$.*

**Definition 2.** The minimum $n$ such that every graph with $n$ vertices contains either a clique of size $p$ or an independent set of size $q$ is the *Ramsey number* $R(p,q)$.

Theorem 3 (or Theorem 2) allows us to make the following conclusion: if $X$ is a hereditary class which does not contain a complete graph $K_m$ and an empty graph $\overline{K}_m$, then graphs in $X$ have less than $R(n,m)$ vertices, i.e. $X$ is finite. More formally, Theorem 3 implies the following conclusion.

**Theorem 4.** *The class of complete graphs and the class of edgeless graphs are the only two minimal infinite hereditary classes of graphs.*

On the other hand, it is not difficult to see that the reverse is also true: Theorem 4 implies Theorem 3. In other words, these two theorems are equivalent.

Theorem 4 characterizes the family of finite hereditary classes in terms of minimal "forbidden" elements, i.e. minimal classes that do not belong to this family. In the present dissertation, we provide a similar characterization for some other families of hereditary classes. In particular, in chapter 3 we show that in the case of classes with polynomial or exponential speed of growth the family of minimal "forbidden" elements is finite. To contrast, in chapter 4 we show that for the case of classes with speeds below the Bell number this family is infinite.

Let us observe that Theorem 4 can be rephrased as follows: the class of complete graphs and the class of edgeless graphs are the only two minimal hereditary classes of graphs with finitely many vertices, i.e. it characterizes the family of classes where a certain graph parameter (the number of vertices) is bounded by means of minimal classes where this parameter is unbounded. The literature contains some other results of this type. For instance, the celebrated theorem of Robertson and Seymour states that the class of planar graphs is the unique minimal minor-closed class of graphs of unbounded tree-width [60]. In the present dissertation, in chapter 5 we show that in the case of clique-width (a notion generalizing tree-width) there are infinitely minimal hereditary classes where this parameter is unbounded. We also identify a new minimal class of unbounded linear clique-width.

In addition to characterizing the families of hereditary classes with polynomial or exponential speeds of growth in terms of minimal "forbidden" classes, we also characterize these families by means of graph parameters which are bounded in such classes. In particular, we show that the speed of a hereditary class $X$ is at most exponential if and only if graphs in $X$ have bounded neighbourhood diversity.

Finally, we obtain a number of other results related to the speed of hereditary properties of graphs. In particular, in chapters 6 and 7 we identify new subclasses of $P_7$-free bipartite graphs with factorial speed of growth and determine the asymptotic speed of so-called word-representable graphs.

In the rest of this chapter we introduce basic definitions and notations, provide some motivation and prove a number of preliminary results.

## 1.1 Preliminaries

All graphs in this dissertation are simple, i.e. undirected, without loops and multiple edges. The vertex set and the edge set of a graph $G$ are denoted by $V(G)$ and $E(G)$, respectively. If $v$ is a vertex of $G$, then $N(v)$ is the neighbourhood of $v$, i.e. the set of vertices adjacent to $v$, and $deg(v) = |N(v)|$ is the degree of $v$.

Given a subset $U \subseteq V(G)$, we denote by $G[U]$ the subgraph of $G$ induced by $U$, the graph with vertex set $U$ in which two vertices are adjacent if and only if they are adjacent in $G$.

### 1.1.1 The speed of hereditary properties and graph coding

As we mentioned earlier, the speed of a hereditary property $X$ is the number $|X_n|$ of $n$-vertex labelled graphs in $X$. Determining $|X_n|$ is important in many respects, in particular, for optimal coding of graphs. By graph coding we mean the problem of representing graphs by words in a finite alphabet, which is important in computer science for representing graphs in computer memory [34, 38, 65]. Without loss of generality we will assume that our alphabet is binary, i.e. consists of two symbols 0 and 1.

More formally, denote $B = \{0, 1\}$ and let $B^*$ be the set of all words over the alphabet $B$. Given a word $\alpha \in B^*$, we denote by $|\alpha|$ the length of $\alpha$ and by $\alpha_j$ the $j$-th letter of $\alpha$. Also, $\lambda$ stands for the empty word (the only word of length 0).

*Coding* of graphs in the class $X$ is a family of bijective mappings $\Phi = \{\phi_n : n = 1, 2, 3, \ldots\}$, where $\phi_n : X \to B^*$. A coding $\Phi$ will be called *asymptotically optimal* if

$$\lim_{n \to \infty} \frac{\max\limits_{G \in X_n} |\phi_n(G)|}{\log |X_n|} = 1.$$

Every labelled graph $G$ with $n$ vertices can be represented by a binary word of length $\binom{n}{2}$, one bit per each pair of vertices, with 1 standing for an edge and 0 for an non-edge. Such a word can be obtained by reading the elements of the adjacency matrix above the main diagonal. The word obtained by reading these elements row by row, starting with the first row, will be called the *canonical coding* of $G$ and will be denoted $\phi_n^c(G)$.

If no priory information about the graph is available, then $\binom{n}{2}$ is the minimum number of bits needed to represent the graph. However, if we know that our graph possesses some special properties, then this knowledge may lead to a shorter representation. For instance,

- if we know that our graph is bipartite, then we do not need to describe the adjacency of vertices that belong to the same part in its bipartition. Therefore, we need at most $n^2/4$ bits to describe the graph, the worst case being a bipartite graph with $n/2$ vertices in each of its parts.

- if we know that our graph is not an arbitrary bipartite graph but *chordal* bipartite, then we can further shorten the code and describe any graph in this class with at most $O(n \log^2 n)$ bits[2] [64].

- a further restriction to trees (a proper subclass of chordal bipartite graphs) enables us to further shorten the code to $(n-2)\log n$ bits, which is the length of binary representation of Prüfer code for trees [58].

How much can the canonical representation be shortened for graphs with a property $X$? For hereditary properties this question can be answered through the notion of entropy.

[2] All logarithms are of base 2

### 1.1.2 Entropy of hereditary properties

In order to represent graphs of size $n$ in a class $X$, we need at least $|X_n|$ different binary words. Therefore, in the worst case the length of a binary code of an $n$-vertex graph in $X$ cannot be shorter than $\log |X_n|$. Thus, the ratio

$$\frac{\log |X_n|}{\binom{n}{2}}$$

can be viewed as the coefficient of compressibility for representing $n$-vertex graphs in $X$. Its limit value, for $n \to \infty$, was called by Alekseev in [3] the *entropy* of $X$. Moreover, in the same paper Alekseev showed that for every hereditary property $X$ the entropy necessarily exists and in [4] he proved that its value takes the following form:

$$\lim_{n\to\infty} \frac{\log |X_n|}{\binom{n}{2}} = 1 - \frac{1}{k(X)}, \tag{1.1}$$

where $k(X)$ is a natural number, called the *index* of $X$. To define this notion let us denote by $\mathcal{E}_{i,j}$ the class of graphs whose vertices can be partitioned into at most $i$ independent sets and $j$ cliques. In particular, $\mathcal{E}_{2,0}$ is the class of bipartite graphs and $\mathcal{E}_{1,1}$ is the class of split graphs. Then $k(X)$ is the largest $k$ such that $X$ contains all $\mathcal{E}_{i,j}$ with $i + j = k$. Independently, this result was obtained by Bollobás and Thomason [14, 15] and is known nowadays as the Alekseev-Bollobás-Thomason Theorem (see e.g. [7]).

### 1.1.3 Coding of graphs in classes of high speed

In [3], Alekseev proposed a universal algorithm which gives an asymptotically optimal coding for graphs in every hereditary class $X$ of index $k > 1$, i.e. of non-zero entropy. Below we present an adapted version of this algorithm with a proof of its optimality.

Let $n > 1$ and let $p$ be a prime number between $\lfloor n/\sqrt{\log n} + 1\rfloor$ and $2\lfloor n/\sqrt{\log n}\rfloor$. Such a number always exists by the Bertrand-Chebyshev theorem (see e.g. [1]). Define $k = \lfloor n/p \rfloor$. Then

$$p \leq 2n/\sqrt{\log n}, \quad k \leq \sqrt{\log n}, \quad n - kp < p. \tag{1.2}$$

Let $G$ be an arbitrary graph with $n$ vertices. Denote by $D_n$ the set of all pairs of vertices of $G$. We split $D_n$ into two disjoint subsets $R_1$ and $R_2$ as follows: $R_1$ consists of the pairs $(a, b)$ such that $a \leq kp$, $b \leq kp$ and $\lfloor (a-1)/p \rfloor \neq \lfloor (b-1)/p \rfloor$, and $R_2$ consists of all the remaining pairs. Let us denote by $\mu^{(1)}$ the binary word consisting of the elements of the canonical code corresponding to the pairs of $R_2$. This word will be included in the code of $G$ we construct.

Now let us take care of the pairs in $R_1$. For all $x, y \in \{0, 1, \ldots, p-1\}$, we define

$$Q_{x,y} = \{pi + 1 + res_p(xi + y) \ : \ i = 0, 1, \ldots, k-1\},$$

where $res_p(z)$ is the remainder on dividing $z$ by $p$. Let us show that every pair of $R_1$ appears in exactly one set $Q_{x,y}$. Indeed, if $(a, b) \in Q_{x,y}$ $(a < b)$, then

$$xi_1 + y \equiv a \pmod{p}, \quad xi_2 + y \equiv b \pmod{p},$$

where $i_1 = \lfloor (a-1)/p \rfloor$, $i_2 = \lfloor (b-1)/p \rfloor$. Since $i_1 \neq i_2$ (by definition of $R_1$), there exists a unique solution of the following system

$$\begin{aligned} x(i_1 - i_2) &\equiv a - b \pmod{p} \\ y(i_1 - i_2) &\equiv ai_2 - bi_1 \pmod{p}. \end{aligned} \tag{1.3}$$

Therefore, by coding the graphs $G_{x,y} = G[Q_{x,y}]$ and combining their codes with the word $\mu^{(1)}$ (that describes the pairs in $R_2$) we obtain a complete description of $G$.

To describe the graphs $G_{x,y} = G[Q_{x,y}]$ we first relabel their vertices according to

$$z \to \lfloor (z-1)/p \rfloor + 1.$$

In this way, we obtain $p^2$ graphs $G'_{x,y}$, each on the vertex set $\{1, 2, \ldots, k\}$. Some of these graphs may coincide. Let $m$ $(m \leq p^2)$ denote the number of pairwise different graphs in this set and $H_0, H_1, \ldots, H_{m-1}$ an (arbitrarily) ordered list of $m$ pairwise different graphs in this set. In other words, for each graph $G'_{x,y}$ there is a unique number $i$ such that $G'_{x,y} = H_i$. We denote the binary representation of this number $i$ by $\omega(x, y)$ and the length of this representation by $\ell$, i.e. $\ell = \lceil \log m \rceil$. Also, denote

$$\mu^{(2)} = \phi_k^c(H_0)\phi_k^c(H_1)\ldots\phi_k^c(H_{m-1}),$$

$$\mu^{(3)} = \omega(0,0)\omega(0,1)\ldots\omega(0,p-1)\omega(1,0)\ldots\omega(p-1,p-1).$$

The word $\mu^{(2)}$ describes all graphs $H_i$ and the word $\mu^{(3)}$ indicates for each pair $x, y$ the interval in the word $\mu^{(2)}$ containing the information about $G'_{x,y}$. Therefore, the words $\mu^{(2)}$

and $\mu^{(3)}$ completely describe all graphs $G_{x,y}$. In order to separate the word $\mu^{(2)}\mu^{(3)}$ into $\mu^{(2)}$ and $\mu^{(3)}$, it suffices to know the number $\ell$, because $|\mu^{(2)}| = \ell p^2$ and the number $p$ is uniquely defined by $n$. Since $m \leq 2^{\binom{k}{2}}$, the number $\ell$ can be described by at most

$$\lceil \log \ell \rceil = \lceil \log \lceil \log m \rceil \rceil \leq \lceil \log \binom{k}{2} \rceil \leq \lceil \log k^2 \rceil \leq \lceil \log \log n \rceil$$

binary bits. Let $\mu^{(0)}$ be the binary representation of the number $\ell$ of length $\lceil \log \log n \rceil$, and let

$$\phi_n^*(G) = \mu^{(0)}\mu^{(1)}\mu^{(2)}\mu^{(3)}, \qquad \Phi^* = \{\phi_n^* \ : \ n = 2, 3, \ldots\}.$$

**Theorem 5.** $\Phi^*$ *is an asymptotically optimal coding for any hereditary class* $X$ *with* $c(X) > 1$.

*Proof.* From the construction of $\Phi^*$ it is clear that any graph is uniquely defined by its code. Therefore, $\Phi^*$ is a coding for any class of graphs. Assume now that our graph $G$ belongs to a hereditary class $X$ with $c(X) > 1$. We denote the entropy of $X$ by $h(X)$, i.e.

$$h(X) = \lim_{n \to \infty} \frac{\log |X_n|}{\binom{n}{2}},$$

and therefore,

$$|X_n| = 2^{\frac{n^2}{2}(h(X)+\varepsilon_n)},$$

where $\varepsilon_n \to 0$ when $n \to \infty$.

Let $n$ be the number of vertices of $G$. We estimate the length of the words in the code $\phi_n^*(G)$ as follows:

$$|\mu^{(0)}| = \lceil \log \log n \rceil,$$
$$|\mu^{(1)}| = k\binom{p}{2} + kp(n-kp) + \binom{n-kp}{2}.$$

Taking into account (1.2), we conclude that

$$|\mu^{(1)}| \leq \frac{3kp^2}{2} + \frac{p^2}{2} \leq \frac{6n^2}{\sqrt{\log n}} + \frac{2n^2}{\log n} = o(n^2).$$

Each graph $H_i$ belongs to $X_k$ and hence the number $m$ of these graphs satisfies

$$m \leq |X_k| = 2^{\frac{k^2}{2}(h(X)+\varepsilon_k)},$$

where $\varepsilon_k \to 0$ when $k \to \infty$. Therefore,

$$|\mu^{(2)}| = m\binom{k}{2} < k^2 2^{\frac{k^2}{2}(h(X)+\varepsilon_k)} \leq n^{\frac{1}{2}(h(X)+\varepsilon_k)} \log n,$$

$$|\mu^{(3)}| = p^2 \lceil \log m \rceil \leq \frac{p^2 k^2}{2}(h(X) + \varepsilon_k).$$

Since $h(X) \leq 1$ and $k \to \infty$ when $k \to \infty$, we conclude that $|\mu^{(2)}| = o(n^2)$. Also, since $kp \leq n$, we have

$$|\mu^{(3)}| \leq \frac{n^2}{2} h(X) + o(n^2).$$

Combining the above arguments, we obtain

$$|\phi_n^*(G)| \leq \frac{n^2}{2} h(X) + o(n^2).$$

Therefore, if $c(X) > 1$ (i.e. if $h(X) > 0$), then

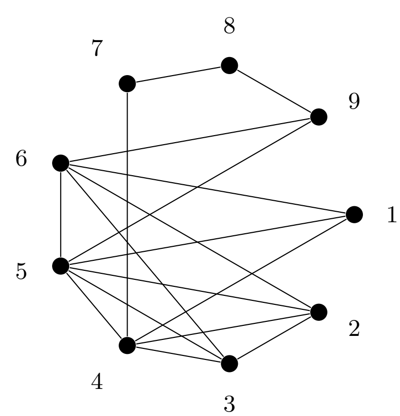
$$\lim_{n\to\infty} \frac{\max\limits_{G \in X_n} |\phi_n^*(G)|}{\log |X_n|} = 1,$$

and hence, $\Phi^*$ is an asymptotically optimal coding for $X$. $\square$

*Example.* Let $G$ be the graph (pictured below) with 9 vertices $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and the following canonical code:

00111000 1111000 111000 10100 1001 001 10 1.

The code is obtained by listing the elements of the adjacency matrix above the main diagonal. For convenience, the elements coming from different rows of the matrix are separated. Let us define $k = p = 3$ (this choice of $k$ and $p$ does not satisfy their definition, but it is not very important. These numbers still satisfy (1.2)).




$G$

The set $R_2$ consists of the pairs of vertices that belong to the subgraphs of $G$ induced by three sets $\{1, 2, 3\}$, $\{4, 5, 6\}$ and $\{7, 8, 9\}$. The word $\mu^{(1)}$ consists of the elements of the canonical code corresponding to these pairs:

$$\mu^{(1)} = 001101101.$$

Now for each pair $(x, y) \in \{0, 1, 2\}^2$, we compute the set $Q_{x,y}$ and the canonical code of the graph $G'_{x,y}$. The results are presented in the table below.

| $x, y$ | $Q_{x,y}$ | $\phi_3^c(G'_{x,y})$ | $\omega(x, y)$ |
|---|---|---|---|
| 0,0 | 1,4,7 | 101 | 0 |
| 0,1 | 2,5,8 | 100 | 1 |
| 0,2 | 3,6,9 | 101 | 0 |
| 1,0 | 1,5,9 | 101 | 0 |
| 1,1 | 2,6,7 | 100 | 1 |
| 1,2 | 3,4,8 | 100 | 1 |
| 2,0 | 1,6,8 | 100 | 1 |
| 2,1 | 2,4,9 | 100 | 1 |
| 2,2 | 3,5,7 | 100 | 1 |

The third column of the table shows that among the graphs $G'_{x,y}$ for various values of $x$ and $y$ there are only two different graphs, i.e. $m = 2$ and hence $\ell = 1$. The canonical codes of these graphs are 101 and 100. These two codes give us the word $\mu^{(2)}$:

$$\mu^{(2)} = 101100.$$

We code the graph with canonical code 101 by 0 and the graph with canonical code 100 by 1. This coding give us the word $\mu^{(3)}$:

$$\mu^{(3)} = 010011111.$$

Thus, the code of $\phi_9^*(G) = \mu^{(0)}\mu^{(1)}\mu^{(2)}\mu^{(3)}$ is represented by the following binary sequence (for convenience, we separate different words of the code):

$$\phi_9^*(G) = 01 \quad 001101101 \quad 101100 \quad 010011111.$$

# Chapter 2

# Graph parameters and Ramsey Theory

In this chapter, we characterize several graph parameters by means of minimal hereditary classes, where these parameters are unbounded. In other words, we prove a number of results of the following form: a graph parameter $p(G)$ is unbounded in a hereditary class $X$ if and only if $X$ contains a class from a family $\mathcal{F}$. According to Ramsey's Theorem, if $p(G)$ is the number of vertices in $G$, then $\mathcal{F}$ consists of two classes: complete graphs and edgeless graphs. That is why we call the results of this form Ramsey-type results. We start with two illustrating results, which will be useful later.

**Proposition 1.** *For any positive integers $s$ and $t$, there exists a positive integer $d = d(s,t)$ such that any graph $G$ of vertex degree at least $d$ contains either a complete graph $K_s$ or an induced star $K_{1,t}$.*

*Proof.* Let $d = R(s-1,t)$ and assume $G$ has a vertex $v$ of degree $d$. Then, by Theorem 3, the neighbourhood of $v$ contains either an independent set $I$ of size $t$, in which case $I \cup \{v\}$ induces a $K_{1,t}$, or a clique of size $s$, i.e. $G$ contains $K_s$. □

This proposition not only uses Theorem 3 for the proof. It is also stated in the style of Theorem 3. Moreover, Proposition 1 admits a reformulation in the style of Theorem 4 as well. To show this, let us introduce the following notation.

$\mathcal{R} = Free(\overline{P}_3, K_3, C_4)$. Since any graph $G$ in this class is $\overline{P}_3$-free, it is complete multipartite. Since $G$ is $K_3$-free, it is complete bipartite. Finally, since $G$ is $C_4$-free, one of the parts in its bipartition contains at most one vertex. Therefore, $G$ is either a star $K_{1,n}$ or an edgeless graph. In other words, $\mathcal{R}$ is the class of all stars and all their induced subgraphs.

Clearly, vertex degree is bounded neither in $\mathcal{R}$ nor in the class of complete graphs. On the other hand, Proposition 1 states that if vertex degree is unbounded in a hereditary class, then it contains either all complete graphs or all stars, which implies the following conclusion.

**Proposition 2.** *The class of complete graphs and the class $\mathcal{R}$ are the only two minimal hereditary classes of graphs of unbounded vertex degree.*

One more Ramsey-type result that will be needed later can be viewed as the bipartite analog of Theorem 2. We give an independent proof of this result derived from the Pigeonhole Principle.

**Theorem 6.** *For every $s$, there is an $n = n(s)$ such that every bipartite graph $G$ with at leas $n$ vertices in each part contains either $K_{s,s}$ or the bipartite complement of $K_{s,s}$.*

*Proof.* Let $n = s2^{2s}$ and let $G = (A, B, E)$ be a bipartite graph with $|A| \geq n$ and $|B| \geq n$. Consider an arbitrary subset $A' \subseteq A$ with $2s$ vertices. We split the vertices of $B$ into at most $2^{2s}$ subsets in accordance with their neighbourhood in $A'$. Since $|B| \geq s2^{2s}$, there must exist a subset $B' \subseteq B$ with at least $s$ vertices. By definition all vertices of $B'$ have the same neighbourhood in $A'$, say $A''$. If $|A''| \geq s$, then $A'' \cup B'$ is a biclique with at least $s$ vertices in each part. Otherwise, $(A' - A'') \cup B'$ is the bipartite complement of a biclique with at least $s$ vertices in each part. $\square$

## 2.1 Independence number, clique number and complex number

As usual, $\alpha(G)$ stands for the independence number and $\omega(G)$ for the clique number of $G$. Now let us introduce a new parameter:

$c(G) = \min(\alpha(G), \omega(G))$ is the *complex number* of $G$.

In what follows we give a Ramsey-type characterization of this parameter, i.e. we characterize it in terms of minimal hereditary classes where the parameter is unbounded. To this end, let us introduce the following notation.

- $\mathcal{S} = Free(P_3, 2K_2)$. In other words, $\mathcal{S}$ is the class of graphs partitionable into a clique and a set of isolated vertices. Indeed, since $G \in \mathcal{S}$ is $P_3$-free, every connected component of $G$ is a clique, and since $G$ is $2K_2$-free, at most one of its components has more than one vertex. Also, let $S_n$ be a graph in $\mathcal{S}$ with a clique of size $n$ and a set of isolated vertices of size $n$. Obviously, $S_n$ is $n$-universal for graphs in $\mathcal{S}$, i.e. it contains every $n$-vertex graph from $\mathcal{S}$ as an induced subgraph.

- $\overline{\mathcal{S}}$ is the class of complements of graphs in $\mathcal{S}$.

**Theorem 7.** *$\mathcal{S}$ and $\overline{\mathcal{S}}$ are the only two minimal hereditary classes of graphs of unbounded complex number.*

*Proof.* Obviously, the complex number of graphs in $\mathcal{S}$ and $\overline{\mathcal{S}}$ can be arbitrarily large. Conversely, let $X$ be a hereditary class containing a graph $G$ with $c(G) \geq k$ for each value of $k$. Then $G$ contains a clique $C$ of size $k$ and an independent set $I$ of size $k$. Since $C$ and $I$ have at most one vertex in common, we may assume without loss of generality that they are disjoint, and since $k$ can be arbitrarily large, in the bipartite graph $G[C, I]$ we can find an arbitrarily large biclique or its bipartite complement (Theorem 6). Therefore, graphs in $X$ contain either $S_n$ or $\overline{S}_n$ for arbitrarily large values of $n$, i.e. $X$ contains either $\mathcal{S}$ or $\overline{\mathcal{S}}$. $\square$

## 2.2 Degree, co-degree and $c$-degree

Let $G$ be a graph and $v$ a vertex of $G$. We denote by $d(v)$ the degree of a vertex $v$ and by $\overline{d}(v)$ the *co-degree* of $v$, i.e. the degree of $v$ in the complement of $G$. The $c$-degree of $v$ is denoted and defined as follows: $cd(v) = \min(d(v), \overline{d}(v))$.

As usual, $\Delta(G)$ is the maximum vertex degree in $G$. Also, we denote by $\overline{\Delta}(G)$ the maximum co-degree and by $c\Delta(G)$ the maximum $c$-degree in $G$. We call $c\Delta(G)$ the *complex* *degree* in $G$. In order to characterize this new parameter by means of minimal hereditary classes where $c\Delta(G)$ is unbounded, let us introduced the following notions.

- $\mathcal{Q} = Free(P_4, C_4, 2K_2, K_3)$. The structure of graphs in this class can be characterized as follows. The vertices of each graph $G \in \mathcal{Q}$ can be partitioned into a set inducing a star and a set of isolated vertices. Indeed, since $G$ is $(P_4, K_3)$-free, every connected component of $G$ is a biclique. Since $G$ is $C_4$-free, every connected component of $G$ is a star. Finally, since $G$ is $2K_2$-free, at most one of its components has more than one vertex. Also, let $Q_n$ be a graph in $\mathcal{Q}$ whose vertices can be partitioned into an induced star $K_{1,n}$ and a set of isolated vertices of size $n$. Obviously, $Q_n$ is $n$-universal for graphs in $\mathcal{Q}$, i.e. it contains every $n$-vertex graph from $\mathcal{Q}$ as an induced subgraph.

- $\overline{\mathcal{Q}}$ is the class of complements of graphs in $\mathcal{Q}$.

**Theorem 8.** *$\mathcal{S}$, $\overline{\mathcal{S}}$, $\mathcal{Q}$ and $\overline{\mathcal{Q}}$ are the only minimal hereditary classes of graphs of unbounded complex degree.*

*Proof.* Obviously, the complex degree of graphs in $\mathcal{S}$, $\overline{\mathcal{S}}$ $\mathcal{Q}$ and $\overline{\mathcal{Q}}$ can be arbitrarily large. Conversely, let $X$ be a hereditary class containing a graph $G$ with $c\Delta(G) \geq k$ for each value of $k$. Then $G$ contains a vertex $v$ such that $d(v) \geq k$ and $\overline{d}(v) \geq k$. Let $A$ be the set of neighbours of $v$ and $B$ the set of its non-neighbours. Since both $A$ and $B$ can be arbitrarily large, each of them contains either a big clique or a big independent set (Ramsey's Theorem), and $G[A, B]$ contains either a big biclique or its bipartite complement (Theorem 6). Therefore, graphs in $X$ contain either $S_n$ or $\overline{S}_n$ or $Q_n$ or $\overline{Q}_n$ for arbitrarily large values of $n$. As a result, $X$ contains at least one of $\mathcal{S}$, $\overline{\mathcal{S}}$, $\mathcal{Q}$ and $\overline{\mathcal{Q}}$. □

## 2.3 Matching number, co-matching number and $c$-matching number

The *matching number* of a graph $G$ is the size of a maximum matching in $G$ and we denote it by $\mu(G)$. The *co-matching number* of $G$ is the size of a maximum matching in the complement of $G$ and we denote it by $\overline{\mu}(G)$. The *c-matching number* of $G$ is defined and denoted as follows: $c\mu(G) = \min(\mu(G), \overline{\mu}(G))$. In this section, we characterize all three parameters in terms of minimal hereditary classes where these parameters are unbounded. To this end, we introduce several more classes of graphs.

$\mathcal{B}$ = $Free(\overline{P}_3, K_3)$. By forbidding $\overline{P}_3$, we exclude all odd cycles of length at least 5. Therefore, $\mathcal{B}$ is a class of bipartite graphs. Every graph in this class with at least one edge is complete bipartite, since otherwise an induced $\overline{P}_3$ arises. With a slight abuse of terminology, we will refer to graphs in this class as *complete bipartite* regardless of whether they contain edges or not. For consistency of notation with previously defined classes, we will denote a biclique with $n$ vertices in each part of its bipartition by $B_n$. Clearly, $B_n$ is $n$-universal for graphs in $\mathcal{B}$, i.e. it contains all $n$-vertex graphs in $\mathcal{B}$ as induced subgraphs.

$\overline{\mathcal{B}}$ is the class of complements of graphs in $\mathcal{B}$.

$\mathcal{M}$ = $Free(P_3, K_3)$. It is not difficult to see that $\mathcal{M}$ is the class of graphs of vertex degree at most 1. By $M_n$ we denote the unique (up to isomorphism) graph from this class with $2n$ vertices each of which has degree 1. Clearly, $M_n$ is $n$-universal for graphs in $\mathcal{M}$.

$\overline{\mathcal{M}}$ is the class of complements of graphs in $\mathcal{M}$.

$\mathcal{Z}$ = $Free(2K_2, K_3, C_5)$. In other words, these are $2K_2$-free bipartite graphs. These graphs are also known under the name of *chain graphs*, because for any bipartition, the neighbourhoods of the vertices in each part form a chain with respect to set-inclusion. By $Z_n$ we denote a chain graph such that for each $i \in \{1, 2, \ldots\}$, each part of the graph contains exactly one vertex of degree $i$. Figure 2.1 represents the graph $Z_n$ for $n = 5$ and Lemma 1 proves that $Z_n$ is $n$-universal for graphs in $\mathcal{Z}$.

**Lemma 1.** *$Z_n$ is an $n$-universal chain graph, i.e. it contains all $n$-vertex chain graphs as induced subgraphs.*

*Proof.* We prove by induction on $n$. For $n = 1, 2$, the statement is trivial. Now let $G$ be a chain graph with $n > 2$ vertices. We consider an arbitrary bipartition $X \cup Y$ of the

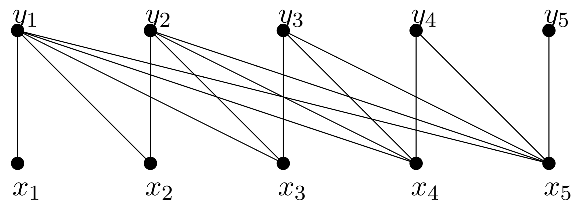



Figure 2.1: The graph $Z_5$

vertices of $G$ and order the vertices in $X$ increasingly and the vertices in $Y$ decreasingly with respect to their degrees, breaking ties arbitrarily. Let $x$ be a vertex with a largest degree in $X$ and $y$ a vertex with a smallest degree in $Y$. If $x$ is not adjacent to $y$, then $y$ is isolated in $G$, and if $x$ is adjacent to $y$, then $x$ dominates $Y$. In the first case, we map $y$ to the vertex $y_n$ of $Z_n$ and embed $G - y$ into $Z_{n-1}$ by induction. In the second case, we map $x$ to the vertex $x_n$ of $Z_n$ and embed $G - x$ into $Z_{n-1}$ by induction. In either case, $G$ is an induced subgraph of $Z_n$. □

**Lemma 2.** *For any positive integers $s, t$, there exists a positive integer $q = q(s,t)$ such that every bipartite graph $G$ with a matching of size $q$ contains either an induced $M_s$ or an induced $B_t$.*

*Proof.* Let us denote $m = 2\max(s,t)$ and $q = R(2,4,m)$, where $R(k,r,p)$ is the number from Ramsey's Theorem (Theorem 1). Consider a matching $M = \{x_1y_1, \ldots, x_qy_q\}$ of size $q$. We color each pair $(x_iy_i, x_jy_j)$ of edges in $M$ $(i < j)$ in one of the four colors as follows:

- color 1 if $G$ contains no edges between $x_iy_i$ and $x_jy_j$,
- color 2 if $G$ contains both edges between $x_iy_i$ and $x_jy_j$,
- color 3 if $G$ contains the edge $x_iy_j$ but not the edge $y_ix_j$,
- color 3 if $G$ contains the edge $y_ix_j$ but not the edge $x_iy_j$.

By Ramsey's Theorem, $M$ contains a monochromatic set $M'$ of edges of size $m$. If the color of each pair in $M'$ is

1 then $M'$ is an induced matching of size $m \geq 2s > s$,

2 then the vertices of $M'$ induce a biclique $B_m$ with $m \geq 2t > t$,

3 or 4, then the vertices of $M'$ induce a $Z_m$ and hence, by Lemma 1, $G$ contains a biclique $B_{m/2}$ with $m/2 \geq t$.

$\square$

**Lemma 3.** *For any natural $s, t, p$, there exists a $Q = Q(s, t, p)$ such that every graph $G$ with a matching of size $Q$ contains either an induced $M_s$ or an induced $B_t$ or a clique $K_p$.*

*Proof.* Let $Q = R(2, 2, \max(R(2, 2, \max(p, q)), p))$, where $q = R(2, 4, 2\max(s, t))$ (i.e. the value defined in the proof Lemma 2). We consider a matching $M$ of size $Q$ in $G$ and color the endpoints of each edge of $M$ in two colors, say white and black, arbitrarily. Since the set of white vertices has size $Q$, it must contain either a clique $K_p$, in which case we are done, or an independent set $A$ of size $R(2, 2, \max(q, p))$. In the latter case, we look at the black vertices matched with the vertices of $A$. According to the size of this set, it must contain either a clique $K_p$, in which case we are done, or an independent set $A'$ of size $q$. In the latter case, we denote by $A''$ the set of white vertices matched with the vertices of $A'$. Then $A'$ and $A''$ induce a bipartite graph with a matching of size $q$, in which case, by Lemma 2, $G$ contains either an induced matching of size $s$ or an induced biclique $B_t$. $\square$

The above sequence of results allow us to make the following conclusions, the first two of which follow directly from Lemma 3.

**Theorem 9.** *$\mathcal{M}$, $\mathcal{B}$ and the class of complete graphs are the only three minimal hereditary classes of graphs of unbounded matching number.*

**Theorem 10.** *$\overline{\mathcal{M}}$, $\overline{\mathcal{B}}$ and the class of edgeless graphs are the only three minimal hereditary classes of graphs of unbounded co-matching number.*

**Theorem 11.** *$\mathcal{M}$, $\mathcal{B}$, $\mathcal{S}$, $\overline{\mathcal{M}}$, $\overline{\mathcal{B}}$ and $\overline{\mathcal{S}}$ are the only six minimal hereditary classes of graphs of unbounded c-matching number.*

*Proof.* Clearly, graphs in $\mathcal{M}$, $\mathcal{B}$, $\mathcal{S}$, $\overline{\mathcal{M}}$, $\overline{\mathcal{B}}$ and $\overline{\mathcal{S}}$ can have arbitrarily large $c$-matching number. Conversely, let $X$ be a hereditary class with unbounded $c$-matching number. Assume $X$ contains none of $\mathcal{M}$, $\mathcal{B}$, $\overline{\mathcal{M}}$, $\overline{\mathcal{B}}$, i.e. there is a values of $p$ such that none of $M_p$, $B_p$, $\overline{M}_p$, $\overline{B}_p$ belongs to $X$. By assumption, $X$ contains a graph $G$ with $c\mu(G) \geq k$ for each value of $k$, i.e. $G$ contains a matching and a co-matching of size $k$. Since $k$ can be arbitrarily large and $M_p$, $B_p$ are forbidden, $G$ contains a large clique (Lemma 3). Similarly, $G$ contains a large independent set. Therefore, $X$ contains graphs with arbitrarily large complex number. But then $X$ contains either $\mathcal{S}$ or $\overline{\mathcal{S}}$ (Theorem 7). $\square$

## 2.4 Neighbourhood diversity

**Definition 3.** Let us say that two vertices $x$ and $y$ are *similar* if there is no vertex $z$ distinguishing them (i.e. if there is no vertex $z$ adjacent to exactly one of $x$ and $y$). Clearly, the similarity is an equivalence relation. We denote by $nd(G)$ the number of similarity classes in $G$ and call it the *neighbourhood diversity* of $G$.

In order to characterize the neighbourhood diversity by means of minimal hereditary classes of graphs where this parameter is unbounded, we need to introduced a few more classes of graphs.

$\mathcal{Y}$ $= Free(K_3, C_5, K_2+2K_1)$. By excluding $K_2+2K_1$ we exclude all cycles of length at least 7. Therefore, every graph in this class is bipartite. Moreover, if $G \in \mathcal{Y}$ contains at least one edge, then every vertex of this graph has at most one non-neighbour in the opposite part. In other words, this is the class of bipartite complements of graphs in $\mathcal{M}$. We denote by $Y_n$ the bipartite complement of $M_n$, i.e. a graph with $2n$ vertices each of which has exactly one non-neighbour in the opposite part. Clearly, $Y_n$ is $n$-universal for graphs in $\mathcal{Y}$.

$\mathcal{M}^*$ $= Free(diamond, 2K_2, C_4, C_5)$. Since a graph $G$ in $\mathcal{M}^*$ is $(2K_2, C_4, C_5)$-free, it is a split graph, i.e. the vertices of $G$ can be partitioned into a clique $C$ and an independent set $I$. Also, since $G$ is *diamond*-free, every vertex of a maximal clique $C$ has at most one neighbour in $I$ and every vertex of $I$ has at most one neighbour in $C$. In other words, $G$ belongs to $\mathcal{M}^*$ if and only if the removal of all edges from $C$ transforms $G$ into a graph in $\mathcal{M}$. We denote by $M_n^*$ the graph obtained from $M_n$ by fixing an arbitrary bipartition of its vertices into two independent sets and then by creating clique in one of the parts. Clearly, $M_n^*$ is $n$-universal for graphs in $\mathcal{M}^*$.

$\mathcal{T} = Free(P_4, C_4, 2K_2)$. This class is known as the class of *threshold graphs*. It is a subclass of split graphs and therefore the vertices of each threshold graph $G$ can be partition into a clique $C$ and an independent set $I$. It is well-known (and not difficult to see) that the removal of all edges from $C$ transforms $G$ into a chain graph, and vice versa. Therefore, the graph $T_n$ obtained from $C^n$ by creating a clique in one of its parts is $n$-universal for graphs in $\mathcal{T}$.

Before we provide a characterization of the neighbourhood diversity, we introduce an auxiliary parameter.

**Definition 4.** A *skew matching* in a graph $G$ is a matching $\{x_1y_1, \ldots, x_qy_q\}$ such that $y_i$ is not adjacent to $x_j$ for all $i < j$. The *complement of a skew matching* is a sequence of pairs of vertices that create a skew matching in the complement of $G$.

**Lemma 4.** *For any positive integer $m$, there exists a positive integer $r = r(m)$ such that any bipartite graph $G = (A, B, E)$ of neighbourhood diversity $r$ contains either a skew matching of size $m$ or its complement.*

*Proof.* Define $r = 2^{2m}$ and let $X$ be a set of pairwise non-similar vertices of size $r/2$ chosen from the same color class of $G$, say from $A$. Let $y_1$ be a vertex in $B$ distinguishing the set $X$ (i.e. $y_1$ has both a neighbour and a non-neighbour in $X$) and let us say that $y_1$ is *big* if the number of its neighbours in $X$ is larger than the number of its non-neighbours, and *small* otherwise. If $y_1$ is small, we arbitrarily choose its neighbour in $X$, denote it by $x_1$ and remover all neighbours of $y_1$ from $X$. If $y$ is big, we arbitrarily choose a non-neighbour of $y_1$ in $X$, denote it by $x_1$ and remove all non-neighbours of $y_1$ from $X$. Observe that $y_1$ does not distinguish the vertices in the updated set $X$.

We apply the above procedure to $X$ $2m - 1$ times and obtain in this way a sequence of $2m - 1$ pairs $x_iy_i$. If $m$ of these pairs contain small vertices $y_i$, then these pairs create a skew matching (of size $m$). Otherwise, there is a set of $m$ pairs containing big vertices $y_i$, in which case these pairs create the complement of a skew matching. □

**Lemma 5.** *For any positive integer $p$, there exists a positive integer $q = q(p)$ such that any bipartite graph $G = (A, B, E)$ of neighbourhood diversity $q$ contains either an induced $M_p$, or an induced $Z_p$ or an induced $Y_p$.*

*Proof.* Let $m = R(p+1)$ (where $R$ is the symmetric Ramsey number) and $q = 2^{2m}$. According to the proof of Lemma 4, $G$ contains a skew matching of size $m$ or its complement. If $G$ contains a skew matching $M$, we color each pair $x_i y_i$, $x_j y_j$ of edges of $M$ in two colors as follows:

- color 1 if $x_i$ is not adjacent to $y_j$,
- color 2 if $x_i$ is adjacent to $y_j$.

By Ramsey's Theorem, $M$ contains a monochromatic set $M'$ of edges of size $p+1$. If the color of each pair of edges in $M'$ is

1 then $M'$ is an induced matching of size $p+1$,

2 then the vertices of $M'$ induce a $Z_{p+1}$.

Analogously, in the case when $G$ contains the complement of a skew matching, we find either an induced $Y_{p+1}$ or an induced $Z_p$ (observe that the bipartite complement of $Z_{p+1}$ contains an induced $Z_p$). □

**Lemma 6.** *For any positive integer $p$, there exists a positive integer $Q = Q(p)$ such that every graph $G$ of neighbourhood diversity $Q$ contains one of the following nine graphs as an induced subgraph: $M_p$, $Y_p$, $Z_p$, $\overline{M}_p$, $\overline{Y}_p$, $\overline{Z}_p$, $M_p^*$, $\overline{M}_p^*$, $T_p$.*

*Proof.* Let $Q = R(q)$, where $q = 2^{2m}$ and $m = R(R(p)+1)$ (where $R$ is the symmetric Ramsey number). We choose one vertex from each similarity class of $G$ and find in the chosen set a subset $A$ of vertices that form an independent set or a clique of size $q = 2^{2m}$. Let us call the vertices of $A$ white. We denote the remaining vertices of $G$ by $B$ and call them black. Let $G' = G[A, B]$. By the choice of $A$, all vertices of this set have pairwise different neighbourhoods in $G'$. Therefore, according to the proof of Lemma 5, $G'$ contains a subgraph $G''$ inducing either $M_n$, $Y_n$, or $Z_n$ with $n = R(p)$. Among the $n$ black vertices of $G''$, we can find a subset $B'$ of vertices that form either a clique or an independent set of size $p$ in the graph $G$. Then $B'$ together with a subset of $A$ of size $p$ induce in $G$ one of the nine graphs listed in the statement of theorem. □

Since the nine graphs of Lemma 6 are universal for their respective classes, we make the following conclusion.

**Theorem 12.** *There exist exactly nine minimal classes of graphs of unbounded neighbourhood diversity:* $\mathcal{M}$, $\mathcal{Y}$, $\mathcal{Z}$, $\overline{\mathcal{M}}$, $\overline{\mathcal{Y}}$, $\overline{\mathcal{Z}}$, $\mathcal{M}^*$, $\overline{\mathcal{M}}^*$, $\mathcal{T}$.

## 2.5 VC-dimension

A set system $(X, S)$ consists of a set $X$ and a family $S$ of subsets of $X$. A subset $A \subseteq X$ is *shattered* if for every subset $B \subseteq A$ there is a set $C \in S$ such that $B = A \cap C$. The VC-dimension of $(X, S)$ is the cardinality of a largest shattered subset of $X$.

The VC-dimension of a graph $G = (V, E)$ was defined in [8] as the VC-dimension of the set system $(V, S)$, where $S$ is the family of closed neighbourhoods of vertices of $G$, i.e. $S = \{N[v] \ : \ v \in V(G)\}$. Let us denote the VC-dimension of $G$ by $vc[G]$.

In this section, we characterize VC-dimension by means of three minimal hereditary classes where this parameter is unbounded. To this end, we first redefine it in terms of open neighbourhoods as follows. Let $vc(G)$ be the size of a largest set $A$ of vertices of $G$ such that for any subset $B \subseteq A$ there is a vertex $v$ *outside of* $A$ with $B = A \cap N(v)$. In other words, $vc(G)$ is the size of a largest subset of vertices shattered by *open* neighbourhoods of vertices of $G$.

We start by showing that the two definitions are equivalent in the sense that they both are either bounded or unbounded in a hereditary class. To prove this, we introduce the following terminology. Let $A$ be a set of vertices which is shattered by a collection of neighbourhoods (open or closed). For a subset $B \subseteq A$ we will denote by $v(B)$ the vertex whose neighbourhood (open or closed) intersect $A$ at $B$. We will say that $B$ is *closed* if $v(B)$ belongs to $B$, and *open* otherwise.

**Lemma 7.** $vc(G) \leq vc[G] \leq vc(G)(vc(G) + 1) + 1$.

*Proof.* The first inequality is obvious. To prove the second one, let $A$ be a subset of $V(G)$ of size $vc[G]$ which is shattered by a collection of closed neighbourhoods. If $A$ has no closed subsets, then $vc[G] = vc(G)$. Otherwise, let $B$ be a closed subset of $A$.

Assume first that $|B| = 1$. Then $B = \{v(B)\}$ and $v(B)$ is isolated in $G[A]$, i.e. it has no neighbours in $A$. Let $C$ be the set of all such vertices, i.e. vertices each of which is a closed subset of $A$. By deleting any vertex $x$ from $C$ we obtain a new set $A$ and may assume that it has no closed subsets of size 1. Indeed, for any vertex $y \in C$ different from $x$, there is a vertex $y' \notin A$ such that $N(y') \cap A = \{x, y\}$. After the deletion of $x$ from $A$, we have $N(y') \cap A = \{y\}$ and hence $\{y\}$ is not a closed subset anymore. This discussion allows us to assume in what follows that $A$ has no closed subsets of size 1, in which case we only need to show that $vc[G] \leq vc(G)(vc(G) + 1)$.

Assume now that $B$ is a closed subset of $A$ of size at least 2. Suppose that $B - v(B)$ contains a closed subset $C$, i.e. $v(C) \in C$. Observe that $v(C)$ is adjacent to $v(B)$, as every vertex of $B - v(B)$ is adjacent to $v(B)$. But then $N[v(C)] \cap A$ contains $v(B)$ contradicting the fact that $N[v(C)] \cap A = C$. This contradiction shows that every subset of $B - v(B)$ is open, i.e. $|B - v(B)| \leq vc(G)$.

The above observation allows us to apply the following procedure: as long as $A$ contains a closed subset $B$ with at least two vertices, delete from $A$ all vertices of $B$ except for $v(B)$. Denote the resulting set by $A^*$. Assume the procedure was applied $p$ times and let $B_1, \ldots, B_p$ be the closed subsets it was applied to. It is not difficult to see that the set $\{v(B_1), \ldots, v(B_p)\}$ has no closed subsets and hence its size cannot be large than $vc(G)$, i.e. $p \leq vc(G)$. Combining, we conclude:

$$vc[G] = |A| = |A^*| + \sum_{i=1}^{p} |B_i - v(B_i)| \leq vc(G) + p \cdot vc(G) \leq vc(G)(vc(G) + 1).$$

□

This lemma allows us to assume that if $A$ is shattered, then there is a set $C$ *disjoint* from $A$ such that for any subset $B \subseteq A$ there is a vertex $v \in C$ with $B = A \cap N(v)$, in which case we will say that $A$ is shattered by $C$, or $C$ shatters $A$.

Let $B_n = (A, B, E)$ be the bipartite graph with $|A| = n$ and $|B| = 2^n$ such that all vertices of $B$ have pairwise different neighbourhood in $A$. Also, let $S_n$ be the split graph obtained from $B_n$ by creating a clique in $A$.

**Lemma 8.** *The graph $B_n$ is an $n$-universal bipartite graph, i.e. it contains every bipartite graph with $n$ vertices as an induced subgraph.*

*Proof.* Let $G$ be a bipartite graph with $n$ and with parts $A$ and $B$ of size $n_1 = |A| \leq |B| = n_2$. By adding at most $n_2$ vertices to $A$, we can guarantee that all vertices of $B$

have pairwise different neighbourhoods in $A$. Then we can extend $B$ by create a vertex $x$ with $N(x) = C$ for each subset $C \subseteq A$ for which such a vertex does not exist. In this way, we transform $G$ into a graph with at most $n + 2^n$ vertices. Clearly, $B_n$ contains this graph and hence it also contains $G$ as an induced subgraph. □

**Corollary 1.** *Every co-bipartite graph with at most $n$ vertices is contained in $\overline{B}_n$ and every split graph with at most $n$ vertices is contained in both $S_n$ and in $\overline{S}_n$.*

**Lemma 9.** *If a set $A$ shatters a set $B$ with $|B| = 2^n$, then $B$ shatters a subset $A^*$ of $A$ with $|A^*| = n$.*

*Proof.* Without loss of generality we assume that $B$ is the set of all binary sequences of length $n$. Let us denote by $B_i$ the subset of $B$ such that $b_1 \dots b_n \in B_i$ if and only if $b_i = 1$. Since $A$ shatters $B$, it contains a vertex $a_i$ with $N(a_i) \cap B = B - i$ for each $i$. We denote $A^* = \{a_1, \dots, a_n\}$ and claim that $B$ shatters $A^*$. To prove this we only need to show that any two distinct vertices $x = (b_1 \dots b_n) \in B$ and $y = (b'_1 \dots b'_n) \in B$ have different neighbourhoods in $A^*$. Since $x$ and $y$ are distinct, there exists an $i$ such that $b_i \neq b'_i$, say $b_i = 1$ and $b'_i = 0$. But then $a_i$ is adjacent to $b_i$ but not to $b'_i$. □

**Lemma 10.** *For every $n$, there exists a $k = k(n)$ such that every graph $G$ with $vc(G) = k$ contains one of $B_n, \overline{B}_n, S_n, \overline{S}_n$ as an induced subgraph.*

*Proof.* Define $k = R(2^{R(n)})$, where $R(n) = R(2, 2, n)$. Since $vc(G) = k$, there are two subsets $A$ and $B$ of $V(G)$ such that $|A| = k$ and $B$ shatters $A$. By definition of $k$, $A$ must have a subset $A'$ of size $2^{R(n)}$ which is a clique or an independent set. Clearly, $B$ shatters $A'$ and hence, by Lemma 9, $A'$ shatters a subset $B'$ of $B$ of size $R(n)$. Then $B'$ must have a subset $B''$ of size $n$ which is either a clique or an independent set. Now $G[A' \cup B'']$ is either bipartite or co-bipartite or split graph, $|B''| = n$ and $A'$ shatters $B''$. Therefore, $G[A' \cup B'']$ contains one of $B_n, \overline{B}_n, S_n, \overline{S}_n$ as an induced subgraph. □

**Theorem 13.** *The classes of bipartite, co-bipartite and split graphs are the only three minimal hereditary classes of graphs of unbounded VC-dimension.*

*Proof.* Clearly these three classes have unbounded VC-dimension, since they contain $B_n, \overline{B}_n, S_n, \overline{S}_n$ with arbitrarily large values of $n$.

Now let $X$ be a hereditary class containing none of these three classes. Therefore, there is a bipartite graph $G_1$, a co-bipartite graph $G_2$ and a split graph $G_3$ which are forbidden for $X$. Denote by $n$ the maximum number of vertices in these graphs.

Assume that VC-dimension is not bounded for graphs in $X$ and let $G \in X$ be a graph with $vc(G) = k$, where $k = k(n)$ from Lemma 10. Then $G$ contains one of $B_n, \overline{B}_n, S_n, \overline{S}_n$, say $B_n$. Since $B_n$ is $n$-universal (Lemma 8), it contains $G_1$ as an induced subgraph, which is impossible because $G_1$ is forbidden for graphs in $X$. This contradiction shows that VC-dimension is bounded in the class $X$. $\square$

# Chapter 3

# Subfactorial properties of graphs

The universal algorithm for graph coding given in Section 1.1.3 is not optimal for classes of index 1, which we call *unitary*, since equality (1.1) does not provide the asymptotic behavior of $|\log X_n|$ in this case. On the other hand, the family of unitary classes can also be split into layers according to their speed, as was shown in [62]. This paper distinguishes four lower layers: constant, polynomial, exponential and factorial. In [11], the authors partition polynomial and exponential layers further and provide each of the sublayers with a structural characterization. Instead, in what follows we give a global characterization of each of the first three layers. Moreover, with polynomial and exponential layers we associate graph parameters responsible for the speed and construct asymptotically optimal codings for classes these layers.

## 3.1 Constant layer

As mentioned earlier, complete graphs and empty (edgeless) graphs contain exactly one $n$-vertex graph for each value of $n$. The family of all hereditary classes for which the number of $n$-vertex graphs is bounded by a constant independent of $n$ is called *constant*. This family constitute the first layer of infinite hereditary classes. The following theorem characterizes the structure of such classes and also provides the minimum classes that do not belong to this layer.

**Theorem 14.** *For a hereditary class $X$, the following statements are equivalent:*

*(1)* $\log |X_n| = O(1)$*;*

*(2)* $X$ *contains finitely many graphs different from complete and empty graphs;*

*(3) none of the following classes is a subclass of* $X$*:*

- $\mathcal{R}$ *the class of graphs each of which is either an edgeless graph or a star (i.e. a graph of the form* $K_{1,n}$ *for some* $n$*),*
- $\mathcal{E}^1$ *the class of graphs with at most one edge,*
- $\overline{\mathcal{R}}$ *the class of complements of graphs in* $\mathcal{R}$*,*
- $\overline{\mathcal{E}}^1$ *the class of complements of graphs in* $\mathcal{E}^1$*.*

*Proof.* Clearly, (2) implies (1). Also, it is not difficult to see that $|\mathcal{R}(n)| = |\overline{\mathcal{R}}(n)| = n+1$ and $|\mathcal{E}^1_n| = |\overline{\mathcal{E}}^1_n| = \binom{n}{2} + 1$. Therefore, (1) implies (3). It remains to show that (3) implies (2).

Assume $X$ satisfies (3). Then there is a number $p$ such that none of the following graphs belongs to $X$: $K_{1,p}$, $\overline{K}_{1,p}$, $K_2 + O_p$, $\overline{K_2 + O_p}$.

Let $G$ be a graph in $X$ which is neither complete nor edgeless. Then $G$ contains a vertex $x$ which has both a neighbour $y$ and a non-neighbour $z$. The remaining vertices of $G$ can be partitioned (with respect to $x, y, z$) into at most eight subsets. For $U \subseteq \{x, y, z\}$, we denote $V_U = \{v \notin \{x, y, z\} \ : \ N(v) \cap \{x, y, z\} = U\}$. To prove the theorem, let us show that each of $V_U$ contains at most $R(p)$ vertices, where $R(p)$ is the symmetric Ramsey number.

If $U = \emptyset$, then $|V_U| < R(p)$ because a clique of size $p$ in $V_U$ together with $x$ create an induced $\overline{K}_{1,p}$, while an independent set of size $p$ in $V_U$ together with $x$ and $y$ create an induced $K_2 + O_p$, which is impossible because both graphs are forbidden in $X$.

If $|U| = 1$, say $U = \{x\}$, then $|V_U| < R(p)$ because a clique of size $p$ in $V_U$ together with $y$ or $z$ create an induced $\overline{K}_{1,p}$, while an independent set of size $p$ in $V_U$ together with $x$ create an induced $K_{1,p}$, which is impossible because both graphs are forbidden in $X$.

For $|U| > 1$, the result follows by complementary arguments. $\square$

## 3.2 Polynomial layer

According to Theorem 14, the four minimal classes above the constant layer are $\mathcal{R}$, $\mathcal{E}^1$, $\overline{\mathcal{R}}$ and $\overline{\mathcal{E}}^1$. Each of them contains polynomially many $n$-vertex labelled graphs and hence the layer following the constant one is polynomial. In order to provide a global characterization of the polynomial layer we use the results and notation of Sections 2.2 and 2.3.

**Theorem 15.** *For a hereditary class $X$, the following statements are equivalent:*

*(1)* $\log|X_n| = \Theta(\log n)$*;*

*(2) there exists a constant $c$ such that every graph in $X_n$ contains a similarity class of size at least $n - c$;*

*(3) complex degree and $c$-matching number are bounded for graphs in $X$;*

*(4) none of the following classes is a subclass of $X$:*

$$\mathcal{B}, \mathcal{S}, \mathcal{Q}, \mathcal{M}, \overline{\mathcal{B}}, \overline{\mathcal{S}}, \overline{\mathcal{Q}}, \overline{\mathcal{M}}.$$

*Proof.* (1) $\Rightarrow$ (4). It is not difficult to see that $|\mathcal{S}_n| = 2^n - n$, $|\mathcal{B}_n| = 2^{n-1}$ and $|\mathcal{Q}_n| = n2^{n-1} - n(n+1)/2 + 1$. Also, the number of $n$-vertex labelled graphs in $\mathcal{M}_n$ is at least $\lfloor n/2 \rfloor !$. Therefore, $X$ cannot contain any of the classes $\mathcal{B}, \mathcal{S}, \mathcal{Q}, \mathcal{M}, \overline{\mathcal{B}}, \overline{\mathcal{S}}, \overline{\mathcal{Q}}, \overline{\mathcal{M}}$.

(2) $\Rightarrow$ (1). The number of labeled graphs on $n$ vertices with a similarity class of size $n - k$ is equal to $\binom{n}{k}2^{\binom{k+1}{2}+1}$ and hence we have that $|\log X_n| = \Theta(\log n)$.

Theorems 8 and 11 imply that (3) and (4) are equivalent. Therefore, it remains to show that (3) and (4) imply (2).

Let $G$ be a graph in $X_n$. Since the $c$-matching number is bounded in $G$, we may assume, without loss of generality, that $\mu(G) \le k$ for a constant $k$. Therefore, $n - 2k$ vertices form an independent set $I$ in $G$. Also, since the complex degree is bounded in $G$, there is a constant $l$ such that every vertex $v$ outside of $I$ has at most $l$ neighbours or at most $l$ non-neighbours in $I$. By removing from $I$ either $l$ neighbours or $l$ non-neighbours of $v$, for each $v \notin I$, we transform $I$ into a similarity class of size $n - c$ where $c \le 2k(l+1)$. $\square$

### 3.2.1 Coding of graphs in classes of polynomial growth

We will now describe another algorithm for coding graphs and provide a proof that it gives an asymptotically optimal coding for graphs in classes of polynomial growth.

Given a labelled graph $G$ on $n$ vertices, define a new labelled graph $G^*$ by choosing a similarity class $K$ in $G$ of greatest size and replacing the vertices in $K$ with a single vertex $z$ labelled $n+1$. The adjacencies of vertices to $z$ are defined by $x \in G^* \setminus \{z\}$ is adjacent to $z$ if $x$ is adjacent to each vertex in $K$, and $x$ is non adjacent to $z$ if $x$ is non adjacent to each vertex in $K$.

Now each vertex in $G$ has a label that can be described with a binary word of length at most $\log n$. Let us denote by $\psi^{(1)}$ the binary word consisting of the labels of the vertices in $G \setminus K$ in increasing order. Let $\psi^{(2)} = \phi^c_{n-|K|+1}(G^*)$ and let $\psi^{(3)} = 0$ if $K$ forms an independent set in $G$ otherwise let $\psi^{(3)} = 1$ if $K$ forms a clique in $G$.[1] We now define

$$\phi^1_n(G) = \psi^{(1)}\psi^{(2)}\psi^{(3)}, \qquad \Phi^1 = \{\phi^1_n \ : \ n = 1, 2, 3, \ldots\}.$$

**Theorem 16.** $\Phi^1$ *is an asymptotically optimal coding for any hereditary class* $X$ *with* $\log |X_n| = \Theta(\log n)$.

*Proof.* Let $X$ be a hereditary class satisfying $|\log X_n| = \Theta(\log n)$. We know by Theorem 15 that there exists a constant $c$ such that every graph in $X_n$ contains a similarity class of size at least $n - c$. Partition $X$ into sets $X^0, X^1, \ldots, X^c$ with $X^i$ defined to be the set of graphs $G \in X$ with the size of the largest similarity class contained in $G$ precisely equal to $n - i$ for $0 \leq i \leq c$. Define $t$ to be the largest number such that the set $X^t$ contains infinitely many graphs, i.e. there exists an $N \in \mathbb{N}$ such that for all $n > N$ every graph in $X_n$ has a similarity class of size at least $n - t$. Hence for $n > N$ we have that

$$\max_{G \in X_n} |\phi^1_n(G)| \leq t \log n + \binom{t+1}{2} + 1.$$

We also know that $|X_n| \geq \binom{n}{t}$ and hence

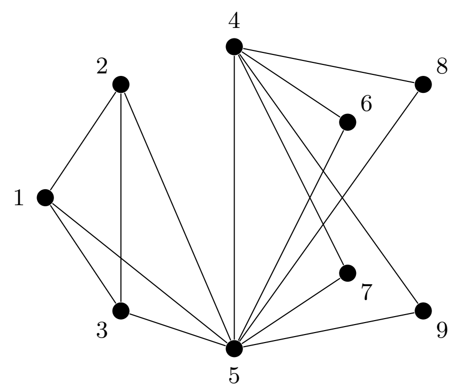
$$\log |X_n| \geq t \log n.$$

Therefore,

$$\lim_{n \to \infty} \frac{\max\limits_{G \in X_n} |\phi^1_n(G)|}{\log |X_n|} \leq 1.$$

[1] To create a unique coding we assume that a similarity class of size one is a clique.

However, each graph in $X_n$ must have a unique coding so one must have length at least $\log |X_n|$, hence the limit is equal to 1. $\square$

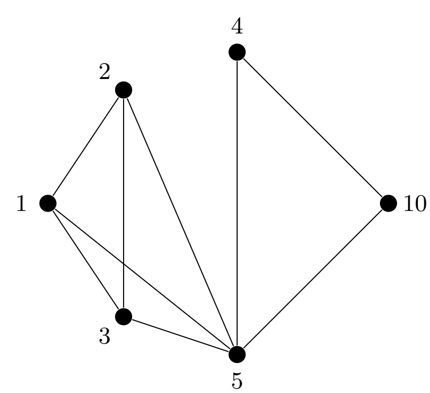



$G$

**Example.** Let $G$ be the graph (pictured above) with 9 vertices $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and the following canonical code:

$$11010000 \quad 1010000 \quad 010000 \quad 11111 \quad 1111 \quad 000 \quad 00 \quad 0.$$

The largest size of a similarity class in $G$ is four, and the vertices $6, 7, 8$ and $9$ form such a similarity class. Hence, the vertices in $G \setminus K$ are $1, 2, 3, 4$ and $5$. This tells us that $\psi^{(1)} = 0001 \quad 0010 \quad 0011 \quad 0100 \quad 0101$. Now replacing the vertices $6, 7, 8$ and $9$ with one vertex labelled 10 gives us $G^*$ (pictured below).




$G^*$

Looking at $G^*$ gives us $\psi^{(2)} = 11010 \quad 1010 \quad 010 \quad 11 \quad 1$ and as the vertices $6, 7, 8$ and $9$ form an independent set in $G$ we have $\psi^{(3)} = 0$. So,

$$\phi_9^1(G) = 00010010001101000101 \quad 110101010010111 \quad 0.$$

In this particular case, the method for coding a graph described in this section gives a binary word of equal length to the one given by the canonical code.

## 3.3 Exponential layer

Theorem 15 tells us that there are eight minimal classes above the polynomial layer, namely $\mathcal{B}, \mathcal{S}, \mathcal{Q}, \mathcal{M}, \overline{\mathcal{B}}, \overline{\mathcal{S}}, \overline{\mathcal{Q}}$, and $\overline{\mathcal{M}}$. Each of these classes contains at least exponentially many $n$-vertex labelled graphs and hence the next layer is the exponential layer. In order to provide a global characterization of this layer we use the results and notation of Sections 2.4.

**Theorem 17.** *For a hereditary class $X$, the following statements are equivalent:*

*(1)* $\log |X_n| = \Theta(n)$*;*

*(2) the neighbourhood diversity is bounded for graphs in $X$;*

*(3) none of the following classes is a subclass of $X$:*

$$\mathcal{M}, \mathcal{Y}, \mathcal{Z}, \overline{\mathcal{M}}, \overline{\mathcal{Y}}, \overline{\mathcal{Z}}, \mathcal{M}^*, \overline{\mathcal{M}}^*, \mathcal{T}.$$

*Proof.* As stated previously, the number of $n$-vertex labelled graphs in $\mathcal{M}_n$ is at least $\lfloor n/2 \rfloor !$. It is not difficult to see that the same is true for all other classes listed in (3). Therefore, (1) implies (3).

By Theorem 12, (3) is equivalent to (2). Finally, the number of labeled graphs on $n$ vertices with at most $k$ similarity classes is at most $k^n 2^{\binom{k}{2}+k}$. Therefore, (2) implies (1). □

### 3.3.1 Coding of graphs in classes of exponential growth

We will now describe a third algorithm for coding graphs and provide a proof that it gives an asymptotically optimal coding for graphs in classes of exponential growth.

Let $G$ be an arbitrary graph with $n$ vertices. Partition $G$ into maximal similarity classes, and let $m$ be the number of these similarity classes. Note that $m \leq n$ and let $\psi^{(4)}$ be the binary representation of $m$ using $\log n$ bits. Now define a new labelled graph $G'$ by contracting each similarity class to a single vertex, hence two vertices $x$ and $y$ in $G'$

are adjacent if and only if all of the vertices in the class represented by $x$ are adjacent to all of the vertices in the class represented by $y$. Now fix an ordering on the vertices of $G'$ and assign to each vertex $v$ of $G$ the binary word $\psi_v^{(5)}$ of length $\log m$ that denotes the label of the vertex which represents the similarity class $v$ belongs to. Finally let $M_{G'}$ be a modified adjacency matrix of $G'$ where the entries on the main diagonal are defined by setting the element $(i,i)$ of the matrix to be 0 if the similarity class in $G$ represented by the vertex $i$ in $G'$ is an independent set and 1 if it is a clique, for $i = 1, 2, \ldots, m$. Now let $\psi^{(6)}$ be the binary word obtained by reading the elements of the modified adjacency matrix above and including the main diagonal, row by row starting with the first. We now define

$$\phi_n^2(G) = \psi^{(4)}\psi_{v_1}^{(5)}\psi_{v_2}^{(5)}\ldots\psi_{v_n}^{(5)}\psi^{(6)}, \qquad \Phi^2 = \{\phi_n^2 \ : \ n = 1, 2, 3, \ldots\}.$$

In order to restore the graph $G$ from its code $\phi_n^2(G)$ first determine the value of $m$ by reading the first $\log n$ bits of the code. We now know that the rest of the code consists of $n$ binary words of length $\log m$ and a binary word of length $\binom{m+1}{2}$ which describes the modified adjacency matrix of $G'$. Now two vertices $x, y \in G$ are adjacent if and only if $M_{G'}(i,j) = 1$ where $i$ is the number represented by $\psi_x^{(5)}$ and $j$ is the number represented by $\psi_y^{(5)}$. We will now show that this coding is asymptotically optimal for any class of graphs with exponential growth.

**Theorem 18.** $\Phi^2$ *is an asymptotically optimal coding for any hereditary class* $X$ *with* $\log|X_n| = \Theta(n)$.

*Proof.* We know by Theorem 17 that there exists a constant $k$ such that every graph in $X_n$ contains $k$ similarity classes. Partition $X$ into sets $X^0, X^1, \ldots, X^k$ with $X^i$ defined to be the set of graphs $G \in X$ which contain at most $i$ maximal similarity classes. Define $t$ to be the largest number such that the set $X^t$ contains infinitely many graphs, i.e. $t$ is the largest number such that for any $s \in \mathbb{N}$ there are infinitely many graphs in $X$ which contain $t$ similarity classes of size at least $s$.

Now pick an $N \in \mathbb{N}$ such that for all $n > N$ every graph in $X_n$ contains at most $t$ maximal similarity classes. Hence for $n > N$ we have that

$$\max_{G \in X_n} |\phi_n^2(G)| \leq \log n + n \log t + \binom{t+1}{2}. \tag{3.1}$$

It is not hard to see that the class $X$ must contain at least

$$\frac{n!}{(\lfloor \frac{n}{t} \rfloor !)^t}$$

different $n$-vertex labelled graphs, for $n > N$. Hence,

$$\log |X_n| \geq \log n! - t \log(\lfloor \frac{n}{t} \rfloor !) \geq \log n! - t \log((\frac{n}{t})^{\frac{n}{t}}) = \log n! - n \log n + n \log t. \quad (3.2)$$

We know by Stirling's approximation that

$$\log n! \sim n \log n - n \log e + \frac{1}{2} \log 2\pi n$$

so

$$\frac{\log n!}{n \log n} \sim 1 - \frac{\log e}{\log n} + \frac{1}{2} \frac{\log 2\pi n}{n \log n}.$$

Therefore,

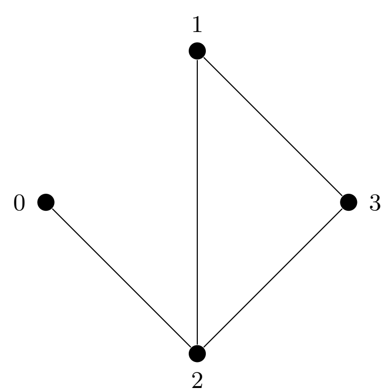
$$\lim_{n \to \infty} \frac{\log n!}{n \log n} = 1 \quad (3.3)$$

So combining (3.1), (3.2) and (3.3) we get

$$\lim_{n \to \infty} \frac{\max\limits_{G \in X_n} |\phi_n^2(G)|}{\log |X_n|} \leq \frac{\log n + n \log t + \binom{t+1}{2}}{\log n! - n \log n + n \log t} = \frac{\log n + n \log t + \binom{t+1}{2}}{n \log t} = 1 + \frac{\log n}{n \log t} + \frac{\binom{t+1}{2}}{n \log t} = 1.$$

However, each graph in $X_n$ must have a unique coding so one must have length at least $\log |X_n|$, hence the limit is equal to 1. $\square$

**Example.** Let $G$ be the graph with 9 vertices $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ used in the example in section 3.2.1. We will now show how this graph can be coded using the method introduced in this section. $G$ has four maximal similarity classes, the sets $\{1, 2, 3\}$, $\{4\}$, $\{5\}$ and $\{6, 7, 8, 9\}$. So $\psi^{(4)} = 0011$. Contracting these classes to a single vertex gives us the graph $G'$ (pictured below).




$G'$

As the graph $G'$ has four vertices, each vertex in $G$ will be assigned one of the four binary words $00, 01, 10$ or $11$. Also looking at $G'$, and noting that the similarity class defined by $\{6, 7, 8, 9\}$ is the only similarity class which is an independent set, gives us $\psi^{(6)} = 1010\ \ 111\ \ 11\ \ 0$. Hence,

$$\phi_9^2(G) = 0011\ \ 00\ \ 00\ \ 00\ \ 01\ \ 10\ \ 11\ \ 11\ \ 11\ \ 11\ \ 1010111110.$$

In this particular case, the method for coding a graph described in this section gives a binary word which uses four less bits than the one given by the canonical code.

# Chapter 4

# Hereditary properties with speeds below the Bell number

Recall that the Bell number $B_n$, defined as the number of ways to partition a set of $n$ labelled elements, satisfies the asymptotic formula $\ln B_n/n = \ln n - \ln\ln n + \Theta(1)$.

Balogh, Bollobás and Weinreich [13] showed that if the speed of a hereditary graph property is at least $n^{(1-o(1))n}$, then it is actually at least $B_n$; hence we call any such property a *property above the Bell number*. Note that this includes hereditary properties whose speed is exactly equal to the Bell numbers (such as the class of disjoint unions of cliques).

### 4.0.1 $(\ell, d)$-graphs and sparsification

Given a graph $G$ and two vertex subsets $U, W \subset V(G)$, define $\Delta(U, W) = \max\{|N(u) \cap W|, |N(w) \cap U| : u \in U, w \in W\}$. With $\overline{N}(u) = V(G) \backslash (N(u) \cup \{u\})$, let $\overline{\Delta}(U, W) = \max\{|\overline{N}(u) \cap W|, |\overline{N}(w) \cap U| : w \in W, u \in U\}$. Note that $\Delta(U, U)$ is simply the maximum degree in $G[U]$.

**Definition 5.** Let $G$ be a graph. A partition $\pi = \{V_1, V_2, \ldots, V_{\ell'}\}$ of $V(G)$ is an $(\ell, d)$-*partition* if $\ell' \leq \ell$ and for each pair of not necessarily distinct integers $i, j \in \{1, 2, \ldots, \ell'\}$ either $\Delta(V_i, V_j) \leq d$ or $\overline{\Delta}(V_i, V_j) \leq d$. We call the sets $V_i$ *bags*. A graph $G$ is an $(\ell, d)$-*graph* if it admits an $(\ell, d)$-partition.

If $\Delta(V_i, V_j) \leq d$, we say $V_i$ is *$d$-sparse* with respect to $V_j$, and if $\overline{\Delta}(V_i, V_j) \leq d$, we say $V_i$ is *$d$-dense* with respect to $V_j$. We will also say that the pair $(V_i, V_j)$ is *$d$-sparse* or *$d$-dense*, respectively. Note that if the bags are large enough (i.e., $\min\{|V_i|\} > 2d+1$), the terms $d$-dense and $d$-sparse are mutually exclusive.

**Definition 6.** A *strong $(\ell, d)$-partition* is an $(\ell, d)$-partition each bag of which contains at least $5 \times 2^\ell d$ vertices; a *strong $(\ell, d)$-graph* is a graph which admits a strong $(\ell, d)$-partition.

Given any strong $(\ell, d)$-partition $\pi = \{V_1, V_2, \ldots, V_{\ell'}\}$ we define an equivalence relation $\sim$ on the bags by putting $V_i \sim V_j$ if and only if for each $k$, either $V_k$ is $d$-dense with respect to both $V_i$ and $V_j$, or $V_k$ is $d$-sparse with respect to both $V_i$ and $V_j$. Let us call a partition $\pi$ *prime* if all its $\sim$-equivalence classes are of size 1. If the partition $\pi$ is not prime, let $p(\pi)$ be the partition consisting of unions of bags in the $\sim$-equivalence classes for $\pi$.

We proceed to showing that the partition $p(\pi)$ of a strong $(\ell, d)$-graph does not depend on the choice of a strong $(\ell, d)$-partition $\pi$. The following three lemmas are the ingredients for the proof of this result.

**Lemma 11.** *Consider any strong $(\ell, d)$-graph $G$ with any strong $(\ell, d)$-partition $\pi$. Then $p(\pi)$ is an $(\ell, \ell d)$-partition with at least $5 \times 2^\ell d$ vertices in each bag.*

*Proof.* Consider two bags $W_1, W_2 \in p(\pi)$. By definition $W_i = \bigcup_{s \in S_i} V_s$ for some $S_i \subset \{1, 2, \ldots, \ell'\}$, $i = 1, 2$. Also, by the definition of the partition, for all $(s_1, s_2) \in S_1 \times S_2$ the pairs $(V_{s_1}, V_{s_2})$ are all either $d$-dense or $d$-sparse. If they are $d$-sparse, then for any $s_1 \in S_1$ we have $\Delta(V_{s_1}, W_2) \leq \sum_{s_2 \in S_2} \Delta(V_{s_1}, V_{s_2}) \leq |S_2|d$. Since this holds for every $s_1 \in S_1$, for all $x \in W_1$ we have that $|N(x) \cap W_2| \leq |S_2|d$. Similarly we conclude that for all $x \in W_2$ we have $|N(x) \cap W_1| \leq |S_1|d$. Therefore, $\Delta(W_1, W_2) \leq \max(|S_1|, |S_2|)d \leq \ell d$. If the pairs of bags are $d$-dense, a similar argument proves that $\overline{\Delta}(W_1, W_2) \leq \ell d$. Hence the partition $p(\pi)$ is an $(\ell, \ell d)$-partition. As it is obtained by unifying some bags from a strong $(\ell, d)$-partition, we conclude that each bag is of size at least $5 \times 2^\ell d$. □

**Lemma 12** ([13, Lemma 10])**.** *Let $G$ be a graph with an $(\ell, d)$-partition $\pi$. If two vertices $x, y \in G$ are in the same bag $V_k$, then the symmetric difference of their neighbourhoods $N(x) \ominus N(y)$ is of size at most $2\ell d$.* □

**Lemma 13.** *Let $G$ be a graph with a strong $(\ell, d)$-partition $\pi$. If two vertices $x, y \in V(G)$ belong to different bags of the partition $p(\pi)$, then the symmetric difference of their neighbourhoods $N(x) \ominus N(y)$ is of size at least $5 \times 2^{\ell} d - 2d$.*

*Proof.* Take any two vertices $x \in V_i$ and $y \in V_j$ with bags $V_i$ and $V_j$ belonging to different $\sim$-equivalence classes. Then there is a bag $V_k$ such that one of the pairs $(V_i, V_k)$ and $(V_j, V_k)$ is $d$-dense and the other one is $d$-sparse; without loss of generality, suppose that $(V_i, V_k)$ is $d$-sparse and $(V_j, V_k)$ is $d$-dense. Then, in particular, $|N(x) \cap V_k| \leq d$ and $|N(y) \cap V_k| \geq |V_k| - d$. Hence $|N(x) \ominus N(y)| \geq |N(y) \setminus N(x)| \geq |V_k| - 2d \geq 5 \times 2^{\ell} d - 2d$. $\square$

We are now ready to prove the uniqueness of $p(\pi)$.

**Theorem 19.** *Let $G$ be a strong $(\ell, d)$-graph with strong $(\ell, d)$-partitions $\pi$ and $\pi'$. Then $p(\pi) = p(\pi')$.*

*Proof.* Assume two vertices $x, y \in V(G)$ are in the same bag of the partition $p(\pi)$. By Lemma 11, $p(\pi)$ is an $(\ell, \ell d)$-partition, so applying Lemma 12 to $p(\pi)$ we obtain $|N(x) \ominus N(y)| \leq 2\ell(\ell d) = 2\ell^2 d < 5 \times 2^{\ell} d - 2d$. Thus by Lemma 13, $x$ and $y$ are in the same bag of $p(\pi')$. Hence, using symmetry, $x$ and $y$ are in the same bag of $p(\pi)$ if and only if they are in the same bag of $p(\pi')$. We deduce that the partitions are the same, i.e., $p(\pi) = p(\pi')$. $\square$

With any strong $(\ell, d)$-partition $\pi = \{V_1, V_2, \ldots, V_{\ell'}\}$ of a graph $G$ we can associate a *density graph* (with loops allowed) $H = H(G, \pi)$: the vertex set of $H$ is $\{1, 2, \ldots, \ell'\}$ and there is an edge joining $i$ and $j$ if and only if $(V_i, V_j)$ is a $d$-dense pair (so there is a loop at $i$ if and only if $V_i$ is $d$-dense).

For a graph $G$, a vertex partition $\pi = \{V_1, V_2, \ldots, V_{\ell'}\}$ of $G$ and a graph with loops allowed $H$ with vertex set $\{1, 2, \ldots, \ell'\}$, we define (as in [11]) the *$H, \pi$-transform* $\psi(G, \pi, H)$ to be the graph obtained from $G$ by replacing $G[V_i, V_j]$ with its bipartite complement for every pair $(V_i, V_j)$ for which $ij$ is an edge of $H$, and replacing $G[V_i]$ with its complement for every $V_i$ for which there is a loop at the vertex $i$ in $H$.

Moreover, if $\pi$ is a strong $(\ell, d)$-partition we define $\phi(G, \pi) = \psi(G, \pi, H(G, \pi))$. Note that $\pi$ is a strong $(\ell, d)$-partition for $\phi(G, \pi)$ and each pair $(V_i, V_j)$ is $d$-sparse in $\phi(G, \pi)$.

We now show that the result of this "sparsification" does not depend on the initial strong $(\ell, d)$-partition.

**Proposition 20.** *Let $G$ be a strong $(\ell, d)$-graph. Then for any two strong $(\ell, d)$-partitions $\pi$ and $\pi'$, the graph $\phi(G, \pi)$ is identical to $\phi(G, \pi')$.*

*Proof.* Suppose that $\pi = \{U_1, U_2, \ldots, U_{\hat{\ell}}\}$ and $\pi' = \{V_1, V_2, \ldots, V_{\hat{\ell}'}\}$. By Theorem 19, $p(\pi) = p(\pi') = \{W_1, W_2, \ldots, W_{\hat{\ell}''}\}$. Consider two vertices $x$, $y$ of $G$. Let $i, j, i', j', i'', j''$ be the indices such that $x \in U_i$, $x \in V_{i'}$, $x \in W_{i''}$, $y \in U_j$, $y \in V_{j'}$, $y \in W_{j''}$. As the partitions have at least $5 \times 2^\ell d$ vertices in each bag, $\ell d$-dense and $\ell d$-sparse are mutually exclusive properties. Hence the pair $(U_i, U_j)$ is $d$-sparse if and only if $(W_{i''}, W_{j''})$ is $\ell d$-sparse if and only if $(V_{i'}, V_{j'})$ is $d$-sparse; and analogously for dense pairs. Therefore $xy$ is an edge of $\phi(G, \pi)$ if and only if it is an edge of $\phi(G, \pi')$. □

Proposition 20 motivates the following definition, originating from [11].

**Definition 7.** For a strong $(\ell, d)$-graph $G$, its *sparsification* is $\phi(G) = \phi(G, \pi)$ for any strong $(\ell, d)$-partition $\pi$ of $G$.

### 4.0.2 Distinguishing number $k_{\mathcal{X}}$

In this section, we discuss the distinguishing number of a hereditary graph property, which is an important parameter introduced by Balogh, Bollobás and Weinreich in [11].

Given a graph $G$ and a set $X = \{v_1, \ldots, v_t\} \subseteq V(G)$, we say that the disjoint subsets $U_1, \ldots, U_m$ of $V(G)$ are *distinguished* by $X$ if for each $i$, all vertices of $U_i$ have the same neighbourhood in $X$, and for each $i \neq j$, vertices $x \in U_i$ and $y \in U_j$ have different neighbourhoods in $X$. We also say that $X$ *distinguishes* the sets $U_1, U_2, \ldots, U_m$.

**Definition 8.** Given a hereditary property $\mathcal{X}$, we define the *distinguishing number* $k_{\mathcal{X}}$ as follows:

- If for all $k, m \in \mathbb{N}$ we can find a graph $G \in \mathcal{X}$ that admits some $X \subset V(G)$ distinguishing at least $m$ sets, each of size at least $k$, then put $k_{\mathcal{X}} = \infty$.
- Otherwise, there must exist a pair $(k, m)$ such that any vertex subset of any graph $G \in \mathcal{X}$ distinguishes at most $m$ sets of size at least $k$. We define $k_{\mathcal{X}}$ to be the minimum value of $k$ in all such pairs.

In [11], Balogh, Bollobás and Weinreich show that the speed of any hereditary property $\mathcal{X}$ with $k_{\mathcal{X}} = \infty$ is above the Bell number. To study the classes with $k_{\mathcal{X}} < \infty$, in the next sections we will need two results from their paper.

**Lemma 14** ([11], Lemma 27)**.** *If $\mathcal{X}$ is a hereditary property with finite distinguishing number $k_{\mathcal{X}}$, then there exist absolute constants $\ell_{\mathcal{X}}$, $d_{\mathcal{X}} \leq k_{\mathcal{X}}$ and $c_{\mathcal{X}}$ such that for all $G \in \mathcal{X}$, the graph $G$ contains an induced subgraph $G'$ such that $G'$ is an $(\ell_{\mathcal{X}}, d_{\mathcal{X}})$-graph and $|V(G) \backslash V(G')| < c_{\mathcal{X}}$.* □

By removing all the small bags with fewer than $5 \times 2^{\ell_{\mathcal{X}}} d_{\mathcal{X}}$ vertices, which affects only the constant $c_{\mathcal{X}}$, we can actually assume that the graph $G'$ is a *strong* $(\ell_{\mathcal{X}}, d_{\mathcal{X}})$-graph. This observation allows us to strengthen Lemma 14 as follows.

**Lemma 15.** *If $\mathcal{X}$ is a hereditary property with finite distinguishing number $k_{\mathcal{X}}$, then there exist absolute constants $\ell_{\mathcal{X}}$, $d_{\mathcal{X}}$ and $c_{\mathcal{X}}$ such that for all $G \in \mathcal{X}$, the graph $G$ contains an induced subgraph $G'$ such that $G'$ is a strong $(\ell_{\mathcal{X}}, d_{\mathcal{X}})$-graph and $|V(G) \backslash V(G')| < c_{\mathcal{X}}$.* □

Finally, we will use this theorem:

**Theorem 21** ([11], Theorem 28)**.** *Let $\mathcal{X}$ be a hereditary property with $k_{\mathcal{X}} < \infty$. Then $\mathcal{X}_n \geq n^{(1+o(1))n}$ if and only if for every $m$ there exists a strong $(\ell_{\mathcal{X}}, d_{\mathcal{X}})$-graph $G$ in $\mathcal{X}$ such that its sparsification $\phi(G)$ has a component of order at least $m$.* □

## 4.1 Structure of minimal classes above the Bell number

In this section, we describe minimal classes with speed above the Bell number. In [13], Balogh, Bollobás and Weinreich characterised all minimal classes with infinite distinguishing number. In Section 4.1.1 we report this result and prove additionally that each of these classes can be characterised by finitely many forbidden induced subgraphs. Then in Section 4.1.2 we move on to the case of finite distinguishing number, which had been left open in [13].

### 4.1.1 Infinite distinguishing number

**Theorem 22** (Balogh–Bollobás–Weinreich [13])**.** *Let $\mathcal{X}$ be a hereditary graph property with $k_{\mathcal{X}} = \infty$. Then $\mathcal{X}$ contains at least one of the following (minimal) classes:*

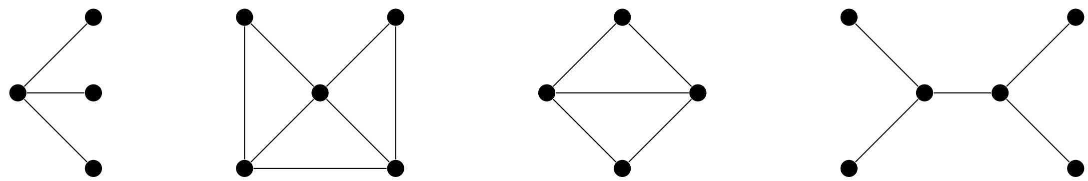
- *the class* $\mathcal{K}_1$ *of all graphs each of whose connected components is a clique;*
- *the class* $\mathcal{K}_2$ *of all star forests;*
- *the class* $\mathcal{K}_3$ *of all graphs whose vertex set can be split into an independent set* $I$ *and a clique* $Q$ *so that every vertex in* $Q$ *has at most one neighbour in* $I$*;*
- *the class* $\mathcal{K}_4$ *of all graphs whose vertex set can be split into an independent set* $I$ *and a clique* $Q$ *so that every vertex in* $I$ *has at most one neighbour in* $Q$*;*
- *the class* $\mathcal{K}_5$ *of all graphs whose vertex set can be split into two cliques* $Q_1$, $Q_2$ *so that every vertex in* $Q_2$ *has at most one neighbour in* $Q_1$*;*
- *the class* $\mathcal{K}_6$ *of all graphs whose vertex set can be split into two independent sets* $I_1$, $I_2$ *so that the neighbourhoods of the vertices in* $I_1$ *are linearly ordered by inclusion (that is, the class of all* chain graphs*);*
- *the class* $\mathcal{K}_7$ *of all graphs whose vertex set can be split into an independent set* $I$ *and a clique* $Q$ *so that the neighbourhoods of the vertices in* $I$ *are linearly ordered by inclusion (that is, the class of all* threshold graphs*);*
- *the class* $\overline{\mathcal{K}_i}$ *of all graphs whose complement belongs to* $\mathcal{K}_i$ *as above, for some* $i \in \{1, 2, \ldots, 6\}$ *(note that the complementary class of* $\mathcal{K}_7$ *is* $\mathcal{K}_7$ *itself).*

Aiming to prove that each of the classes above is defined by forbidding finitely many induced subgraphs, we first state an older result by Fldes and Hammer about split graphs of which we make use in our proof. A *split graph* is a graph whose vertex set can be split into an independent set and a clique.

**Theorem 23** ([33])**.** *The class of all split graphs is exactly the class* $\mathrm{Free}(2K_2, C_4, C_5)$*.*

Before showing the characterisation of the classes $\mathcal{K}_1$–$\mathcal{K}_6$ in terms of forbidden induced subgraphs, we introduce some of the less commonly appearing graphs: the *claw* $K_{1,3}$, the 3-*fan* $F_3$, the *diamond* $K_4^-$, and the graph $H_6$ (Fig. 4.1).

**Theorem 24.** *Each of the classes of Theorem 22 is defined by finitely many forbidden induced subgraphs.*



$K_{1,3}$ $F_3$ $K_4^-$ $H_6$

Figure 4.1: Some small graphs

*Proof.* First, observe that if we define $\overline{\mathcal{X}}$ as the class of the complements of all graphs in $\mathcal{X}$, then $\overline{\text{Free}(\mathcal{F})} = \text{Free}(\overline{\mathcal{F}})$. Hence if each class $\mathcal{K}_i$ is defined by finitely many forbidden induced subgraphs, then so is each $\overline{\mathcal{K}_i}$.

(a) $\mathcal{K}_1 = \text{Free}(P_3)$: It is trivial to check that $P_3$ does not belong to $\mathcal{K}_1$, and any graph not containing an induced $P_3$ must be a collection of cliques.

(b) $\mathcal{K}_2 = \text{Free}(K_3, P_4, C_4)$: Obviously, none of the graphs $K_3$, $P_4$, $C_4$ belongs to $\mathcal{K}_2$. Let $G \in \text{Free}(K_3, P_4, C_4)$. Since every cycle of length at least 5 contains $P_4$, $G$ does not contain any cycles; thus $G$ is a forest. The absence of a $P_4$ implies that the diameter of any component of $G$ is at most 2, hence $G$ is a star forest.

(c) $\mathcal{K}_3 = \text{Free}(\mathcal{F})$ for $\mathcal{F} = \{2K_2, C_4, C_5, K_{1,3}, F_3\}$: It is easy to check that none of the forbidden graphs belong to $\mathcal{K}_3$. Let $G \in \text{Free}(\mathcal{F})$. By Theorem 23, $G$ is a split graph. Split $G$ into a maximal clique $Q$ and an independent set $I$. Suppose, for the sake of contradiction, that $Q$ contains a vertex $u$ with two neighbours $a, b \in I$. As we took $Q$ to be a maximal clique, $a$ has a non-neighbour $v$ and $b$ has a non-neighbour $w$ in $Q$. If $a$, $w$ are not adjacent, then the vertices $a, b, u, w$ induce a claw in $G$; if $b$, $v$ are not adjacent, then the vertices $a, b, u, v$ induce a claw in $G$; otherwise the vertices $a, b, u, v, w$ induce a 3-fan in $G$. In either case we get a contradiction.

(d) $\mathcal{K}_4 = \text{Free}(\mathcal{F})$ for $\mathcal{F} = \{2K_2, C_4, C_5, K_4^-\}$: Again, it is easy to check that none of the forbidden graphs belong to $\mathcal{K}_4$. Let $G \in \text{Free}(\mathcal{F})$. By Theorem 23, $G$ is a split graph. Just like before, split $G$ into a maximal clique $Q$ and an independent set $I$. Suppose that some vertex $u$ in $I$ has two neighbours $a, b$ in $Q$. By maximality of $Q$, $u$ also has a non-neighbour $c$ in $Q$. But then the vertices $a, b, c, u$ induce a $K_4^-$ in $G$, a contradiction.

(e) The class $\overline{\mathcal{K}_5}$ of the complements of the graphs in $\mathcal{K}_5$ is characterised as the class of all (bipartite) graphs whose vertex set can be split into independent sets $I_1$, $I_2$ so that each vertex in $I_2$ has at most one non-neighbour in $I_1$. We show that $\overline{\mathcal{K}_5} = \text{Free}(\mathcal{F})$ for $\mathcal{F} = \{K_3, C_5, P_4 + K_1, 2K_2 + K_1, C_4 + K_2, C_4 + 2K_1, H_6\}$. The reader will kindly check that indeed no graph in $\mathcal{F}$ belongs to $\overline{\mathcal{K}_5}$.

Consider some $G \in \text{Free}(\mathcal{F})$; we will show that $G \in \overline{\mathcal{K}_5}$. Observe that $\mathcal{F}$ prevents $G$ from having an odd cycle, thus $G$ is bipartite. We distinguish three cases depending on the structure of the connected components of $G$.

First, suppose that $G$ has at least two non-trivial connected components (that is, connected components that are not just isolated vertices). Because $G$ is $(2K_2 + K_1)$-free, it only has two connected components in all. Being $C_4$- and $P_4$-free, each component is necessarily a star. Observe that any graph consisting of one or two stars belongs to $\overline{\mathcal{K}_5}$.

Next assume that $G$ has only one non-trivial component and some isolated vertices. The non-trivial component is bipartite and $P_4$-free, so it is a biclique. If this biclique contains $C_4$, then $G$ only contains one other isolated vertex; any graph consisting of a biclique and one isolated vertex is in $\overline{\mathcal{K}_5}$. Otherwise the biclique is a star; any graph consisting of a star and one or more isolated vertices belongs to $\overline{\mathcal{K}_5}$.

Finally, consider $G$ that is connected. We will show that for any two vertices of $G$ in different parts, one of them must have at most one non-neighbour in the opposite part. Suppose this is not true and there are $x, y \in V(G)$ in different parts such that both $x$ and $y$ have at least two non-neighbours in the opposite part. Assume first that $x$ and $y$ are adjacent. Let $a$ and $b$ be two non-neighbours of $x$, and let $c$ and $d$ be two non-neighbours of $y$. Then the graph induced by $a, b, c$ and $d$ cannot be a $C_4$, $P_4$, $P_3 + K_1$, $2K_2$ or $K_2 + 2K_1$, because $G$ is $(P_4 + K_1, 2K_2 + K_1, C_4 + K_2)$-free. Hence $a, b, c$ and $d$ must induce a $4K_1$. As $G$ is connected, $a$ must have a neighbour, say $w$. However, the vertices $x, y, a, c$ and $w$ induce a $P_4 + K_1$ if $y$ and $w$ are adjacent and they induce a $2K_2 + K_1$ if $y$ and $w$ are not adjacent. Therefore, $x$ and $y$ must be non-neighbours.

By assumption, $x$ has another non-neighbour $a \neq y$ in the opposite part, and $y$ has another non-neighbour $b \neq x$ in the opposite part. As $G$ is connected, $x$ must have a neighbour, say $u$. If $a$ and $b$ are adjacent, then $x, y, u, a$ and $b$ induce a $2K_2 + K_1$ if $u$ is not adjacent to $b$, and they induce a $P_4 + K_1$ if $u$ is adjacent to $b$. Both cases lead to a contradiction as $G$ is $(P_4 + K_1, 2K_2 + K_1)$-free, hence $a$ and $b$ cannot be adjacent. Now, as $G$ is connected, $y$ must also have a neighbour, say $v$. If $u$ is not adjacent to $b$, then $x, y, u, v$ and $b$ induce either a $2K_2 + K_1$ or a $P_4 + K_1$, hence $u$ and $b$ must be adjacent. By a symmetric argument, $v$ is adjacent to $a$. Now $u$ and $v$ must be non-adjacent: otherwise $x, y, u, v, a$ and $b$ induce an $H_6$.

This argument shows that any neighbour of $x$ must also be a neighbour of $b$, any neighbour of $y$ must also be a neighbour of $a$, and that any neighbour of $x$ cannot be adjacent to any neighbour of $y$. This means that the shortest induced path between $x$ and $y$ must contain a $P_6$, which is a contradiction as $G$ is $(P_4+K_1)$-free. Therefore, either $x$ or $y$ must have at most one non-neighbour. This implies that $G$ can be split into two independent sets $I_1$, $I_2$ such that every vertex in $I_2$ has at most one non-neighbour in $I_1$, so $G$ belongs to $\overline{\mathcal{K}_5}$.

(f) Chain graphs are characterised by finitely many forbidden induced subgraphs by a result of Yannakakis [66]; namely, $\mathcal{K}_6 = \mathrm{Free}(2K_2, K_3, C_5)$.

(g) Threshold graphs are characterised by finitely many forbidden induced subgraphs by a result of Chvátal and Hammer [20]; namely, $\mathcal{K}_7 = \mathrm{Free}(2K_2, P_4, C_4)$. $\square$

### 4.1.2 Finite distinguishing number

In this section we provide a characterisation of the minimal classes for the case of finite distinguishing number $k_{\mathcal{X}}$. It turns out that these minimal classes consist of $(\ell_{\mathcal{X}}, d_{\mathcal{X}})$-graphs, that is, the vertex set of each graph is partitioned into at most $\ell_{\mathcal{X}}$ bags and dense pairs are defined by a density graph $H$ (see Lemma 15). The condition of Theorem 21 is enforced by long paths (indeed, an infinite path in the infinite universal graph). Thus actually $d_{\mathcal{X}} \leq 2$ for the minimal classes $\mathcal{X}$.

Let $A$ be a finite alphabet. A *word* is a mapping $w : S \to A$, where $S = \{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$ or $S = \mathbb{N}$; $|S|$ is the *length* of $w$, denoted by $|w|$. We write $w_i$ for $w(i)$, and we often use the notation $w = w_1w_2w_3 \ldots w_n$ or $w = w_1w_2w_3 \ldots$. For $n \leq m$ and $w = w_1w_2 \ldots w_n$, $w' = w'_1w'_2 \ldots w'_m$ (or $w' = w'_1w'_2 \ldots$), we say that $w$ is a *factor* of $w'$ if there exists a non-negative integer $s$ such that $w_i = w'_{i+s}$ for $1 \leq i \leq n$; $w$ is an *initial segment* of $w'$ if we can take $s = 0$.

Let $H$ be an undirected graph with loops allowed and with vertex set $V(H) = A$, and let $w$ be a (finite or infinite) word over the alphabet $A$. For any increasing sequence

$u_1 < u_2 < \cdots < u_m$ of positive integers such that $u_m \leq |w|$, define $G_{w,H}(u_1, u_2, \ldots, u_m)$ to be the graph with vertex set $\{u_1, u_2, \ldots, u_m\}$ and an edge between $u_i$ and $u_j$ if and only if

- either $|u_i - u_j| = 1$ and $w_{u_i} w_{u_j} \notin E(H)$,
- or $|u_i - u_j| > 1$ and $w_{u_i} w_{u_j} \in E(H)$.

Let $G = G_{w,H}(u_1, u_2, \ldots, u_m)$ and define $V_a = \{u_i \in V(G) : w_{u_i} = a\}$ for any $a \in A$. Then $\pi = \pi_w(G) = \{V_a : a \in A\}$ is an $(|A|, 2)$-partition, and so $G$ is an $(|A|, 2)$-graph. Moreover, $\psi(G, \pi, H)$ is a linear forest whose paths are formed by the consecutive segments of integers within the set $\{u_1, u_2, \ldots, u_m\}$. This partition $\pi_w(G)$ is called the *letter partition* of $G$.

**Definition 9.** Let $H$ be an undirected graph with loops allowed and with vertex set $V(H) = A$, and let $w$ be an infinite word over the alphabet $A$. Define $\mathcal{P}(w, H)$ to be the hereditary class consisting of the graphs $G_{w,H}(u_1, u_2, \ldots, u_m)$ for all finite increasing sequences $u_1 < u_2 < \cdots < u_m$ of positive integers.

As we shall see later, all classes $\mathcal{P}(w, H)$ are above the Bell number. More importantly, all minimal classes above the Bell number have the form $\mathcal{P}(w, H)$ for some $w$ and $H$. Our goal here is firstly to describe sufficient conditions on the word $w$ under which $\mathcal{P}(w, H)$ is a minimal class above the Bell number; moreover, we aim to prove that any hereditary class above the Bell number with finite distinguishing number contains the class $\mathcal{P}(w, H)$ for some word $w$ and graph $H$. We start by showing that these classes indeed have finite distinguishing number.

**Lemma 16.** *For any word $w$ and graph $H$ with loops allowed, the class $\mathcal{X} = \mathcal{P}(w, H)$ has finite distinguishing number.*

*Proof.* Put $\ell = |H|$ and let $G$ be a graph in $\mathcal{X}$. Consider the letter partition $\pi = \pi_w(G) = \{V_a : a \in V(H)\}$ of $G$, which is an $(\ell, 2)$-partition. Choose an arbitrary set of vertices $X \subseteq V(G)$ and let $\{U_1, U_2, \ldots, U_k\}$ be the sets distinguished by $X$. If there are subsets $U_i$, $U_j$ and $V_a$ such that $|V_a \cap U_i| \geq 3$ and $|V_a \cap U_j| \geq 3$, then some vertex of $X$ has at least three neighbours and at least three non-neighbours in $V_a$, which contradicts the fact that $\pi$ is an $(\ell, 2)$-partition. Therefore, in the partition $\{V_a \cap U_i : a \in V(H),\ 1 \leq i \leq k\}$ we have at most $\ell$ sets of size at least 3. Note that every set $U_i$ of size at least $2\ell + 1$ must contain at least one such set. Hence the family $\{U_1, U_2, \dots, U_k\}$ contains at most $\ell$ sets of size at least $2\ell + 1$. Since the set $X$ was chosen arbitrarily, we conclude that $k_{\mathcal{X}} \leq 2\ell + 1$, as required. $\square$

The graphs $G_{w,H}(u_1, u_2, \dots, u_n)$ defined on a sequence of consecutive integers will play a special role in our considerations.

**Definition 10.** If $u_1, u_2, \dots, u_m$ is a sequence of consecutive integers (i.e., $u_{k+1} = u_k + 1$ for each $k$), we call the graph $G_{w,H}(u_1, u_2, \dots, u_m)$ an *$|H|$-factor*. Notice that each $|H|$-factor is an $(|H|, 2)$-graph; if its letter partition is a strong $(|H|, 2)$-partition, we call it a *strong $|H|$-factor*.

Note that if $G = G_{w,H}(u_1, u_2, \dots, u_m)$ is a strong $\ell$-factor, then its sparsification $\phi(G) = \psi(G, \pi_w(G), H)$ is an induced path with $m$ vertices.

**Proposition 25.** *If $w$ is an infinite word over a finite alphabet $A$ and $H$ is a graph on $A$, with loops allowed, then the class $\mathcal{P}(w, H)$ is above the Bell number.*

*Proof.* We may assume that every letter of $A$ appears in $w$ infinitely many times: otherwise we can remove a sufficiently long starting segment of $w$ to obtain a word $w'$ satisfying this condition, replace $H$ with its induced subgraph $H'$ on the alphabet $A'$ of $w'$, and obtain a subclass $\mathcal{P}(w', H')$ of $\mathcal{P}(w, H)$ with that property. For sufficiently large $k$, the $|A|$-factor $G_k = G_{w,H}(1, \dots, k)$ is a strong $|A|$-factor; thus $\phi(G_k)$ is an induced path of length $k - 1$. Having a finite distinguishing number by Lemma 16, the class $\mathcal{P}(w, H)$ is above the Bell number by Theorem 21. $\square$

**Definition 11.** An infinite word $w$ is called *almost periodic* if for any factor $f$ of $w$ there is a constant $k_f$ such that any factor of $w$ of length at least $k_f$ contains $f$ as a factor.

The notion of an almost periodic word plays a crucial role in our characterisation of minimal classes above the Bell number. First, let us show that if $w$ is almost periodic, then $\mathcal{P}(w, H)$ is a minimal property above the Bell number. To prove this, we need an auxiliary lemma.

**Lemma 17.** *Consider $G = G_{w,H}(u_1, \ldots, u_n)$. If $G$ is a strong $(\ell, d)$-graph and $\phi(G)$ contains a connected component $C$ such that $|C| \geq \left(2d'^2\ell^2|H|^2 + 1\right)(m-1) + 1$, where $d' = \max\{d, 2\}$, then $V(C)$ contains a sequence of $m$ consecutive integers.*

*Proof.* Let $\pi = \{U_1, U_2, \ldots, U_{\ell'}\}$ be a strong $(\ell, d)$-partition of $G$, so that $\ell' \leq \ell$ and $\phi(G) = \phi(G, \pi)$; let $\pi' = \{V_a : a \in V(H)\}$ be the letter partition of $G$, given by $V_a = \{u_j \in V(G) : w_{u_j} = a\}$. Put $k = |H|$. Note that $\pi'$ is an $(k, 2)$-partition, hence also an $(k, d')$-partition.

Let $E = E(\phi(G)) \setminus E(\psi(G, \pi', H))$ be the set of all the edges of $\phi(G)$ that are not edges of $\psi(G, \pi', H)$, that is, that do not join two consecutive integers. We will now upper-bound the number of such edges. Observe that $E$ consists of

- the edges between $U_i \cap V_a$ and $U_j \cap V_b$ where $(U_i, U_j)$ is $d'$-sparse and $(V_a, V_b)$ is $d'$-dense, and
- the non-edges between $U_i \cap V_a$ and $U_j \cap V_b$ where $(U_i, U_j)$ is $d'$-dense and $(V_a, V_b)$ is $d'$-sparse.

Consider the partition $\rho = \{U_i \cap V_a \colon 1 \leq i \leq \ell',\ a \in V(H)\}$ of $G$, which is an $(\ell' k, d')$-partition. Let $(U_i \cap V_a, U_j \cap V_b)$ be a pair of non-empty sets such that $(U_i, U_j)$ is $d'$-sparse but $(V_a, V_b)$ is $d'$-dense. Each such pair is both $d'$-sparse and $d'$-dense, and consequently we have $|U_i \cap V_a| \leq 2d'$ and $|U_j \cap V_b| \leq 2d'$. Moreover, there are at most $2d'^2$ edges between $U_i \cap V_a$ and $U_j \cap V_b$. Similarly, for any pair $(U_i \cap V_a, U_j \cap V_b)$ where $(U_i, U_j)$ is $d'$-dense but $(V_a, V_b)$ is $d'$-sparse, we can show that there are at most $2d'^2$ non-edges between $U_i \cap V_a$ and $U_j \cap V_b$. We conclude that $|E| \leq 2d'^2(\ell' k)^2$.

Any edge of $\phi(G)$ that is not in $E$ joins two consecutive integers. Hence any connected component $C$ of $\phi(G)$ consists of at most $|E|+1$ segments of consecutive integers connected by edges from $E$. If $C$ does not contain a sequence of $m$ consecutive integers, it consists of at most $|E|+1 \leq 2d'^2(\ell' k)^2+1$ segments of consecutive integers, each of length at most $m-1$; it can therefore contain at most $\left(2d'^2(\ell' k)^2 + 1\right)(m-1) \leq \left(2d'^2\ell^2|H|^2 + 1\right)(m-1)$ vertices. $\square$

**Theorem 26.** *If $w$ is an almost periodic infinite word and $H$ is a finite graph with loops allowed, then $\mathcal{P}(w, H)$ is a minimal hereditary property above the Bell number.*

*Proof.* The class $\mathcal{P} = \mathcal{P}(w, H)$ is above the Bell number by Proposition 25. Thus we only need to show that any proper hereditary subclass $\mathcal{X}$ of $\mathcal{P}$ is below the Bell number. Suppose $\mathcal{X} \subset \mathcal{P}$ and let $F \in \mathcal{P} \setminus \mathcal{X}$. By definition of $\mathcal{P}(w, H)$, the graph $F$ is of the form $G_{w,H}(u_1, \ldots, u_n)$ for some positive integers $u_1 < \cdots < u_n$. Let $w'$ be the finite word $w' = w_{u_1} w_{u_1+1} w_{u_1+2} \ldots w_{u_n-1} w_{u_n}$. As $w$ is almost periodic, there is an integer $m$ such that any factor of $w$ of length $m$ contains $w'$ as factor. Assume, for the sake of contradiction, that $\mathcal{X}$ is hereditary and above the Bell number. By Lemma 16, the distinguishing number of $\mathcal{P}$, and hence of $\mathcal{X}$, is finite, and therefore, by Lemma 15 and Theorem 21, there exists a strong $(\ell_{\mathcal{X}}, d_{\mathcal{X}})$-graph $G = G_{w,H}(u'_1, u'_2, \ldots, u'_{n'}) \in \mathcal{X}$ such that $\phi(G)$ has a connected component $C$ of order at least $\left(2d^2\ell^2|H|^2 + 1\right)(m-1) + 1$, where $\ell = \ell_{\mathcal{X}}$ and $d = \max\{d_{\mathcal{X}}, 2\}$. By Lemma 17, the vertices of $C$ contain a sequence of $m$ consecutive integers, i.e., $V(G) \supseteq V(C) \supseteq \{u', u'+1, \ldots, u'+m-1\}$. However, the word $w_{u'} w_{u'+1} \ldots w_{u'+m-1}$ contains $w'$; therefore $G$ contains $F$, a contradiction. $\square$

The existence of minimal classes does not necessarily imply that every class above the Bell number contains a minimal one. However, in our case this turns out to be true, as we proceed to show next. Moreover, this will also imply that the minimal classes described in Theorem 26 are the *only* minimal classes above the Bell number with $k_{\mathcal{X}}$ finite. To prove this, we first show in the next two lemmas that any class $\mathcal{X}$ above the Bell number with $k_{\mathcal{X}}$ finite contains arbitrarily large strong $\ell_{\mathcal{X}}$-factors.

**Lemma 18.** *Let $\mathcal{X}$ be a hereditary class with speed above the Bell number and with finite distinguishing number $k_{\mathcal{X}}$. Then for each $m$, the class $\mathcal{X}$ contains an $\ell_{\mathcal{X}}$-factor of order $m$.*

*Proof.* From Theorem 21 it follows that for each $m$ there is a graph $G_m \in \mathcal{X}$ which admits a strong $(\ell_{\mathcal{X}}, d_{\mathcal{X}})$-partition $\{V_1, V_2, \ldots, V_{\ell_m}\}$ with $\ell_m \leq \ell_{\mathcal{X}}$ such that the sparsification $\phi(G_m)$ has a connected component $C_m$ of order at least $(\ell_{\mathcal{X}} d_{\mathcal{X}})^m$. Fix an arbitrary vertex $v$ of $C_m$. As $C_m$ is an induced subgraph of $\phi(G_m)$, the maximum degree in $C_m$ is bounded by $d = \ell_{\mathcal{X}} d_{\mathcal{X}}$. Hence for any $k > 0$, in $C_m$ there are at most $d(d-1)^{k-1}$ vertices at distance $k$ from $v$; so there are at most $1 + \sum_{k=1}^{m-2} d(d-1)^{k-1} < d^m$ vertices at distance at most $m-2$ from $v$. As $C_m$ has order at least $d^m$, there exists a vertex $v'$ of distance $m-1$ from $v$. Therefore $C_m$ contains an induced path $v = v_1, v_2, \ldots, v_m = v'$ of length $m-1$.

Let $A = \{1, 2, \ldots, \ell_m\}$ and let $H$ be the graph with vertex set $A$ and edge between $i$ and $j$ if and only if $V_i$ is $d_{\mathcal{X}}$-dense with respect to $V_j$. Let $w_i \in A$ be such that $v_i \in V_{w_i}$ and define the word $w = w_1 w_2 \ldots w_m$. The induced subgraph $G_m[v_1, v_2, \ldots, v_m] \cong G_{w,H}(1, 2, \ldots, m)$ is an $\ell_{\mathcal{X}}$-factor of order $m$ contained in $\mathcal{X}$. $\square$

**Lemma 19.** *Let $\ell$ and $B$ be positive integers such that $B \geq 5 \times 2^{\ell+1}$. Then any $\ell$-factor $G_{w,H}(1, 2, \ldots, |w|)$ of order at least $B^\ell$ contains a strong $\ell$-factor $G_{w',H}(1, 2, \ldots, |w'|)$ of order at least $B$ such that $w'$ is a factor of $w$.*

*Proof.* We will prove by induction on $r \in \{1, 2, \ldots, \ell\}$ that any $\ell$-factor $G_{w,H}(1, 2, \ldots, B^r)$ on $B^r$ vertices with at most $r$ bags in the letter partition contains a strong $\ell$-factor on at least $B$ vertices. For $r = 1$ the statement holds because any $\ell$-factor with one bag in the letter partition of order $B \geq 5 \times 2^{\ell+1}$ is a strong $\ell$-factor. Suppose $1 < r \leq \ell$. Then either each letter of $w = w_1 w_2 \ldots w_{B^r}$ appears at least $B$ times, in which case we are done, or there is a letter $a = w_i$ which appears less than $B$ times in $w$. Consider the maximal factors of $w$ that do not contain the letter $a$. Because the number of occurrences of the letter $a$ in $w$ is less than $B$, there are at most $B$ such factors of $w$ and the sum of their orders is at least $B^r - B + 1$. By the pigeonhole principle, one of these factors has order at least $B^{r-1}$; call this factor $w''$. Now $w''$ contains at most $r - 1$ different letters; thus $G'' = G_{w'',H}(1, 2, \ldots, |w''|)$ is an $\ell$-factor of order at least $B^{r-1}$ for which the letter partition has at most $(r - 1)$ bags. By induction, $G''$ contains a strong $\ell$-factor $G_{w',H}(1, 2, \ldots, |w'|)$ of order at least $B$ such that $w'$ is a factor of $w''$ which is a factor of $w$. Hence $w'$ is a factor of $w$ and we are done. $\square$

**Theorem 27.** *Suppose $\mathcal{X}$ is a hereditary class above the Bell number with $k_{\mathcal{X}}$ finite. Then $\mathcal{X} \supseteq \mathcal{P}(w, H)$ for an infinite almost periodic word $w$ and a graph $H$ of order at most $\ell_{\mathcal{X}}$ with loops allowed.*

*Proof.* From Lemmas 18 and 19 it follows that each class $\mathcal{X}$ with speed above the Bell number with finite distinguishing number $k_{\mathcal{X}}$ contains an infinite set $\mathcal{S}$ of strong $\ell_{\mathcal{X}}$-factors of increasing order. For each $H$ on $\{1, 2, \ldots, \ell\}$ with $1 \leq \ell \leq \ell_{\mathcal{X}}$, let $\mathcal{S}_H = \{G_{w,H}(1, \ldots, m) \in \mathcal{S}\}$ be the set of all $\ell_{\mathcal{X}}$-factors in $\mathcal{S}$ whose adjacencies are defined using the density graph $H$. Then for some (at least one) fixed graph $H_0$ the set $\mathcal{S}_{H_0}$ is infinite.

Hence also $L = \{w : G_{w,H_0}(1, \dots, m) \in \mathcal{X}\}$ is an infinite language. As $\mathcal{X}$ is a hereditary class, the language $L$ is closed under taking word factors (it is a *factorial language*).

It is not hard to see that any infinite factorial language contains an inclusion-minimal infinite factorial language. So let $L' \subseteq L$ be a minimal infinite factorial language. It follows from the minimality that $L'$ is well quasi-ordered by the factor relation, because otherwise removing one word from any infinite antichain and taking all factors of the remaining words would generate an infinite factorial language strictly contained in $L'$. Thus there exists an infinite chain $w^{(1)}, w^{(2)}, \dots$ of words in $L'$ such that for any $i < j$, the word $w^{(i)}$ is a factor of $w^{(j)}$. More precisely, for each $i$ there is a non-negative integer $s_i$ such that $w_k^{(i)} = w_{k+s_i}^{(i+1)}$. Let $g(i,k) = k + \sum_{j=1}^{i-1} s_j$. Now we can define an infinite word $w$ by putting $w_k = w_{g(i,k)}^{(i)}$ for the least value of $i$ for which the right-hand side is defined. (Without loss of generality we get that $w$ is indeed an infinite word; otherwise we would need to take the reversals of all the words $w^{(i)}$.)

Observe that any factor of $w$ is a factor of some $w^{(i)}$ and hence in the language $L'$. If $w$ is not almost periodic, then there exists a factor $f$ of $w$ such that there are arbitrarily long factors $f'$ of $w$ not containing $f$. These factors $f'$ generate an infinite factorial language $L'' \subset L'$ which does not contain $f \in L'$. This contradicts the minimality of $L'$ and proves that $w$ is almost periodic.

Because any factor of $w$ is in $L$, any $G_{w,H_0}(u_1, \dots, u_m)$ is an induced subgraph of some $\ell_{\mathcal{X}}$-factor in $\mathcal{X}$. Therefore $\mathcal{P}(w, H_0) \subseteq \mathcal{X}$. $\square$

Combining Theorems 26 and 27 we derive the main result of this section.

**Corollary 28.** *Let $\mathcal{X}$ be a class of graphs with $k_{\mathcal{X}} < \infty$. Then $\mathcal{X}$ is a minimal hereditary class above the Bell number if and only if there exists a finite graph $H$ with loops allowed and an infinite almost periodic word $w$ over $V(H)$ such that $\mathcal{X} = \mathcal{P}(w, H)$.* $\square$

Lastly, note that – similarly to the case of infinite distinguishing number – each of the minimal classes $\mathcal{P}(w, H)$ has an infinite universal graph: $G_{w,H}(1, 2, 3, \dots)$.

# Chapter 5

# Hereditary properties of unbounded clique-width

Clique-width is a graph parameter which is important in theoretical computer science, because many algorithmic problem that are generally NP-hard become polynomial-time solvable when restricted to graphs of bounded clique-width [24]. This is a relatively new notion and it generalizes another important graph parameter, tree-width, studied in the literature for decades. Clique-width is stronger than tree-width in the sense that graphs of bounded tree-width have bounded clique-width, but not necessarily vice versa. For instance, both parameters are bounded for trees, while for complete graphs only clique-width is bounded.

When we study classes of graphs of bounded tree-width, we may assume without loss of generality that together with every graph $G$ our class contains all minors of $G$, as the tree-width of a minor can never be larger than the tree-width of the graph itself. In other words, when we try to identify classes of graphs of bounded tree-width, we may restrict ourselves to minor-closed graph classes. However, when we deal with clique-width this restriction is not justified, as the clique-width of a minor of $G$ can be much larger than the clique-width of $G$ [25]. On the other hand, the clique-width of $G$ is never smaller than the clique-width of any of its induced subgraphs [25]. This allows us to be restricted to hereditary classes, i.e. those that are closed under taking induced subgraphs.

One of the most remarkable outcomes of the graph minor project of Robertson and

Seymour is the proof of Wagner's conjecture stating that the minor relation is a well-quasi-order [61]. This implies, in particular, that in the world of minor-closed graph classes there exist minimal classes of unbounded tree-width and the number of such classes is finite. In fact, there is just one such class (the planar graphs), which was shown even before the proof of Wagner's conjecture [60].

In the world of hereditary classes the situation is more complicated, because the induced subgraph relation is not a well-quasi-order. It contains infinite antichains, the set of cycles for example. Hence, there may exist infinite strictly decreasing sequences of graph classes with no minimal one. In other words, even the existence of minimal hereditary classes of unbounded clique-width is not an obvious fact. This fact was recently confirmed in [46]. However, whether the number of such classes is finite or infinite remained an open question. We will settle this question by showing that the family of minimal hereditary classes of unbounded clique-width is infinite. Moreover, we will prove that the same is true with respect to *linear* clique-width.

## 5.1 Preliminaries

Let $G$ be a graph and $U \subseteq V(G)$ a subset of its vertices. Two vertices of $U$ will be called $U$-similar if they have the same neighborhood outside $U$. Clearly, $U$-similarity is an equivalence relation. The number of equivalence classes of $U$ will be denoted $\mu(U)$.

The notion of clique-width of a graph was introduced in [23]. The clique-width of a graph $G$ is denoted $\mathrm{cwd}(G)$ and is defined as the minimum number of labels needed to construct $G$ by means of the following four graph operations:

- creation of a new vertex $v$ with label $i$ (denoted $i(v)$),
- disjoint union of two labeled graphs $G$ and $H$ (denoted $G \oplus H$),
- connecting vertices with specified labels $i$ and $j$, with $i \neq j$ (denoted $\eta_{i,j}$) and
- renaming label $i$ to label $j$ (denoted $\rho_{i \to j}$).

Every graph can be defined by an algebraic expression using the four operations above. This expression is called a $k$-expression if it uses $k$ different labels. For instance, the cycle $C_5$ on vertices $a, b, c, d, e$ (listed along the cycle) can be defined by the following 4-expression:

$$\eta_{4,1}(\eta_{4,3}(4(e) \oplus \rho_{4\to3}(\rho_{3\to2}(\eta_{4,3}(4(d) \oplus \eta_{3,2}(3(c) \oplus \eta_{2,1}(2(b) \oplus 1(a)))))))).$$

Alternatively, any algebraic expression defining $G$ can be represented as a rooted tree, whose leaves correspond to the operations of vertex creation, the internal nodes correspond to the $\oplus$-operations, and the root is associated with $G$. The operations $\eta$ and $\rho$ are assigned to the respective edges of the tree. Figure 5.1 shows the tree representing the above expression defining a $C_5$.

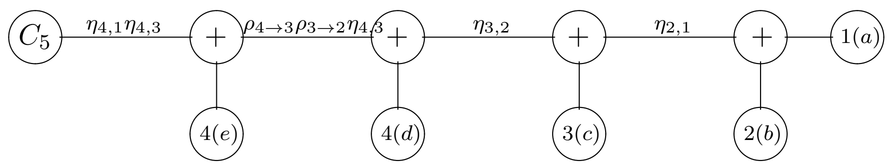



Figure 5.1: The tree representing the expression defining a $C_5$

Let us observe that the tree in Figure 5.1 has a special form known as a *caterpillar tree* (i.e. a tree that becomes a path after the removal of vertices of degree 1). The minimum number of labels needed to construct a graph $G$ by means of caterpillar trees is called the *linear clique-width* of $G$ and is denoted $\mathrm{lcwd}(G)$. Clearly, $\mathrm{lcwd}(G) \geq \mathrm{cwd}(G)$ and there are classes of graphs for which the difference between clique-width and linear clique-width can be arbitrarily large (see e.g. [18]).

A notion which is closely related to clique-width is that of *rank-width* (denoted $\mathrm{rwd}(G)$), which was introduced by Oum and Seymour in [54]. They showed that rank-width and clique-width are related to each other by proving that if the clique-width of a graph $G$ is $k$, then

$$\mathrm{rwd}(G) \leq k \leq 2^{\mathrm{rwd}(G)+1} - 1.$$

Therefore a class of graphs has unbounded clique-width if and only if it also has unbounded rank-width.

For a graph $G$ and a vertex $v$, the *local complementation* at $v$ is the operation that replaces the subgraph induced by the neighbourhood of $v$ with its complement. A graph

$H$ is a *vertex-minor* of $G$ if $H$ can be obtained from $G$ by a sequence of local complementations and vertex deletions. In [53] it was proved that if $H$ is a vertex-minor of $G$, then the rank-width of $H$ is at most the rank-width of $G$.

Finally, we introduce some language-theoretics terminology and notation. Given a word $\alpha$, we denote by $\alpha(k)$ the $k$-th letter of $\alpha$ and by $\alpha^k$ the concatenation of $k$ copies of $\alpha$. A *factor* of $\alpha$ is a contiguous subword of $\alpha$, i.e. a subword $\alpha(i)\alpha(i+1)\dots\alpha(i+k)$ for some $i$ and $k$. An infinite word $\alpha$ is *periodic* if there is a finite number $k$ such that $\alpha(i) = \alpha(i+k)$ for all $i$.

## 5.2 Minimal classes of graphs of unbounded clique-width

In this section, we describe an infinite family of graph classes of unbounded clique-width (Subsections 5.2.1 and 5.2.2). The fact that each of them is a minimal hereditary class of unbounded clique-width will be proved in Subsection 5.2.3.

Each class in our family is defined through a universal element, i.e. a infinite graph that contains all graphs from the class as induced subgraphs. All constructions start from the graph

- $\mathcal{P}$ the disjoint union of infinitely many of infinite paths. We assume that the vertices of each path are labelled by natural numbers listed along the path. We denote the set of vertices of $\mathcal{P}$ with label $j$ by $V_j$ and refer to it as the $j$-th column of the graph. Also, the set of vertices of the $i$-th path will be called the $i$-th row of $\mathcal{P}$. The vertex of $\mathcal{P}$ in the $i$-th row and $j$-th column will be denoted $v_{i,j}$.

Let $\alpha = \alpha_1\alpha_2\dots$ be an infinite binary word, i.e. an infinite word such that $\alpha_j \in \{0,1\}$ for each natural $j$. The graph $\mathcal{P}^\alpha$ is obtained from $\mathcal{P}$ by complementing the edges between two consecutive columns $V_j$ and $V_{j+1}$ if and only if $\alpha_j = 1$. In other words, we apply bipartite complementation to the bipartite graph induced by $V_j$ and $V_{j+1}$. In particular, if $\alpha$ does not contain 1s, then $\mathcal{P}^\alpha = \mathcal{P}$.

Finally, by $\mathcal{G}_\alpha$ we denote the class of all finite induced subgraphs of $\mathcal{P}^\alpha$. By definition, $\mathcal{G}_\alpha$ is a hereditary class. In what follows we show that $\mathcal{G}_\alpha$ is a minimal hereditary class of unbounded clique-width for infinitely many values of $\alpha$.

### 5.2.1 The basic class

Our first example constitutes the basis for infinitely many other constructions. It deals with the class $\mathcal{G}_{1^\infty}$, where $1^\infty$ stands for the infinite word of all 1s. Let us denote by

$F_{n,n}$ the subgraph of $\mathcal{P}^{1^\infty}$ induced by $n$ consecutive columns and any $n$ rows.

In order to show that $\mathcal{G}_{1^\infty}$ is a class of unbounded clique-width, we will prove the following lemma.

**Lemma 20.** *The clique-width of $F_{n,n}$ is at least $\lfloor n/2 \rfloor$.*

*Proof.* Let $\mathrm{cwd}(F_{n,n}) = t$. Denote by $\tau$ a $t$-expression defining $F_{n,n}$ and by $tree(\tau)$ the rooted tree representing $\tau$. The subtree of $tree(\tau)$ rooted at a node $x$ will be denoted $tree(x, \tau)$. This subtree corresponds to a subgraph of $F_{n,n}$, which will be denoted $F(x)$. The label of a vertex $v$ of the graph $F_{n,n}$ at the node $x$ is defined as the label that $v$ has immediately prior to applying the operation $x$.

Let $a$ be a lowest $\oplus$-node in $tree(\tau)$ such that $F(a)$ contains a full column of $F_{n,n}$. Denote the children of $a$ in $tree(\tau)$ by $b$ and $c$. Let us colour all vertices in $F(b)$ blue and all vertices in $F(c)$ red, and the remaining vertices of $F_{n,n}$ yellow. Note that by the choice of $a$ the graph $F_{n,n}$ contains a non-yellow column (i.e. a column each vertex of which is non-yellow), but none of its columns are entirely red or blue. We denote a non-yellow column of $F_{n,n}$ by $r$. Without loss of generality we assume that $r \leq \lceil n/2 \rceil$ and that the column $r$ contains at least $n/2$ red vertices, since otherwise we could consider the columns in reverse order and swap the colours red and blue.

Observe that edges of $F_{n,n}$ between different coloured vertices are not present in $F(a)$. Therefore, if a non-red vertex distinguishes two red vertices $u$ and $v$, then $u$ and $v$ must have different labels at the node $a$. We will use this fact to show that $F(a)$ contains a set $U$ of at least $\lfloor n/2 \rfloor$ vertices with pairwise different labels at the node $a$. Such a set can be constructed by the following procedure.

1. Set $j = r,\ U = \emptyset$ and $I = \{i : v_{i,r} \text{ is red}\}$.
2. Set $K = \{i \in I : v_{i,j+1} \text{ is non-red}\}$.
3. If $K \neq \emptyset$, add the vertices $\{v_{k,j} : k \in K\}$ to $U$. Remove members of $K$ from $I$.

4. If $I = \emptyset$, terminate the procedure.

5. Increase $j$ by 1. If $j = n$, choose an arbitrary $i \in I$, put $U = \{v_{i,m} : r \leq m \leq n-1\}$ and terminate the procedure.

6. Go back to Step 2.

It is not difficult to see that this procedure must terminate. To complete the proof, it suffices to show that whenever the procedure terminates, the size of $U$ is at least $\lfloor n/2 \rfloor$ and the vertices in $U$ have pairwise different labels at the node $a$

First, suppose that the procedure terminates in Step 5. Then $U$ is a subset of red vertices from at least $\lfloor n/2 \rfloor$ consecutive columns of row $i$. Consider two vertices $v_{i,l}$ , $v_{i,m} \in U$ with $l < m$. According to the above procedure, $v_{i,m+1,}$ is red. Since $F_{n,n}$ does not contain an entirely red column, there must exist a non-red vertex $w$ in the column $m+1$. According to the structure of $F_{n,n}$, vertex $w$ is adjacent to $v_{i,m}$ and non-adjacent to $v_{i,l}$. We conclude that $v_{i,l}$ and $v_{i,m}$ have different labels. Since $v_{i,l}$ and $v_{i,m}$ have been chosen arbitrarily, the vertices of $U$ have pairwise different labels.

Now suppose that the procedure terminates in Step 4. By analyzing Steps 2 and 3, it is easy to deduce that $U$ is a subset of red vertices of size at least $\lfloor n/2 \rfloor$. Suppose that $v_{i,l}$ and $v_{k,m}$ are two vertices in $U$ with $l \leq m$. The procedure certainly guarantees that $i \neq k$ and that both $v_{i,l+1}$ and $v_{k,m+1}$ are non-red. If $m \in \{l, l+2\}$, then it is clear that $v_{i,l+1}$ distinguishes vertices $v_{i,l}$ and $v_{k,m}$, and therefore these vertices have different labels. If $m \notin \{l, l+2\}$, we may consider vertex $v_{k,m-1}$ which must be red. Since $F_{n,n}$ does not contain an entirely red column, the vertex $v_{k,m}$ must have a non-red neighbor $w$ in the column $m-1$. But $w$ is not a neighbour of $v_{i,l}$, trivially. We conclude that $v_{i,l}$ and $v_{k,m}$ have different labels, and therefore, the vertices of $U$ have pairwise different labels. This shows that the clique-width of the graph $F_{n,n}$ is at least $\lfloor n/2 \rfloor$. $\square$

### 5.2.2 Other classes

In this section, we discover more hereditary classes of graphs of unbounded clique-width by showing that for all $n \in \mathbb{N}$ such classes have graphs containing $F_{n,n}$ as a vertex-minor.

**Lemma 21.** *Let $\alpha$ be an infinite binary word containing infinitely many* 1*s. Then the clique-width of graphs in the class $\mathcal{G}_\alpha$ is unbounded.*

*Proof.* First fix an even number $n$. Let $\beta$ be a factor of $\alpha$ containing precisely $n$ occurrences of 1, starting and ending with 1. We denote the length of $\beta$ by $\ell$ and consider the subgraph $G_n$ of $P^\alpha$ induced by $\ell + 1$ consecutive columns corresponding to $\beta$ and by any $n$ rows. We will now show that $G_n$ contains the graph $F_{n,n}$ defined in Theorem 20 as a vertex-minor.

If $\beta$ contains 00 as a factor, then there are three columns $V_i, V_{i+1}, V_{i+2}$ such that each of $V_i \cup V_{i+1}$ and $V_{i+1} \cup V_{i+2}$ induces a 1-regular graph. We apply a local complementation to each vertex of $G_n$ in column $V_{i+1}$ and then delete the vertices of $V_{i+1}$ from $G_n$. Under this operation, our graph transforms into a new graph where column $V_{i+1}$ is absent, while columns $V_i$ and $V_{i+2}$ induce a 1-regular graph. In terms of words, this operation is equivalent to removing one 0 from the factor 00. Applying this transformation repeatedly, we can reduce $G_n$ to an instance corresponding to a word $\beta$ with no two consecutive 0s.

Now assume $\beta$ contains 01 as a factor, and let $V_j, V_{j+1}$ and $V_{j+2}$ be three consecutive columns such that $V_j \cup V_{j+1}$ induces a 1-regular graph, while the edges between $V_{j+1}$ and $V_{j+2}$ form the bipartite complement of a 1-regular graph. We apply a local complementation to each vertex of $V_{j+1}$ in turn and then delete the vertices of $V_{j+1}$ from $G_n$. It is not difficult to see that in the transformed graph the edges between $V_j$ and $V_{j+2}$ form the bipartite complement of a matching. Looking at the vertices in $V_{j+2}$ we see that for any two vertices $x$ and $y$ in this column, when a local complementation is applied at $z \in V_{j+1}$ the adjacency between $x$ and $y$ is complemented if and only if both $x$ and $y$ are adjacent to $z$. Since $|V_{j+2}| = n$ is even, we conclude that after $n$ applications of local complementation $V_{j+2}$ remains an independent set. In terms of words, this operation is equivalent to removing 0 from the factor 01. Applying this transformation repeatedly, we can reduce $G_n$ to an instance corresponding to a word $\beta$ which is free of 0s.

The above discussion shows that $G_n$ can be transformed by a sequence of local complementations and vertex deletions into $F_{n,n}$. Therefore, $G_n$ contains the graph $F_{n,n}$ as a vertex-minor. Since $n$ can be arbitrarily large, we conclude that the rank-width, and hence the clique-width, of graphs in $\mathcal{G}_\alpha$ is unbounded. $\square$

### 5.2.3 Minimality of classes $\mathcal{G}_\alpha$ with a periodic $\alpha$

In the previous section, we proved that any class $\mathcal{G}_\alpha$ with infinitely many 1s in $\alpha$ has unbounded clique-width. In the present section, we will show that if $\alpha$ is periodic, then $\mathcal{G}_\alpha$ is a minimal hereditary class of graphs of unbounded clique-width, provided that $\alpha$ contains at least one 1. In other words, we will show that in any proper hereditary subclass of $\mathcal{G}_\alpha$ the clique-width is bounded. Moreover, we will show that proper hereditary subclasses of $\mathcal{G}_\alpha$ have bounded *linear* clique-width. To this end, we first prove a technical lemma, which strengthens a similar result given in [46] from clique-width to linear clique-width. Let us repeat that by $\mu(U)$ we denote the number of similarity classes with respect to an equivalence relation defined in Section 5.1.

**Lemma 22.** *Let $m \geq 2$ and $\ell$ be positive integers. Suppose that the vertex set of $G$ can be partitioned into sets $U_1, U_2, \ldots$ where for each $i$,*

(1) $\mathrm{lcwd}(G[U_i]) \leq m$*,*

(2) $\mu(U_i) \leq \ell$ *and* $\mu(U_1 \cup \cdots \cup U_i) \leq \ell$*.*

*Then* $\mathrm{lcwd}(G) \leq \ell(m+1)$*.*

*Proof.* If $G[U_1]$ can be constructed with at most $m$ labels and $\mu(U_1) \leq \ell$, then $G[U_1]$ can be constructed with at most $m\ell$ different labels in such a way that in the process of construction any two vertices in different equivalence classes of $U_1$ have different labels, and by the end of the process any two vertices in the same equivalence class of $U_1$ have the same label. In other words, we build $G[U_1]$ with at most $m\ell$ labels and finish the process with at most $\ell$ labels corresponding to the equivalence classes of $U_1$.

Now assume we have constructed the graph $G_i = G[U_1 \cup \cdots \cup U_i]$ using $m\ell$ different labels making sure that the construction finishes with a set $A$ of at most $\ell$ different labels corresponding to the equivalence classes of $U_1 \cup \cdots \cup U_i$. By assumption, it is possible to construct $G[U_{i+1}]$ using a set $B$ of at most $m\ell$ different labels such that we finish the process with at most $\ell$ labels corresponding to the equivalence classes of $U_{i+1}$. We choose labels so that $A$ and $B$ are disjoint. As we construct $G[U_{i+1}]$ join each vertex to its neighbours in $G_i$ to build the graph $G_{i+1} = G[U_1 \cup \cdots \cup U_i \cup U_{i+1}]$. Notice that any two vertices in the same equivalence class of $U_1 \cup \cdots \cup U_i$ or $U_{i+1}$ belong to the same equivalence class of $U_1 \cup \cdots \cup U_i \cup U_{i+1}$. Therefore, the construction of $G_{i+1}$ can be completed with a set of at most $\ell$ different labels corresponding to the equivalence classes of the graph. The conclusion now follows by induction. $\square$

Now let $\alpha$ be an infinite binary periodic word of period $p$ with at least one 1. For constants $k$ and $t$, let us denote by

$H_{k,t}$ the subgraph of $\mathcal{P}^\alpha$ induced by the first $k$ rows and any $t$ columns.

It is not difficult to see that

**Lemma 23.** *A graph with $n$ vertices in $\mathcal{G}_\alpha$ is an induced subgraph of $H_{k,t}$ for any $k \geq n$ and any $t \geq n(p+1)$.*

Now, with the help of Lemma 22 we derive the following conclusion.

**Lemma 24.** *The linear clique-width of $H_{k,t}$ is at most $4t$.*

*Proof.* Denote by $U_i$ the $i$-th row of $H_{k,t}$. Since each row induces a path forest (i.e. a disjoint union of paths), it is clear that $\mathrm{lcwd}(G[U_i]) \leq 3$ for every $i$. Trivially, $\mu(U_i) \leq t$, since $|U_i| = t$. Also, denoting $W_i := U_1 \cup \ldots \cup U_i$, it is not difficult to see that $\mu(W_i) \leq t$ for every $i$, since the vertices of the same column are $W_i$-similar. Now the conclusion follows from Lemma 22. $\square$

Next we use Lemmas 22, 23 and 24 to prove the following result.

**Lemma 25.** *For any fixed $k \geq 1$, the linear clique-width of $H_{k,k}$-free graphs in the class $\mathcal{G}_\alpha$ is bounded by a function of $k$.*

*Proof.* Let $G$ be an $H_{k,k}$-free graph in $\mathcal{G}_\alpha$. By Lemma 23, the graph $G$ is an induced subgraph of $H_{n,n}$ for some $n$. For convenience, assume that $n$ is a multiple of $k$, say $n = tk$. We fix an arbitrary embedding of $G$ into $H_{n,n}$ and call the vertices of $H_{n,n}$ that induce $G$ *black*. The remaining vertices of $H_{n,n}$ will be called *white*.

For $1 \leq i \leq t$, let us denote by $W_i$ the subgraph of $H_{n,n}$ induced by the $k$ consecutive columns $(i-1)k+1, (i-1)k+2, \ldots, ik$. We partition the vertices of $G$ into subsets $U_1, U_2, \ldots, U_t$ according to the following procedure:

1. For $1 \leq j \leq t$, set $U_j = \emptyset$. Add every black vertex of $W_1$ to $U_1$. Set $i = 2$.

2. For $j = 1, \ldots, n$,

   - if row $j$ of $W_i$ is entirely black, then add the first vertex of this row to $U_{i-1}$ and the remaining vertices of the row to $U_i$.

   - otherwise, add the (black) vertices of row $j$ preceding the first white vertex to $U_{i-1}$ and add the remaining black vertices of the row to $U_i$.

3. Increase $i$ by 1. If $i = t + 1$, terminate the procedure.

4. Go back to Step 2.

Let us show that the partition $U_1, U_2, \ldots, U_t$ given by the procedure satisfies the assumptions of Lemma 22 with $m$ and $\ell$ depending only on $k$.

The procedure clearly assures that each $G[U_i]$ is an induced subgraph of $G[V(W_i) \cup V(W_{i+1})]$. By Lemma 24, we have $\mathrm{lcwd}(G[V(W_i) \cup V(W_{i+1})]) = \mathrm{lcwd}(F_{n,2k}) \leq 8k$. Since the linear clique-width of an induced subgraph cannot exceed the linear clique-width of the parent graph, we conclude that $\mathrm{lcwd}(G[U_j]) \leq 8k$, which shows condition (1) of Lemma 22.

To show condition (2) of Lemma 22, let us call a vertex $v_{j,m}$ of $U_i$ boundary if either $v_{j,m-1}$ belongs to $U_{i-1}$ or $v_{j,m+1}$ belongs to $U_{i+1}$ (or both). It is not difficult to see that a vertex of $U_i$ is boundary if it belongs either to the second column of an entirely black row of $W_i$ or to the first column of an entirely black row of $W_{i+1}$. Since the graph $G$ is $H_{k,k}$-free, the number of rows of $W_i$ which are entirely black is at most $k - 1$. Therefore, the boundary vertices of $U_i$ introduce at most $2(k - 1)$ equivalence classes in $U_i$.

Now consider two non-boundary vertices of $U_i$ from the same column. It is not difficult to see that these vertices have the same neighborhood outside of $U_i$. Therefore, the non-boundary vertices of the same column of $U_i$ are $U_i$-similar and hence the non-boundary vertices give rise to at most $2k$ equivalence classes in $U_i$. Thus, $\mu(U_i) \leq 4k - 2$ for all $i$.

Similar argument show that $\mu(U_1 \cup \ldots \cup U_i) \leq 3k - 1 \leq 4k - 2$ for all $i$. Therefore, by Lemma 22, we conclude that $\mathrm{lcwd}(G) \leq (4k - 2)(8k + 1)$, which completes the proof. $\square$

**Theorem 29.** *Let $\alpha$ be an infinite binary periodic word containing at least one* 1*. Then the class $\mathcal{G}_\alpha$ is a minimal hereditary class of graphs of unbounded clique-width and linear* *clique-width.*

*Proof.* By Lemma 21, the clique-with of graphs in $\mathcal{G}_\alpha$ is unbounded. Therefore, linear clique-width is unbounded too. To prove the minimality, consider a proper hereditary subclass $X$ of $\mathcal{G}_\alpha$ and let $G \in \mathcal{G}_\alpha - X$. By Lemma 23, $G$ is an induced subgraph of $H_{k,k}$ for some finite $k$. Therefore, each graph in $X$ is $H_{k,k}$-free. Observe that the value of $k$ is the same for all graphs in $X$. It depends only on $G$ and the period of $\alpha$. Therefore, by Lemma 25, the linear clique-width (and hence clique-width) of graphs in $X$ is bounded by a constant. $\square$

## 5.3 More classes of graphs of unbounded clique-width

In this section, we extend the alphabet from $\{0, 1\}$ to $\{0, 1, 2\}$ in order to construct more classes of graphs of unbounded clique-width. Let $\alpha$ be an infinite word over the alphabet $\{0, 1, 2\}$. We remind the reader that the letter 1 stands for the operation of bipartite complementation between two consecutive columns $V_j$ and $V_{j+1}$ of the graph $\mathcal{P}$, i.e. if $\alpha_j = 1$, then two vertices $v_{i,j} \in V_j$ and $v_{k,j+1} \in V_{j+1}$ are adjacent in $\mathcal{P}^\alpha$ if and only if they are not adjacent in $\mathcal{P}$.

The new letter 2 will represent the operation of "forward" complementation, i.e. if $\alpha_j = 2$, then two vertices $v_{i,j} \in V_j$ and $v_{k,j+1} \in V_{j+1}$ with $i < k$ are adjacent in $\mathcal{P}^\alpha$ if and only if they are not adjacent in $\mathcal{P}$. In other words, this operation adds edges between $v_{i,j}$ and $v_{k,j+1}$ with $i < k$. The bipartite graph induced by two consecutive columns corresponding to the letter 2 is known in the literature as a *chain graph*.

Of special interest for the topic of this paper is the word $2^\infty = 222\ldots$. The class $\mathcal{G}_{2^\infty}$ is also known as the class of bipartite permutation graphs and this is one of the first two minimal classes of graphs of unbounded clique-width discovered in the literature [46]. We will denote by

$X_{n,n}$ the subgraph of $\mathcal{P}^{2^\infty}$ induced by $n$ consecutive columns and and any $n$ rows. Figure 5.2 represents an example of the graph $X_{n,n}$ with $n = 6$.

The unboundedness of clique-width in the class $\mathcal{G}_{2^\infty}$ follows from the following result proved in [16].

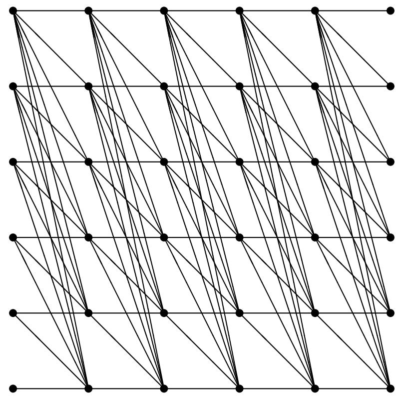


Figure 5.2: The graph $X_{6,6}$

**Lemma 26.** *The clique-width of $X_{n,n}$ is at least $n/6$.*

In what follows, we will prove that every class $\mathcal{G}_\alpha$ with infinitely many 2s in $\alpha$ has unbounded clique-width by showing that graphs in this class contain $X_{n,n}$ as a vertex minor for arbitrarily large values of $n$. We start with the case when the letter 1 appears finitely many times in $\alpha$.

**Lemma 27.** *Let $\alpha$ be an infinite word over the alphabet $\{0, 1, 2\}$, containing the letter 2 infinitely many times and the letter 1 finitely many times. Then the class $\mathcal{G}_\alpha$ has unbounded clique-width.*

*Proof.* First fix a constant $n$. Let $\beta$ be a factor of $\alpha$ containing precisely $n$ instances of the letter 2, starting and ending with the letter 2 and containing no instances of the letter 1 (since letter 2 appears infinitely many times and letter 1 finitely many times in $\alpha$, we can always find such a factor). We denote the length of $\beta$ by $\ell$ and consider the subgraph $G_n$ of $P^\alpha$ induced by $\ell + 1$ consecutive columns corresponding to $\beta$ and by any $n2^{n-1}$ rows. We will now show that $G_n$ contains the graph $X_{n,n}$ as a vertex-minor.

Using arguments identical to those in Theorem 21, we can show that any instance of 00 can be replaced by 0 with the help of local complementations and vertex deletions.

Now each instance of 0 is surrounded by 2s in $\beta$. Consider any factor 02 of $\beta$ and let $V_j, V_{j+1}, V_{j+2}$ be three columns such that $V_j \cup V_{j+1}$ induces a 1-regular graph and $V_{j+1} \cup V_{j+1}$ induces a chain graph. If we apply a local complementation to each vertex of $V_{j+1}$ in turn, it is easy to see that the edges between $V_j$ and $V_{j+2}$ form a chain graph.

Looking at the vertices in the column $V_{j+2}$ we see that for any two vertices $x$ and $y$, when a local complementation is applied at $z \in V_{j+1}$ the edge between $x$ and $y$ is complemented if and only if both $x$ and $y$ are adjacent to $z$. Therefore, $x$ and $y$ are adjacent if and only if $\min\{|N(x) \cap V_{j+1}|, |N(y) \cap V_{j+1}|\}$ is odd. Hence the vertices of $V_{j+2}$ in the even rows induce an independent set. So, applying a local complementation to each vertex of $V_{j+1}$ in turn and then deleting column $V_{j+1}$ together with the odd rows allows us to reduce the factor 02 to 2. This transformation also reduces the number of rows two times. Since the factor 02 can appear at most $n-1$ times, in at most $n-1$ transformations we reduced $G_n$ to a graph containing $X_{n,n}$. Therefore, $G_n$ contains $X_{n,n}$ as a vertex minor.

Since $n$ can be arbitrarily large, we conclude with the help of Lemma 26 that graphs in $\mathcal{G}_\alpha$ can have arbitrarily large clique-width. □

To extend the last lemma to a more general result, we again refer to [53], which introduces another useful transformation, called *pivoting*. For a graph $G$ and an edge $xy$, the graph obtained by pivoting $xy$ is defined to be the graph obtained by applying local complementation at $x$, then at $y$ and then at $x$ again. They then show that in the case of bipartite graphs pivoting $xy$ is identical to complementing the edges between $N(x) \setminus \{y\}$ and $N(y) \setminus \{x\}$. We will use this transformation to prove the following result.

**Lemma 28.** *Let $\alpha$ be an infinite word over the alphabet $\{0, 1, 2\}$, containing the letter 2 infinitely many times. Then the class $\mathcal{G}_\alpha$ has unbounded clique-width.*

*Proof.* First, fix a constant $n$. Let $\beta$ be a factor of $\alpha$ containing precisely $n$ instances of the letter 2, starting and ending with the letter 2. Let $G_n$ be the subgraph of $\mathcal{P}^\alpha$ induced by the columns corresponding to $\beta$ and by any $n2^n + n^2$ rows. To prove the lemma, it is enough to show that $G_n$ contains either $F_{n,n}$ or $X_{n,n}$ as a vertex minor.

Consider any two consecutive appearances of 2 in $\beta$ and denote the word between them by $\gamma$. In other words, $\gamma$ is a (possibly empty) word in the alphabet $\{0, 1\}$. If $\gamma$ contains at least $n$ instances of 1, then by Lemma 21 $G_n$ contains $F_{n,n}$ as a vertex minor. Therefore, we assume that the number of 1s in $\gamma$ is at most $n-1$. If $\gamma$ contains no instance of 1, then we apply the idea of Lemma 27 to reduce it to the empty word. If $\gamma$ contains at least one instance of 1, we apply the idea of Lemma 21 to eliminate all 0s from it.

Suppose that after this transformation $\gamma$ contains at least two 1s, i.e. $\beta$ contains 211 as a factor. Let $V_j, V_{j+1}, V_{j+2}$ and $V_{j+3}$ be the four columns such that $V_{j+1} \cup V_{j+2}$ and $V_{j+2} \cup V_{j+3}$ induce bipartite complements of 1-regular graph and $V_j \cup V_{j+1}$ induces a chain graph. Let $x$ be the vertex in the first row of column $V_{j+1}$ and $y$ be the vertex in the last row of column $V_{j+2}$. It is not difficult to see that if we pivot the edge $xy$ and delete the first and the last row, then the graphs induced by $V_{j+1} \cup V_{j+2}$ and by $V_{j+2} \cup V_{j+3}$ become a 1-regular. In other words, we transform the factor 211 into 200. Then we apply the idea of Lemma 21 to further transform it into 2.

Repeated applications of the above transformation allows us to assume that $\gamma$ contains exactly one 1, i.e. $\beta$ contains 212 as a factor. Let $V_j, V_{j+1}, V_{j+2}$ and $V_{j+3}$ be the four columns such that $V_j \cup V_{j+1}$ and $V_{j+2} \cup V_{j+3}$ induce chain graphs and $V_{j+1} \cup V_{j+2}$ induces the bipartite complement of a 1-regular graph. Let $x$ be the vertex in the first row of column $V_{j+1}$ and $y$ be the vertex in the last row of column $V_{j+2}$. It is not difficult to see that if we pivot the edge $xy$ and delete the first and the last row, then the graph induced by $V_{j+1} \cup V_{j+2}$ becomes 1-regular, while the graphs induced by $V_j \cup V_{j+1}$ and by $V_{j+2} \cup V_{j+3}$ remain chain graphs. In other words, we transform the factor 212 into 202. Then we apply the idea of Lemma 27 to further transform it into 22.

The above procedure applied at most $n-1$ times allows us to transform $\beta$ into the word of $n$ consecutive 2s. In terms of graphs, $G_n$ transforms into a sequence of $n$ chain graphs. Moreover, it is not difficult to see that if initially $G_n$ contains $n2^n + n^2$ rows, then the resulting graph has at least $n$ rows, i.e. it contains $X_{n,n}$ as a vertex minor. $\square$

## 5.4 Conclusion and open problems

In the preceding sections, we have described a new family of hereditary classes of graphs of unbounded clique-width. For many of them, we proved the minimality. Our results allow us to make the following conclusion.

**Theorem 30.** *There exist infinitely many minimal hereditary classes of graphs of unbounded clique-width and linear clique-width.*

*Proof.* Let $n$ be a natural number and $\alpha^{(n)} = (0^n 1)^\infty$. Since $\alpha^{(n)}$ is an infinite periodic word, by Theorem 29 $\mathcal{G}_{\alpha^{(n)}}$ is a minimal class of unbounded clique-width and linear clique-width.

If $n < m$, then $\mathcal{G}_{\alpha^{(n)}}$ and $\mathcal{G}_{\alpha^{(m)}}$ do not coincide, since $\mathcal{G}_{\alpha^{(n)}}$ contains an induced $C_{2(n+2)}$, while $\mathcal{G}_{\alpha^{(m)}}$ does not (which it is not difficult to see). Therefore, $\mathcal{G}_{\alpha^{(1)}}, \mathcal{G}_{\alpha^{(2)}}, \ldots$ is an infinite sequence of minimal hereditary classes of graphs of unbounded clique-width and linear clique-width. $\square$

A full description of minimal classes of the form $\mathcal{G}_\alpha$ remains an open question. To propose the conjecture addressing this question, we first define the notion of almost periodic word. An infinite word $\alpha$ is *almost periodic* if for each factor $\beta$ of $\alpha$ there exists a constant $\ell(\beta)$ such that every factor of $\alpha$ of length at least $\ell(\beta)$ contains $\beta$ as a factor.

**Conjecture 1.** Let $\alpha$ be an infinite word over the alphabet $\{0, 1, 2\}$ with at least one appearance of 1 or 2. The class $\mathcal{G}_\alpha$ is a minimal hereditary class of unbounded clique-width if and only if $\alpha$ is almost periodic.

Note that the conditions of Conjecture 1 imply that either 1 or 2 appears in $\alpha$ infinitely many times. It is not hard to verify that this condition is necessary for the class $\mathcal{G}_\alpha$ to have unbounded clique-width, in other words if $\alpha$ contains finitely many 1s and 2s the class $\mathcal{G}_\alpha$ has bounded clique-width.

We conclude this section by discussing an intriguing relationship between clique-width in a hereditary class $X$ and the existence of infinite antichains in $X$ with respect to the induced subgraph relation. In particular, the following question was asked in [28]: is it true that if the clique-width in $X$ is unbounded, then it necessarily contains an infinite antichain? Recently, this question was answered negatively in [49]. However, in the case of so-called *labelled* induced subgraphs, the question remains open.

> *Labelled induced subgraphs.* We define this notion for two labels (or colors), which is the simplest case where the above question is open. Assume we deal with graphs whose vertices are colored by two colors, say white and black. We say that a graph $H$ is a *labelled induced subgraph* of $G$ if there is an embedding of $H$ into $G$ that respects the colors. With this strengthening of the induced subgraph relation, some graphs that are comparable without labels may become incomparable if equipped

with labels. Consider, for instance, two chordless paths $P_k$ and $P_n$. Without labels, one of them is an induced subgraph of the other. Now imagine that we color the endpoints of both paths black and the remaining vertices white. Then clearly they become incomparable with respect to the labelled induced subgraph relation (if $k \neq n$). Therefore, the set of all paths colored in this way create an infinite labelled antichain. Let us denote it by $A_0$.

In [28], it was conjectured that hereditary classes of graphs of unbounded clique-width necessarily contain infinite labelled antichains. We believe this is true. Moreover, we propose the following strengthening of the conjecture from [28].

**Conjecture 2.** Every minimal hereditary class of graphs of unbounded clique-width contains a canonical infinite labelled antichain.

The notion of a canonical antichain was introduced by Guoli Ding in [29] and can be defined for hereditary classes as follows. An infinite antichain $A$ in a hereditary class $X$ is canonical if any hereditary subclass of $X$ containing only finitely many graphs from $A$ has no infinite antichains. In other words, speaking informally, an antichain is canonical if it is unique in the class.

To support Conjecture 2, let us observe that it is valid for all minimal classes $\mathcal{G}_\alpha$ described in Theorem 29. Indeed, all of them contain arbitrarily large chordless paths and hence all of them contained the infinite labelled antichain $A_0$ defined above. Moreover, this antichain is canonical, because by forbidding all paths of length greater than $k$ for some fixed $k$, we are left with subgraphs of $P^\alpha$ occupying at most $k$ consecutive columns, in which case the clique-width of such graphs is at most $4k$ by Lemma 24.

There exist many other infinite labelled antichains, but all available examples are obtained from the antichain $A_0$ by various transformations. We believe that any infinite labelled antichain can be transformed from $A_0$ in a certain way and that any minimal hereditary class of unbounded clique-width can be transformed from $\mathcal{P}^\alpha$ (for some $\alpha$) in a similar way. Describing the set of these transformations is a challenging research task.

# Chapter 6

# Factorial properties and implicit representations of graphs

Every simple graph on $n$ vertices can be represented by a binary word of length $\binom{n}{2}$ (half of the adjacency matrix), and if no a priory information about the graph is known, this representation is best possible in terms of its length. However, if we know that our graph belongs to a particular class (possesses a particular property), this representation can be shortened. For instance, the Prüfer code allows representing a labelled tree with $n$ vertices by a word of length $(n-2)\log n$ (in binary encoding)[1]. For labelled graphs, i.e. graphs with vertex set $\{1, 2, \ldots, n\}$, we need $\log n$ bits for each vertex just to represent its label. That is why graph representations requiring $O(\log n)$ bits per vertex have been called in [38] *implicit*.

Throughout this section by representing a graph we mean its coding, i.e. representing by a word in a finite alphabet (in our case the alphabet is always binary). Moreover, we assume that different graphs are mapped to different words (i.e. the mapping is injective) and that the graph can be restored from its code. For an implicit representation, we additionally require that the code of the graph consists of the codes of its vertices, each of length $O(\log n)$, and that the adjacency of two vertices, i.e. the element of the adjacency matrix corresponding to these vertices, can be computed from their codes.

Not every class of graphs admits an implicit representation, since a bound on the total

[1]All logarithms in this section are of base 2

length of the code implies a bound on the number of graphs admitting such a representation. More precisely, only classes containing $2^{O(n \log n)}$ graphs with $n$ vertices can admit an implicit representation. However, this restriction does not guarantee that graphs in such classes can be represented implicitly. A simple counter-example can be found in [65]. Even with further restriction to *hereditary classes*, i.e. those that are closed under taking induced subgraphs, the question is still not so easy. The authors of [38], who introduced the notion of an implicit representation, conjectured that every hereditary class with $2^{O(n \log n)}$ graphs on $n$ vertices admits an implicit representation, and this conjecture is still open.

In the terminology of [11], hereditary classes containing $2^{O(n \log n)}$ labelled graphs on $n$ vertices are at most *factorial*, i.e. have at most factorial speed of growth. Classes with speeds lower than factorial are well studied and have a very simple structure. The family of factorial classes is substantially richer and the structure of classes in this family is more diverse. It contains many classes of theoretical or practical importance, such as line graphs, interval graphs, permutation graphs, threshold graphs, forests, planar graphs and, even more generally, all proper minor-closed graph classes [52], all classes of graphs of bounded vertex degree, of bounded clique-width [6], etc.

In spite of the crucial importance of the family of factorial classes, except the definition very little can be said about this family in general, and the membership in this family is an open question for many specific graph classes. To simplify the study of this question, we will introduce a number of tools and apply them to reveal new members of this family. For some of them, we do even better and find an implicit representation.

## 6.1 Tools

### 6.1.1 Modular decomposition

Given a graph $G$ and a subset $U \subset V(G)$, we say that a vertex $x$ outside of $U$ distinguishes $U$ if it has both a neighbour and a non-neighbour in $U$. A proper subset of $V(G)$ is called a *module* if it is indistinguishable by the vertices outside of the set. A module is *trivial* if it consists of a single vertex, no vertices or every vertex. A graph every module of which is trivial is called *prime*.

It is well-known (and not difficult to see) that a graph $G$ which is connected and co-

connected (the complement to a connected graph) admits a unique partition into maximal modules. Moreover, for any two maximal modules $M_1$ and $M_2$, the graph $G$ contains either all possible edges between $M_1$ and $M_2$ or none of them. Therefore, by contracting each maximal module into a single vertex we obtain a graph which is prime (due to the maximality of the modules). This property allows a reduction of various graph problems from the set of all graphs in a hereditary class $X$ to prime graphs in $X$. In what follows, we show that the question of deciding whether a hereditary class is at most factorial also allows such a reduction. We start with the following technical lemma.

**Lemma 29.** *For any positive integers $k < n$, and $n_1, n_2, \ldots, n_k$ such that $n_1 + n_2 + \cdots + n_k = n$, the following inequality holds:*

$$k \log k + n_1 \log n_1 + n_2 \log n_2 + \ldots + n_k \log n_k \leq n \log n.$$

*Proof.* For $k = 1$, the statement is trivial. Let $k > 1$. The derivative of $f_a(x) = x \log x + (a-x)\log(a-x)$ is $\log x - \log(a-x)$, which is non-negative for $x \geq \frac{a}{2}$. In particular, this implies that for any two integers $m \geq n > 1$ we have $f_{n+m}(m) \leq f_{n+m}(m+1)$. Hence,

$$m \log m + n \log n \leq (m+1)\log(m+1) + (n-1)\log(n-1). \tag{6.1}$$

Denote $n_0 = k$ and let $s$ be a number in $\{0, 1, \ldots, k\}$ such that $n_s \geq n_i$ for all $i = 0, 1, \ldots, k$. Applying inequality (6.1) $(n_0 - 1) + \ldots + (n_k - 1) - (n_s - 1) = n - n_s$ times we obtain:

$$n_0 \log n_0 + \ldots + n_k \log n_k \leq (n_s + n - n_s)\log(n_s + n - n_s) + 1 \log 1 + \ldots + 1 \log 1 = n \log n. \quad \square$$

**Theorem 31.** *Let $X$ be a hereditary class of graphs. If the number of prime $n$-vertex graphs in $X$ is $2^{O(n \log n)}$, then the number of all $n$-vertex graphs in $X$ is $2^{O(n \log n)}$.*

*Proof.* For convenience, let us extend the notion of prime graphs by including in it all complete and all empty graphs. For each $n > 2$, this extension adds to the set of prime graphs just two graphs, so we may assume the number of graphs in our class is at most $2^{cn \log n}$ for a constant $c > 0$.

For $n \geq 2$, let $f_n$ be an injection from the set of prime $n$-vertex graphs in $X$ to the binary sequences of length at most $cn \log n$. For each prime graph $P$ on $n \geq 2$ vertices, let $f(P) = |n^{bin}|f_n(P)|$, where $n^{bin}$ is the binary expression of $n$. Thus, $f$ is an injection from the set of prime graphs in $X$ to the set of ternary words (i.e. words in the alphabet of three symbols $\{0, 1, |\}$). For each $n$-vertex prime graph $P \in X$ the length of the word $f(P)$ is at most $cn \log n + \log n + 3$. Observe that $cn \log n + \log n + 3 \leq (c+2)n \log n$ for $n \geq 2$. Therefore, each $n$-vertex prime graph in $X$ is represented by a ternary word of length at most $(c+2)n \log n$ for $n \geq 2$. We claim that all the graphs in $X_n$ can be represented by different ternary words of length at most $(c+3)n \log n + n$.

Given a graph $G \in X_n$ we construct a modular decomposition tree $T$ of $G$ in which each node $x$ corresponds to an induced subgraph of $G$, denoted $G_x$, and has a label, denoted $L_x$. For the root, we define $G_x = G$. To define the children of $x$ and its label, we proceed as follows.

- Assume $G_x$ has at least two vertices, then
  - If $G_x$ is disconnected, we decompose it into connected components, associate each connected component with a child of $x$, and define $L_x = f(O_k)$, where $k$ is the number of connected components.
  - If $G_x$ is the complement to a disconnected graph, then we decompose it into co-components (connected components of the complement), associate each co-component with a child of $x$, and define $L_x = f(K_k)$, where $k$ is the number of co-components.
  - If both $G_x$ and its complement are connected, then we decompose $G$ into maximal modules, associate each module with a child of $x$, and define $L_x = f(G_x^*)$, where $G_x^*$ is the prime graph obtained from $G_x$ by contracting each maximal module into a single vertex.
- Assume $G_x$ has just one vertex, and let $j \in \{1, 2, \ldots, n\}$ be the label of that vertex in $G$. Then we define $x$ to be a leaf in $T$ and $L_x = j^{bin}$, where $j^{bin}$ is the binary expression of $j$ of length $\log n$.

If $x$ is a non-leaf node of $T$, then it has $k \geq 2$ children, in which case its label has length at most $(c+2)k \log k$. Otherwise $x$ is a leaf and its label has length $\log n$. Let $f(G)$ be the concatenation of the labels of all the nodes of $T$ in the order they appear in the depth-first search algorithm applied to $T$. Since the labels record the number of children for each node, it is not hard to see that we can reconstruct the original tree $T$ from the word $f(G)$, and hence we can reconstruct the graph $G$ from $f(G)$, i.e. $f$ is an injection.

Let us prove that the length of the word $f(G)$ is at most $(c+3)n\log n + n$. The leaf nodes of $T$ contribute $n\log n$ bits to $f(G)$. Now by induction on $n \geq 2$ we show that the remaining nodes of $T$ contribute at most $(c+2)n\log n + n$ symbols to $f(G)$. For $n = 2$, this follows from the first part of the proof. Now assume $n > 2$. Let the root of the tree $T$ have $k$ children corresponding to induced subgraphs $G_1, \ldots, G_k$ of $G$ of sizes $n_1, n_2, \ldots n_k$ with $n_1 + n_2 + \cdots + n_k = n$. Since $n_i < n$, by the induction hypothesis the internal nodes of $T_{G_i}$ contribute at most $(c+2)n_i \log n_i + n_i$ symbols to $f(G_i)$, where $T_{G_i}$ is the subtree of $T$ rooted at the vertex corresponding to subgraph $G_i$. Also, the label of the root has length at most $(c+2)k\log k$. Clearly the set of internal (non-leaf) nodes of $T$ coincides with the union of internal nodes of $T_{G_1} \ldots T_{G_k}$ and the root of $T$. Hence, by Lemma 29, the internal nodes of $T$ contribute at most $(c+2)n\log n + n$ symbols to $f(G)$.

Since we used 3 letters to represent graphs from $X$, the number of graphs in $X_n$ is at most $3^{(c+3)n\log n+n} \leq 3^{(c+4)n\log n} = 2^{c'n\log n}$, where $c' = (c+4)\log 3$, i.e. $|X_n| = 2^{O(n\log n)}$. □

**Corollary 32.** *If the set of prime graphs in a hereditary class $X$ belongs to a class which is at most factorial, then $X$ is at most factorial.*

### 6.1.2 Functional vertices

In this section, we introduce one more tool which is helpful in deciding whether a given class of graphs is factorial or not. We repeat that by $m(x, y)$ we denote the element of the adjacency matrix corresponding to vertices $x$ and $y$.

**Definition 12.** For a graph $G = (V, E)$, we say that a vertex $y \in V$ is a function of a set of vertices $x_1, \ldots, x_k \in V$ if there exists a Boolean function $f : B^k \to B$ of $k$ variables such that for any vertex $z \in V \setminus \{y, x_1, \ldots, x_k\}$,

$$m(y, z) = f(m(x_1, z), \ldots, m(x_k, z)).$$

**Theorem 33.** *Let $X$ be a hereditary class of graphs and $c$ be a constant. If for every graph $G$ in $X$ there is a vertex $y$ and two disjoint sets $U$ and $R$ each of at most $c$ vertices* *such that $y$ is a function of $U$ in the graph $G \setminus R$, then $|X_n| = 2^{O(n \log n)}$.*

*Proof.* To prove the theorem, we will show by induction on $n$ that any $n$-vertex graph in this class can be described by $(2c+1)n \log n + (2^c + 2c)n$ bits. This is clearly true for $n = 1$ or $n = 2$, so assume that every $(n-1)$-vertex graph in $X$ admits a description by a binary word of length at most $(2c+1)(n-1)\log(n-1) + (2^c + 2c)(n-1)$. Let $G$ be a graph in $X$ with $n$ vertices, $y$ a vertex in $G$ and $U = \{x_1, \ldots, x_k\}, R$ two sets as described in the statement of the theorem. For ease of notation we will call $y$ the functional vertex.

To obtain a description of $G$, we start by describing the label of $y$ by a binary word of length $\log n$. Next, we list each of the labels of the vertices in $R$, following each with a 0 if $y$ is not adjacent to the vertex and a 1 if $y$ is adjacent to the vertex. As there are at most $c$ vertices in $R$, this requires at most $c \log n + c$ bits. Next, we list each of the labels of the vertices in $U$, following each with a 0 if $y$ is not adjacent to the vertex and a 1 if $y$ is adjacent to the vertex. Similarly, this requires at most $c \log n + c$ bits. Then, as we know that $y$ is a function of the vertices in $U$ in the graph $G \setminus R$, there is a Boolean function that describes the adjacencies of the vertices in $G \setminus \{U \cup R \cup y\}$ to $y$. List the image of this function next. This requires at most $2^c$ bits, as there are at most $c$ vertices in $U$. Finally, append the description of the graph $G \setminus \{y\}$ which requires at most $(2c+1)(n-1)\log(n-1) + (2^c + 2c)(n-1)$ bits by induction. So we have a description of $G$ by a binary word of length at most $(2c+1)\log n + (2^c + 2c) + (2c+1)(n-1)\log(n-1) + (2^c + 2c)(n-1)$ bits. Finally we see that

$$(2c+1)\log n + (2^c + 2c) + (2c+1)(n-1)\log(n-1) + (2^c + 2c)(n-1) \leq$$

$$(2c+1)\log n + (2^c + 2c) + (2c+1)(n-1)\log n + (2^c + 2c)(n-1) =$$

$$(2c+1)n \log n + (2^c + 2c)n$$

hence the result holds by induction.

For any two different vertices in $G$ this description can be used to identify if they are adjacent or not. If both vertices are different from $y$, their adjacency can be determined from the description of the graph $G \setminus \{y\}$. Assume now that one of the vertices is $y$ and let $z$ be the other vertex. If $z \in R \cup U$, then the bit $m(y, z)$ is explicitly included in the description of $G$. If $z \in V(G) \setminus \{U \cup R \cup y\}$, then $m(y, z) = f(m(x_1, z), \ldots, m(x_k, z))$.

Note that the bits $m(x_1, z), \dots, m(x_k, z)$ can be determined from the description of the graph $G \setminus \{y\}$, while the value of the function $f$ can be found in the description of $G$. This completes the proof of the theorem. $\square$

A trivial example of a functional vertex is a vertex of bounded degree or co-degree, in which case the set $U$ of variable vertices is empty. In this case, we can make a stronger conclusion.

**Lemma 30.** *Let $X$ be a hereditary class and $d$ a constant. If every graph in $X$ has a vertex of degree or co-degree at most $d$, then $X$ admits an implicit representation.*

*Proof.* Since $X$ is hereditary, every graph $G$ in $X$ admits a linear order $P = (v_{i_1}, \dots, v_{i_n})$ of its vertices so that $v_{i_j}$ has degree or co-degree at most $d$ in the subgraph induced by vertices $(v_{i_j}, v_{i_{j+1}}, \dots, v_{i_n})$. Then an implicit representation for $G$ can be obtained by recording for each vertex $v$ its position in the linear order $P$ and at most $d$ of its neighbours or non-neighbours among the vertices following $v$ in $P$. One more bit is needed to indicate whether $v$ has at most $d$ neighbours or at most $d$ non-neighbours. Clearly, this description completely defines the graph and hence provides an implicit representation for $G$. $\square$

The case when $y$ is a functional vertex with $U = \{x\}$ and $f$ being a Boolean function of one variable mapping 0 to 0 and 1 to 1 can be described as follows: $|N(x)\Delta N(y)| \leq c$, where $\Delta$ denotes the symmetric difference of two sets. This observation implies the following corollary which will be frequently used in the subsequent sections.

**Corollary 34.** *Let $X$ be a hereditary class of graphs and $c$ be a constant. If for every graph $G$ in $X$ there exist two vertices $x, y$ such that $|N(x)\Delta N(y)| \leq c$, then $X$ is at most factorial.*

### 6.1.3 Covering of graphs

#### Locally bounded covering

The idea of locally bounded coverings was introduced in [48] to study factorial properties of graphs. This idea can be described as follows.

Let $G$ be a graph. A set of graphs $H_1, \dots, H_k$ is called a *covering* of $G$ if the union of $H_1, \dots, H_k$ coincides with $G$, i.e. if $V(G) = \bigcup_{i=1}^{k} V(H_i)$ and $E(G) = \bigcup_{i=1}^{k} E(H_i)$.

**Lemma 31** ([48])**.** *Let $X$ be a class of graphs and $c$ a constant. If every graph $G \in X$ can be covered by graphs from a class $Y$ with $\log Y_n = O(n \log n)$ in such a way that every vertex of $G$ is covered by at most $c$ graphs, then $\log X_n = O(n \log n)$.*

Now we derive a similar result for implicit representations of graphs.

**Lemma 32.** *Let $X$ be a class of graphs and $c$ a constant. If every graph $G \in X$ can be covered by graphs from a class $Y$ admitting an implicit representation in such a way that every vertex of $G$ is covered by at most $c$ graphs, then graphs in $X$ also admit an implicit representation.*

*Proof.* Let $H_1, \ldots, H_k \in Y$ be a covering of a graph $G \in X$ such that every vertex of $G$ is covered by at most $c$ graphs, where $c$ is a constant independent of $G$. Denote $n_i = |V(H_i)|$, $n = |V(G)|$. Then

$$k \leq \sum_{i=1}^{k} n_i \leq cn. \tag{6.2}$$

Let $\phi_i$ be an implicit representation of $H_i$, i.e. a binary word containing for each vertex of $H_i$ a code of length $O(\log n_i)$ so that the adjacency of two vertices can be computed from their codes.

Now we construct an implicit representation of $G$ as follows. To each vertex $j \in V(G)$ we assign a binary word $\psi_j$ containing for each graph $H_i$ covering $j$ the index $i$ and the code of vertex $j$ in the representation $\phi_i$ of $H_i$. Clearly, the adjacency of two vertices $j, k \in V(G)$ can be determined from their codes $\psi_j$ and $\psi_k$, because they are adjacent in $G$ if and only if there is a graph $H_i$ which covers both of them and in which these vertices are adjacent. Since each vertex $j \in V(G)$ is covered by at most $c$ graphs, the length of $\psi_j$ is at most $c \log k + \sum_{i=1}^{c} O(\log n_{j_i}) = c(\log k + O(\log n))$. Together with (6.2) this implies that $|\psi_j| = O(\log n)$ and hence $\{\psi_j \ : \ j = 1, \ldots, n\}$ is an implicit representation of $G$. $\square$

### Partial covering

One more tool was introduced in [27] and can be stated as follows.

**Lemma 33.** *Let $X$ be a hereditary class. If there is a constant $d \in \mathbb{N}$ and a hereditary class $Y$ with at most factorial speed of growth such that every graph $G = (V, E) \in X$*

*contains a non-empty subset $A \subseteq V$ such that $G[A] \in Y$ and each vertex $a \in A$ has either at most $d$ neighbours or at most $d$ non-neighbours in $V - A$, then $X$ is at most factorial.*

Now we derive a similar result for implicit representations. This result can be viewed as a generalization of Lemma 30.

**Lemma 34.** *Let $X$ be a hereditary class. If there is a constant $d \in \mathbb{N}$ and a hereditary class $Y$ which admits an implicit representation such that every graph $G = (V, E) \in X$ contains a non-empty subset $A \subseteq V$ such that $G[A] \in Y$ and each vertex $a \in A$ has either at most $d$ neighbours or at most $d$ non-neighbours in $V - A$, then $X$ admits an implicit representation.*

*Proof.* First, we represent $G[A]$ implicitly (which is possible, because $G[A] \in Y$ and $Y$ admits an implicit representation) and then add to the code of each vertex $v$ of $G[A]$ the list of at most $d$ neighbours or non-neighbours of $v$ in the rest of the graph. This describes $G[A]$ and its adjacency to the rest of the graph implicitly, i.e. with $O(\log n)$ bits per each vertex of $A$. Then the set $A$ can be deleted (or simply ignored) and the procedure can be applied to the rest of the graph, which is possible because $X$ is a hereditary class. Eventually, we obtain an ordered sequence of sets $A_0 = A, A_1, A_2, \ldots, A_k$ $(k \leq n)$ such that for each $i \geq 0$, the graph $G[A_i]$ and its adjacency to the vertices in $A_{i+1}, \ldots, A_k$ are described implicitly. To complete the description of $G$, we assign to each vertex $v \in V(G)$ the index the set $A_i$ it belongs to. Now the adjacency of two vertices $u, v \in V(G)$ can be tested as follows: if both of them belong to the same set $A_i$, then their adjacency can be determined through their codes in the implicit representation of $G[A_i]$, and if $u \in A_i$ and $v \in A_j$ with $i < j$, then their adjacency can be determined by looking at the list of neighbours (or non-neighbours) of $u$ which is stored in the label of $u$. $\square$

### 6.1.4 Remarks

In Theorem 33, Lemma 30, Corollary 34, Lemmas 33 and 34, to prove the corresponding statements for a class $X$, we require that *every* graph in $X$ has a subset of vertices (or a single vertex) satisfying certain properties. This requirement can be relaxed if some graphs in $X$ belong to a class $Z$ that satisfy conditions of the corresponding statement. In this case, the existence of a subset (or a vertex) with a particular property can be required only for graphs in $X - Z$. For easy reading, we do not introduce this relaxation into the text of the corresponding results. But we keep it in mind when we apply these results in the next section.

## 6.2 Applications

In this section, we apply the tools developed in the previous one in order to reveal new factorial classes of graphs. In some cases, we show that these classes admit an implicit representation. To simplify the study of factorial graph properties, in [47] the following conjecture was proposed.

**Conjecture 3.** A hereditary graph property $X$ is factorial if and only if the fastest of the following three properties is factorial: bipartite graphs in $X$, co-bipartite graphs in $X$, split graphs in $X$.

To justify this conjecture we observe that if in the text of the conjecture we replace the word "factorial" by any of the lower layers (constant, polynomial or exponential), then the text becomes a valid statement. Also, the "only if" part of the conjecture is true, because all minimal factorial classes are subclasses of bipartite, co-bipartite or split graphs. Also, in [47] this conjecture was verified for all hereditary classes defined by forbidden induced subgraphs with at most 4 vertices.

The above conjecture reduces the question of deciding the membership in the factorial layer from the family of all hereditary properties to those which are bipartite, co-bipartite and split. Taking into account the obvious relationship between bipartite, co-bipartite and split graphs, this question can be further reduced to hereditary properties of bipartite graphs only.

When we talk about bipartite graphs, we assume that each graph is given together with a bipartition of its vertex set into two parts (independent sets), say top and bottom, and we denote a bipartite graph with parts $A$ and $B$ by $G = (A, B, E)$, where $E$, as before, stands for the set of edges. The *bipartite complement* of a bipartite graph $G = (A, B, E)$ is the bipartite graph $\widetilde{G} = (A, B, E')$, where two vertices $a \in A$ and $b \in B$ are adjacent in

$G$ if and only if they are not adjacent in $\widetilde{G}$. By $O_{n,m}$ we denote the bipartite complement of $K_{n,m}$.

For connected graphs, the bipartition into two independent sets is unique (up to symmetry). A disconnected bipartite graph can admit several different bipartitions, and this distinction between different bipartitions can be crucial if our graph is *forbidden*. Consider, for instance, the graph $2K_{1,2}$ (a disjoint union of two copies of $K_{1,2}$). Up to symmetry, it admits two different bipartitions and by forbidding one of them we obtain a subclass of bipartite graphs which is factorial, while by forbidding the other we obtain a superfactorial subclass of bipartite graphs. This is because the *bipartite complement* of one of them does not contain any cycle, while the *bipartite complement* of the other contains a $C_4$. More generally, in [5] the following result was proved.

**Theorem 35.** *Let $G$ be a bipartite graph. If either $G$ or its bipartite complement contains a cycle, then the class of $G$-free bipartite graphs is superfactorial. If both $G$ and its bipartite complement are acyclic and $G$ is different from $P_7$, then the class of $G$-free bipartite graphs is at most factorial.*

Moreover, for most bipartite graphs $G$ such that neither $G$ nor its bipartite complement contains a cycle, paper [5] proves a stronger result. To state this result, let us observe that when we say that a bipartite graph $G$ contains a bipartite graph $H$ as an induced subgraph, we do not specify which part of $H$ is mapped to which part of $G$. However, sometimes this specification is important and if all induced copies of $H$ appear in $G$ with all bottom parts of $H$ being in the same part of $G$, then we say that $H$ is contained in $G$ *one-sidedly*. If at least one of the two possible appearances of $H$ is missing in $G$, we say that $G$ contains no *one-sided* copy of $H$.

**Theorem 36** ([5])**.** *If both $G$ and its bipartite complement are acyclic and $G$ is different from $P_7$, $S_{1,2,3}$, $S_{1,2,2}$ and from the bipartite complement of $S_{1,2,2}$, then the class of bipartite graphs containing no one-sided copy of $G$ is at most factorial.*

According to Theorem 35, the class of $P_7$-free bipartite graphs is the only subclass of bipartite graphs defined by a single forbidden induced subgraph for which the membership in the factorial layer is unknown. However, more recently in [50] it was proved that the class of $P_7$-free bipartite graphs is indeed at most factorial.

To better understand the structure of $P_7$-free bipartite graphs, in this section we study subclasses of this class defined by one additional forbidden induced subgraph and prove, using different methods to those in [50], that for every graph $G$ with at most 6 vertices the class of $(P_7, G)$-free bipartite graphs is at most factorial.

Many of our results can be extended, with no extra work, to the more general case of bipartite graphs of bounded chordality, i.e. $(C_k, C_{k+1}, \ldots)$-free bipartite graphs for a constant $k$ (the *chordality* of a graph is the length of a longest chordless cycle). For $k = 4$, the class of $(C_k, C_{k+1}, \ldots)$-free bipartite graphs coincides with forests and this class is factorial. However, for any $k > 4$, the class of $(C_k, C_{k+1}, \ldots)$-free bipartite graphs is superfactorial. In particular, the class of $(C_6, C_8, \ldots)$-free bipartite graphs, also known as chordal bipartite graphs, is superfactorial, as the number of $n$-vertex labelled graphs in this class is $2^{\Theta(n \log^2 n)}$ [64]. Moreover, the class of chordal bipartite graphs is not a minimal superfactorial class, which is due to the following result proved in [27], where $2C_4$ denotes the disjoint union of 2 copies of $C_4$, and $2C_4 + e$ is the graph obtained from $2C_4$ by adding one edge between the two copies of $C_4$.

**Lemma 35.** *The class of* $(2C_4, 2C_4 + e)$*-free chordal bipartite graphs is superfactorial.*

On the other hand, most of the hereditary subclasses of chordal bipartite graphs studied in the literature, such as forests, bipartite permutation, convex graphs, are factorial. Also, several results on factorial properties of chordal bipartite graphs were obtained in [27] and [51]. In particular, in [51] the following result was proved.

**Lemma 36.** *For any forest $F$, the class of $F$-free chordal bipartite graphs is at most factorial.*

This result cannot be extended to $(C_k, C_{k+1}, \ldots)$-free bipartite graphs for $k > 6$, because, for instance, $(C_{10}, C_{11}, \ldots)$-free bipartite graphs contain all $P_8$-free bipartite graphs, which is a superfactorial class (as the bipartite complement of $P_8$ contains a $C_4$), and $(C_8, C_{10}, \ldots)$-free bipartite graphs contain $P_7$-free bipartite graphs, a class for which the membership in the factorial layer is an open question.

However, for some graphs $G$ containing a cycle, it is possible to prove the membership of $(G, C_k, C_{k+1}, \ldots)$-free bipartite graphs in the factorial layer for any value of $k$. For

$k = 6$ (i.e. for chordal bipartite graphs) several results of this type have been obtained in [27]. In Section 6.2.1, we extend these results to bipartite graphs of chordality at most $k$ for arbitrary value of $k$. We also obtain a number of new results for such classes.

In Section 6.2.2, we restrict ourselves further and consider subclasses of $P_7$-free bipartite graphs, which is a special case of $(C_8, C_{10}, \ldots)$-free bipartite graphs. We systematically study subclasses of $P_7$-free bipartite graphs defined by one additional forbidden induced subgraph and show that for every graph $G$ with at most 6 vertices the class of $(P_7, G)$-free bipartite graphs is at most factorial.

### 6.2.1 Bipartite graphs of small chordality

In this section, we study $(C_k, C_{k+1}, \ldots)$-free bipartite graphs. For $k = 6$, this class is known as chordal bipartite graphs and is known to be superfactorial [64]. Therefore, bipartite graphs of chordality at most $k$ constitute a superfactorial class for all $k \geq 6$. Various factorial properties of chordal bipartite graphs were studied in [27]. In the present section, we generalize most of them to arbitrary values of $k$ and obtain a number of new results for such classes. We start with the following general result.

**Lemma 37.** *For any natural numbers $p \geq 2$ and $k \geq 6$, the class of $(K_{p,p}, C_k, C_{k+1}, \ldots)$-free bipartite graphs admits an implicit representation and hence is at most factorial.*

*Proof.* In [45], it was shown that for every graph $H$ and for every natural $p$, there exists $d = d(H, p)$ such that every graph of average degree at least $d$ contains either a $K_{p,p}$ as a (not necessarily induced) subgraph or an induced subdivision of $H$. This implies that every $(K_{p,p}, C_k, C_{k+1}, \ldots)$-free bipartite graph $G$ contains a vertex of degree less than $d(C_k, p)$, since otherwise the average degree of $G$ is at least $d(C_k, p)$, in which case it must contain either an induced subdivision of $C_k$ (which is forbidden) or a $K_{p,p}$ as a subgraph (which is forbidden either, else an induced copy of $K_{p,p}$ or $K_3$ arises). This implies, by Lemma 30, that the class of $(K_{p,p}, C_k, C_{k+1}, \ldots)$-free bipartite graphs admits an implicit representation and hence is at most factorial. □

For $k = 6$, i.e. for chordal bipartite graphs, the result of Lemma 37 was derived, by different arguments, in [27]. In particular, in that paper it was proved that $K_{p,p}$-free chordal bipartite graphs have bounded tree-width. This is a stronger conclusion and we believe that the same conclusion holds for $K_{p,p}$-free bipartite graphs of chordality at most $k$ for each value of $k$. More generally we conjecture:

**Conjecture**. *For all $r$, $p$ and $k$, there is a $t = t(r, p, k)$ such that any $(K_r, K_{p,p})$-free graph of chordality at most $k$ has tree-width at most $t$.*

We leave this conjecture for future research. In this section, we extend the result of Lemma 37 in a different way. In [27], it was proved that the class of chordal bipartite graphs containing no induced $K_{p,p} + K_1$ is at most factorial by showing that every $K_{p,p} + K_1$-free bipartite graph containing a $K_{s,s}$ with $s = p(2^{p-1} + 1)$ contains a vertex which has at most $2p - 2$ non-neighbours in the opposite part. Together with Lemma 30, this immediately implies the following extension of Lemma 37.

**Lemma 38.** *For any natural $p \geq 2$ and $k \geq 6$, the class of $(K_{p,p} + K_1, C_k, C_{k+1}, \ldots)$-free bipartite graphs admits an implicit representation and hence is at most factorial.*

Below we further extend this result and obtain a number of other results for subclasses of bipartite graphs of bounded chordality.

### $Q(p)$-free bipartite graphs of bounded chordality

We denote by $Q(p)$ the graph obtained from $K_{p,p} + K_1$ by adding a new vertex to the smaller part of the graph and connecting it to every vertex in the opposite side. The graph $Q(2)$ is represented in Figure 6.1.

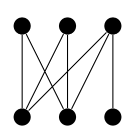


Figure 6.1: Graph $Q(2)$

**Theorem 37.** *For any natural $k$ and $p$, the class of $Q(p)$-free bipartite graphs of chordality at most $k$ admits an implicit representation and hence is at most factorial.*

*Proof.* Let $G$ be a $Q(p)$-free bipartite graph of chordality at most $k$. If $G$ contains no $K_{p^2,p^2}$, it admits an implicit representation by Lemma 37. Therefore, we assume that $G$ contains a $K_{p^2,p^2}$. Moreover, by Lemma 32 we may assume that $G$ is connected.

We denote the two parts in the bipartition of $G$ by $A$ and $B$ and extend the $K_{p^2,p^2}$ contained in $G$ to a maximal (with respect to set inclusion) complete bipartite graph $H$ with parts $A_0 \subseteq A$ and $B_0 \subseteq B$. The set $A - A_0$ can further be split into the set $A_1$ of vertices that have neighbours in $B_0$ and the set $A_2$ of vertices that have no neighbours in $B_0$. Observe that due to the maximality of $H$ each vertex of $A_1$ has at least one non-neighbour in $B_0$. We further split $A_1$ into the set $A_1'$ of vertices with at most $p-1$ non-neighbours in $B_0$ and the set $A_1''$ of vertices with at least $p$ non-neighbours in $B_0$. The set $B - B_0$ can be split into $B_1'$, $B_1''$ and $B_2$ analogously. We claim that

(1) $A_1'' = B_1'' = \emptyset$. Suppose this is not true and let $x$ be a vertex in $A_1''$ (without loss of generality). By definition $x$ must have a neighbour $y$ and $p$ non-neighbours in $B_0$. Then these vertices together with any $p$ vertices in $A_0$ induce a $Q(p)$.

(2) $A_2 = B_2 = \emptyset$. Suppose to the contrary that $A_2$ contains a vertex $x$. Then because of Claim (1) and due to the connectedness of $G$, vertex $x$ must have a neighbour $y \in B_1'$. Since $y$ has at most $p-1$ non-neighbours in $A_0$, it has at least $p$ neighbours in $A_0$. Then these $p$ neighbours together with $x, y$ and any $p$ vertices in $B_0$ induce a $Q(p)$.

(3) The subgraph of $G$ induced by $A_1' \cup B_1'$ is $K_{p,p}+K_1$-free. Assume $G[A_1' \cup B_1']$ contains an induced $K_{p,p}+K_1$ and let, without loss of generality, the $p+1$ vertices of this graph belong to $A_1'$. Each of this $p+1$ vertices has at most $p-1$ non-neighbours in $B_0$ and since $B_0$ contains at least $p^2$ vertices we conclude that there must be a vertex in $B_0$ adjacent to each of the $p+1$ vertices of the copy of $K_{p,p}+K_1$. But then together (that vertex and the copy of $K_{p,p}+K_1$) induce a $Q(p)$ in $G$.

Claim (3) implies by Lemma 38 that $G[A_1' \cup B_1']$ admits an implicit representation. Besides, every vertex of $A_1' \cup B_1'$ has at most $p-1$ non-neighbours in the rest of the graph. Therefore, by Lemma 34 (as well as by Lemma 32) we conclude that $G$ admits an implicit representation. □

### $L(s,p) + O_{0,1}$-free bipartite graphs of bounded chordality

By $L(s,p)$ we denote a bipartite graph obtained from $K_{2,p}$ by adding $s$ pendant edges to one of the vertices of degree $p$. By adding an isolated vertex to the bottom part of the graph, we obtain $L(s,p) + O_{0,1}$ (see example of $L(2,2) + O_{0,1}$ in Figure 6.2).

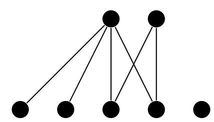


Figure 6.2: Graph $L(2,2) + O_{0,1}$

**Theorem 38.** *For any natural $k, s, p$, the class of $L(s,p) + O_{0,1}$-free bipartite graphs of chordality at most $k$ is at most factorial.*

*Proof.* Let $G$ be an $L(s,p)+O_{0,1}$-free bipartite graph of chordality at most $k$. If $G$ contains no $K_{2,p} + O_{0,1}$, it admits an implicit representation by Lemma 38. Therefore, we assume that $G$ contains an induced copy of $K_{2,p}+O_{0,1}$ and let $x, y$ be the two vertices of degree $p$ in that copy. Vertex $x$ cannot have $s$ or more private neighbours (i.e. neighbours which are not adjacent to $y$), since otherwise any $s$ of these neighbours together with the $K_{2,p}+O_{0,1}$ would induce an $L(s,p) + O_{0,1}$. The analogous statement also holds for $y$. Therefore, $|N(x)\Delta N(y)| \leq 2(s-1)$ and hence, by Corollary 34, the class of $L(s,p) + O_{0,1}$-free bipartite graphs of chordality at most $k$ is at most factorial. $\square$

Observe that the result of Theorem 38 is best possible in the sense that by increasing either of the indices of the second term in the definition of the forbidden graph we obtain a superfactorial class. More precisely:

**Observation 1.** *For any $s, p \geq 1$ and $k \geq 8$, the classes of $L(s,p) + O_{1,1}$-free, $L(s,p) + O_{0,2}$-free and $L(s,p)+O_{2,0}$-free bipartite graphs of chordality at most $k$ are superfactorial.*

This conclusion follows from the fact that $L(s,p) + O_{1,1}$ , $L(s,p) + O_{0,2}$ and $L(s,p) + O_{2,0}$, as well as all bipartite cycles of length more than 8 contain $\widetilde{C}_4$, and hence the corresponding classes contain all $\widetilde{C}_4$-free bipartite graphs, which form a superfactorial class by Theorem 35.

### $M(p)$-free bipartite graphs of bounded chordality

By $M(p)$ we denote the graph obtained from $L(1,p)$ by adding one vertex which is adjacent only to the vertex of degree 1 in the $L(1,p)$. Figure 6.3 represents the graph $M(3)$.

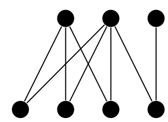


Figure 6.3: Graph $M(3)$

**Theorem 39.** *For any natural $k$ and $p$, the class of $M(p)$-free bipartite graphs of chordality at most $k$ is at most factorial.*

*Proof.* Let $G$ be an $M(p)$-free bipartite graph of chordality at most $k$. If $G$ contains no $K_{p^2,p^2}$, it admits an implicit representation by Lemma 37. Therefore, we assume that $G$ contains a $K_{p^2,p^2}$. Moreover, by Theorem 31 we may assume that $G$ is prime and hence is connected. We denote the two parts in the bipartition of $G$ by $A$ and $B$ and split them into $A_0, A_1', A_1'', A_2$ and $B_0, B_1', B_1'', B_2$ as in Theorem 37.

Let $M^*(p)$ denote the subgraph of $M(p)$ obtained by deleting the vertex of degree $p+1$ (i.e. the only vertex in the smaller part which dominates the other part). By Theorem 36,

(1) *The class of bipartite graphs containing no one-sided copy of $M^*(p)$ is at most factorial.*

The rest of the proof will follow from a series of claims.

(2) $A_2 = B_2 = \emptyset$. Suppose this is not true, then as $G$ is connected there must be a vertex $x \in A_2 \cup B_2$ with a neighbour $y \in A_1 \cup B_1$. Without loss of generality assume that $x \in A_2$, then $x, y$, a neighbour and a non-neighbour of $y$ in $A_0$, and any $p$ vertices in $B_0$ induce an $M(p)$ in $G$.

(3) *No vertex in $A_1''$ has a neighbour in $B_1$.* Indeed, if a vertex $x \in A_1''$ is adjacent to a vertex $y \in B_1$, then $x, y$ together with $p$ non-neighbours of $x$ in $B_0$, a neighbour and a non-neighbour of $y$ in $A_0$ induce an $M(p)$ in $G$.

(4) *No vertex in $B_1''$ has a neighbour in $A_1$* by analogy with (3).

(5) *The subgraph of $G$ induced by $A_1'$ and $B_1'$ is $M^*(p)$-free.* Assume $G[A_1' \cup B_1']$ contains an induced $M^*(p)$ and let, without loss of generality, the $p+1$ vertices of this graph belong to $A_1'$. Each of this $p+1$ vertices has at most $p-1$ non-neighbours in $B_0$ and since $B_0$ contains at least $p^2$ vertices we conclude that there must be a vertex in $B_0$ adjacent to each of the $p+1$ vertices of the copy of $M^*(p)$. But then together (that vertex and the copy of $M^*(p)$) induce an $M(p)$ in $G$.

(6) *The graphs $G[A_0 \cup B_1]$ and $G[B_0 \cup A_1]$ do not contain a one-sided copy of $M^*(p)$.* Indeed, if, say, $G[A_0 \cup B_1]$ contains a one-sided copy of $M^*(p)$ with $p+1$ vertices in $A_0$, then this copy together with any vertex in $B_0$ induces an $M(p)$ in $G$.

This structure obtained for graphs in the class of $M(p)$-free bipartite graphs containing a $K_{p^2,p^2}$ implies that such graphs can be covered by finitely many graphs from a finite union of classes with at most factorial speed of growth. By Lemma 31 we conclude that the class of $M(p)$-free bipartite graphs of chordality at most $k$ is at most factorial for any values of $k$ and $p$. □

**$N(p)$-free bipartite graphs of bounded chordality**

By $N(p)$ we denote the graph $L(1,p)+O_{1,0}$, i.e. the graph obtained from $L(1,p)$ by adding an isolated vertex to the smaller part of graph. Figure 6.4 represents the graph $N(3)$.

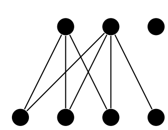


Figure 6.4: Graph $N(3)$

**Theorem 40.** *For any natural $k$ and $p$, the class of $N(p)$-free bipartite graphs of chordality at most $k$ is at most factorial.*

*Proof.* Let $G$ be an $N(p)$-free bipartite graph of chordality at most $k$. If $G$ contains no $K_{p^2,p^2}$, it admits an implicit representation by Lemma 37. Therefore, we assume that $G$

contains a $K_{p^2,p^2}$. Moreover, by Theorem 31 we may assume that $G$ is prime and hence is connected. We denote the two parts in the bipartition of $G$ by $A$ and $B$ and split them into $A_0, A_1', A_1'', A_2$ and $B_0, B_1', B_1'', B_2$ as in Theorem 37.

Let $N^*(p)$ denote the subgraph of $N(p)$ obtained by deleting the vertex of degree $p+1$ (i.e. the only vertex in the smaller part which dominates the other part). By Theorem 36,

(1) *The class of bipartite graphs containing no one-sided copy of $N^*(p)$ is at most factorial.*

The rest of the proof will follow from a series of claims.

(2) *The subgraph of $G$ induced by $A_0$ and $B_1 \cup B_2$ contains no one-sided copy of $N^*(p)$.*

Assume by contradiction that this subgraph contains a copy of $N^*(p)$ with the larger part belonging to $A_0$. Then this copy together with any vertex of $B_0$ induce an $N(p)$.

(3) *Every vertex in $A_1$ is adjacent to every vertex in $B_1'' \cup B_2$.*

To prove this, assume a vertex $x \in A_1$ has a non-neighbour $y \in B_1'' \cup B_2$. By definition of $B_1''$ and $B_2$, vertex $y$ has at least $p$ non-neighbours in $A_0$, while $x$ has a neighbour and a non-neighbour in $B_0$. But then $x, y$, a neighbour and a non-neighbour of $x$ in $B_0$ and any $p$ non-neighbours of $y$ in $A_0$ induce an $N(p)$.

(4) *Every vertex in $B_1$ is adjacent to every vertex in $A_1'' \cup A_2$* by analogy with (3).

(5) *The subgraph of $G$ induced by $A_2$ and $B_2$ contains no one-sided copy of $N^*(p)$.*

To show this, we first observe that if $A_2 \cup B_2$ is not empty, then $A_1 \cup B_1$ is not empty, since otherwise the graph $G$ is disconnected. Therefore, if $A_2 \cup B_2$ is not empty, we may consider a vertex $x \in A_1$ (without loss of generality). Then the subgraph $G[A_2 \cup B_2]$ contains no copy of $N^*(p)$ with the larger part belonging to $B_2$, since otherwise this copy together with vertex $x$ induce an $N(p)$.

(6) *The subgraph of $G$ induced by $A_1'$ and $B_1'$ is $N^*(p)$-free.*

Assume $G[A_1' \cup B_1']$ contains an induced $N^*(p)$ and let the $p+1$ vertices of this graph belong to $A_1'$. Each of this $p+1$ vertices has at most $p-1$ non-neighbours in $B_0$ and since $B_0$ contains at least $p^2$ vertices we conclude that there must be a vertex in $B_0$ adjacent to each of the $p+1$ vertices of the copy of $N^*(p)$. But then together (that vertex and the copy of $N^*(p)$) induce an $N(p)$ in $G$.

This structure obtained for graphs in the class of $N(p)$-free bipartite graphs containing a $K_{p^2,p^2}$ implies that such graphs can be covered by finitely many graphs from a finite union of classes with at most factorial speed of growth. By Lemma 31 we conclude that the class of $N(p)$-free bipartite graphs of chordality at most $k$ is at most factorial for any values of $k$ and $p$. $\square$

According to Observation 1, the result obtained in Theorem 40 is, in a sense, best possible.

**$\mathcal{A}$-free bipartite graphs of bounded chordality**

By $\mathcal{A}$ we denote the graph represented in Figure 7.2.

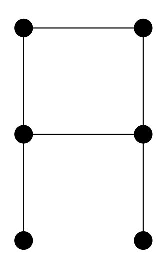


Figure 6.5: The graph $\mathcal{A}$

**Theorem 41.** *For each natural $k$, the class of $\mathcal{A}$-free bipartite graphs of chordality at most $k$ is at most factorial.*

*Proof.* Let $G$ be an $\mathcal{A}$-free bipartite graph of chordality at most $k$. If $G$ contains no $C_4$, it admits an implicit representation by Lemma 37. Therefore, we assume that $G$ contains a $C_4$. Moreover, by Theorem 31 we may assume that $G$ is prime.

We extend the $C_4$ contained in $G$ to a maximal (with respect to set inclusion) complete bipartite graph $H$ with parts $A$ and $B$. Observe that $|A| \geq 2$ and $|B| \geq 2$, since $H$ contains a $C_4$. We denote by $C$ the set of neighbours of $B$ outside $A$ (i.e. the set of vertices outside

$A$ each of which has at least one neighbour in $B$) and by $D$ the set of neighbours of $A$ outside $B$. Notice that

(1) *$C$ and $D$ are non-empty*, since otherwise $B$ or $A$ is a non-trivial module, contradicting the primality of $G$;

(2) *each vertex of $C$ has a non-neighbour in $B$ and each vertex of $D$ has a non-neighbour in $A$* due to the maximality of $H$.

We also claim that

(3) *$C \cup D$ induces a complete bipartite graph.* Indeed, assume there are two non-adjacent vertices $c \in C$ and $d \in D$. Consider a neighbour $b_1$ and a non-neighbour $b_2$ of $c$ in $B$, and a neighbour $a_1$ and a non-neighbour $a_2$ of $d$ in $A$. Then the six vertices $a_1, a_2, b_1, b_2, c, d$ induce an $\mathcal{A}$ in $G$, a contradiction.

(4) $V(G) = A \cup B \cup C \cup D$. To show this, assume there is a vertex $x \notin A \cup B \cup C \cup D$. Without loss of generality we may assume that $x$ is adjacent to a vertex $c \in C$ (since $G$ is prime and hence is connected). Let $d$ be any vertex of $D$, $b$ any neighbour of $c$ in $B$, and $a_1, a_2$ a neighbour and a non-neighbour of $d$ in $A$. Then the six vertices $a_1, a_2, b, c, d, x$ induce an $\mathcal{A}$ in $G$, a contradiction.

(5) *every vertex of $D$ has at most one non-neighbour in $A$.* Assume, by contradiction, that a vertex $d \in D$ has two non-neighbours $a_1, a_2$ in $A$. Since $G$ is prime, there must exist a vertex distinguishing $a_1$ and $a_2$ (otherwise $\{a_1, a_2\}$ is a non-trivial module). Let $d'$ be such a vertex. Clearly, $d'$ belongs to $D$. Finally, consider any vertex $c \in C$ and any of its neighbours $b \in B$. Then the six vertices $a_1, a_2, b, c, d, d'$ induce an $\mathcal{A}$ in $G$, a contradiction.

(6) *every vertex of $A$ has at most one non-neighbour in $D$.* Assume, to the contrary, that a vertex $a \in A$ has two non-neighbours $d_1, d_2$ in $D$. Then, by (3) and (5), $a$ is the only non-neighbour of $d_1$ and $d_2$. But then $\{d_1, d_2\}$ is a non-trivial module, contradicting the primality of $G$.

Claims (5) and (6) show that the bipartite complement of $G[A \cup D]$ is a graph of vertex degree at most 1. Moreover, in this graph at most one vertex of $A$ and at most one vertex of $D$ have degree less than 1 (since $G$ is prime). By symmetry, the bipartite complement of $G[B \cup C]$ is a graph of degree at most 1 with at most one vertex of degree 0 in each part. Therefore, $G$ can be covered by at most 4 graphs each of which admits an implicit representation (by Lemma 30). As a result, by Lemma 32, $G$ admits an implicit representation and hence the class under consideration is at most factorial. $\square$

### 6.2.2 $P_7$-free bipartite graphs

As we mentioned earlier, the class of $P_7$-free bipartite graphs was only recently proven to be a member of the factorial layer. To better understand this case, in this section we systematically study subclasses of $P_7$-free bipartite graphs obtained by forbidding one more graph.

First, we observe that all the results obtained in the previous section are applicable to $P_7$-free bipartite graphs, because these graphs are $(C_8, C_9, \ldots)$-free.

Next, we list a number of subclasses of $P_7$-free bipartite graphs for which the membership in the factorial layer is either known or easily follows from some known results. In particular, from Theorem 35 we know that $(P_7, G)$-free bipartite graphs constitute a factorial class for any graph $G \neq P_7$ such that neither $G$ nor its bipartite complement contains a cycle. Also, two more results follow readily from Lemma 36.

**Corollary 42.** *The classes of $(P_7, C_6)$-free and $(P_7, 3K_2)$-free bipartite graphs are factorial.*

*Proof.* Both classes contain $2K_2$-free bipartite graphs, which proves a lower bound. To show an upper bound, we observe that the class of $(P_7, C_6)$-free bipartite graphs coincides with $P_7$-free chordal bipartite graphs and hence is at most factorial by Lemma 36. Also, the class of $(P_7, 3K_2)$-free bipartite graphs coincides with the bipartite complements of $(P_7, C_6)$-free bipartite graphs and hence is at most factorial too. $\square$

#### $(P_7, S_{p,p})$-free bipartite graphs

By $S_{p,q}$ we denote a double star, i.e. the graph obtained from two stars $K_{1,p}$ and $K_{1,q}$ by connecting their central vertices with an edge.

**Theorem 43.** *For any $p$, the class of $(P_7, S_{p,p})$-free bipartite graphs admits an implicit representation and hence is at most factorial.*

*Proof.* Let $G$ be a $(P_7, S_{p,p})$-free bipartite graph. By Lemma 32 we assume that $G$ is connected. If $G$ does not contain $K_{p,p}$ as an induced subgraph, then $G$ can be described implicitly by Lemma 37. So suppose $G$ contains a $K_{p,p}$. If $G = K_{p,p}$, then obviously it can be described implicitly. Therefore, we assume that the set of neighbours of the $K_{p,p}$ is non-empty. We denote this set by $A$ and apply Lemma 34 (keeping in mind remarks of Section 6.1.4).

First, we show that $G[A]$ can be represented implicitly. To this end, for each edge $uv$ of the $K_{p,p}$, we denote by $H_{uv}$ the subgraph of $A$ induced by the neighbours of $u$ and the neighbours of $v$. This subgraph must be $O_{p,p}$-free, since otherwise any copy of this subgraph together with $u$ and $v$ would induce an $S_{p,p}$. Clearly, every pair of vertices of $A$ (from different parts of the bipartition) belongs to at least one subgraph $H_{uv}$ and hence the set of all these subgraphs gives a covering of $G[A]$. Also each vertex of $A$ is covered by at most $p^2$ subgraphs in the covering, because $p^2$ is the total number of such subgraphs. Finally, we observe that each $H_{uv}$ admits an implicit representation, because each of them is the bipartite complement of a $(P_7, K_{p,p})$-free bipartite graph, which admit such a representation by Lemma 37. Therefore, by Lemma 32, $G[A]$ admits an implicit representation.

Second, we show that each vertex of $A$ has at most $2p - 1$ neighbours outside of this set. Indeed, each vertex of $A$ has at most $p$ neighbours in the $K_{p,p}$. Now assume a vertex $u \in A$ has at least $p$ neighbours outside of $K_{p,p} \cup A$. Observe that $u$ must also have a neighbour $v$ in the $K_{p,p}$. But then vertices $u$ and $v$ together with the neighbours of $v$ in the $K_{p,p}$ and the $p$ neighbours of $u$ outside of $K_{p,p} \cup A$ induce an $S_{p,p}$. This contradiction shows that each vertex $u$ of $A$ has at most $p - 1$ neighbours outside of $K_{p,p} \cup A$ and hence at most $2p - 1$ neighbours outside of $A$.

Combining the two facts above, we conclude by Lemma 34 that $G$ can be represented implicitly. □

To conclude this section, we observe that the result of Theorem 43 cannot be extended to graphs of bounded chordality, because

*Remark* 1. For any $p \geq 2$, the class of $(P_8, S_{p,p})$-free bipartite graphs is superfactorial.

This conclusion follows from the fact that for any $p \geq 2$, the class of $(P_8, S_{p,p})$-free bipartite graphs contains all $\widetilde{C}_4$-free graphs.

**$(P_7, K_{p,p} + O_{0,p})$-free bipartite graphs**

By $B(p, q)$ we denote the bipartite Ramsey number, i.e. the minimum number such that every bipartite graph with at least $B(p, q)$ vertices in each of the parts contains either $K_{p,q}$ or $O_{p,q}$ as an induced subgraph.

**Lemma 39.** *For every $p \in \mathbb{N}$, any $(K_{p,p} + O_{0,p})$-free bipartite graph $G = (A, B, E)$ is either $K_{t,t}$-free or $O_{t,t}$-free, where $t = B(p, p) + p - 1$.*

*Proof.* Suppose for the contradiction that $G$ contains $K_{t,t}$ and $O_{t,t}$ as an induced subgraphs. Denote by $A_K \subseteq A$, $B_K \subseteq B$ the parts of the $K_{t,t}$ and by $A_O \subseteq A$, $B_O \subseteq B$ the parts of the $O_{t,t}$.

Obviously, either $A_K \cap A_O = \emptyset$ or $B_K \cap B_O = \emptyset$. Without loss of generality, assume that $A_K \cap A_O = \emptyset$. If $|B_K \cap B_O| \geq p$ then any $p$ vertices from $A_K$, any $p$ vertices from $B_K \cap B_O$ and any $p$ vertices from $A_0$ induce forbidden $K_{p,p} + O_{0,p}$. If $|B_K \cap B_O| < p$, then $|B_K \setminus B_O| \geq B(p, p)$ and $|B_O \setminus B_K| \geq B(p, p) > p$. Therefore, $G[B_K \setminus B_O \cup A_O]$ contains either $K_{p,p}$ or an induced $O_{p,p}$. In the former case $G[B_K \setminus B_O \cup A_O \cup B_O \setminus B_K]$ contains $K_{p,p} + O_{0,p}$ as an induced subgraph and in the latter case $G[A_K \cup B_K \setminus B_O \cup A_O]$ contains the forbidden induced subgraph. This contradiction proves the lemma. □

**Theorem 44.** *For every $p \in \mathbb{N}$, the class of $(P_7, K_{p,p} + O_{0,p})$-free bipartite graphs is at most factorial.*

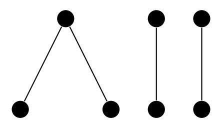
*Proof.* From Lemma 39 it follows that the class of $(P_7, K_{p,p} + O_{0,p})$-free bipartite graphs is contained in the union $Free(P_7, K_{t,t}) \cup Free(\widetilde{P}_7, O_{t,t})$, where $t = B(p, p) + p - 1$. Since $P_7 = \widetilde{P}_7$, from Lemma 37 it follows that both classes in the union are at most factorial. Therefore, the class of $(P_7, K_{p,p} + O_{0,p})$-free bipartite graphs also is at most factorial. □

By analogy with Remark 1, we conclude that

*Remark* 2. For any $p \geq 2$, the class of $(P_8, K_{p,p} + O_{0,p})$-free bipartite graphs is superfactorial.

Therefore, Theorem 44 cannot be extended to graphs of bounded chordality.

**$(P_7, K_{1,2} + 2K_2)$-free bipartite graphs**



$K_{1,2} + 2K_2$

**Lemma 40.** *Let $G = (V, E)$ be a $(P_7, K_{1,2} + 2K_2)$-free bipartite graph. Then $G$ either is $3K_2$-free or has two vertices $a, b$ such that $|N(a)\Delta N(b)| = 2$.*

*Proof.* We will show that if $G$ contains $3K_2$ as an induced subgraph, then it has two vertices $a, b$ such that $|N(a)\Delta N(b)| = 2$.

Suppose that a set of vertices $M = \{x_1, y_1, x_2, y_2, x_3, y_3\} \subseteq V$ induces a $3K_2$ such that $(x_i, y_i) \in E$ for $i = 1, 2, 3$. If some vertex $v \notin M$ has a neighbour in $A = \{x_1, x_2, x_3\}$, then it has at least two neighbours in this set, because otherwise $v, x_1, y_1, x_2, y_2, x_3, y_3$ would induce a forbidden $K_{1,2} + 2K_2$.

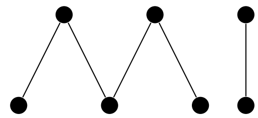
If two vertices $v$ and $w$ have exactly two neighbours in $A$, then $N(v) \cap A = N(w) \cap A$. Indeed, if say $v$ is adjacent to $x_1, x_2$ and $w$ is adjacent to $x_2, x_3$, then $y_1, x_1, v, x_2, w, x_3, y_3$ induce a $P_7$, which is impossible.

Thus each vertex outside $M$ either has no neighbours in $A$, or is adjacent to all vertices of $A$, or is adjacent to exactly two particular vertices, say $x_1, x_2$. This implies that $N(x_1)\Delta N(x_2) = \{y_1, y_2\}$. $\square$

This lemma together with Corollaries 34, 42 and remarks of Section 6.1.4 imply the following conclusion.

**Theorem 45.** *The class of $(P_7, K_{1,2} + 2K_2)$-free bipartite graphs is factorial.*

**($P_7, P_5 + K_2$)-free bipartite graphs**



$P_5 + K_2$

**Lemma 41.** *Every $2K_2$-free bipartite graph with at least three vertices has two vertices $x$ and $y$ which are in the same part and $|N(x)\Delta N(y)| \leq 1$.*

*Proof.* Let $G = (V_1, V_2, E)$ be a $2K_2$-free bipartite graph. It is known that the vertices in each of the parts can be ordered linearly with respect to inclusion of their neighbourhoods. Suppose that $|V_1| = n_1 \geq |V_2|$ and let $V_1 = \{x_1, x_2, \ldots, x_{n_1}\}$ such that $N(x_1) \supseteq N(x_2) \ldots \supseteq N(x_{n_1})$. If $|N(x_1)\Delta N(x_2)| > 1$ then there are at least two vertices in $N(x_1) \setminus N(x_2)$. All these vertices have the same neighbourhood $\{x_1\}$ and hence any two of them meet the condition of the statement. $\square$

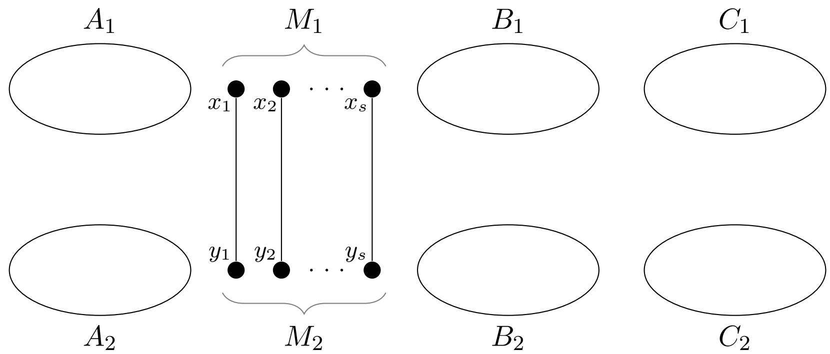
**Lemma 42.** *Let $G = (V_1, V_2, E)$ be a $\{P_7, P_5 + K_2\}$-free bipartite graph. Then $G$ either is $3K_2$-free or has two vertices $a, b$ such that $|N(a)\Delta N(b)| \leq 4$.*

*Proof.* Suppose that $G$ contains $3K_2$ as an induced subgraph. We will show that $G$ has two vertices $a, b$ such that $|N(a)\Delta N(b)| \leq 4$.

Let $\{(x_1, y_1), (x_2, y_2), \ldots, (x_s, y_s)\} \subseteq E, s \geq 3$ be a maximum induced matching in $G$ such that $M_1 = \{x_1, \ldots, x_s\} \subseteq V_1$ and $M_2 = \{y_1, \ldots, y_s\} \subseteq V_2$.

Every vertex $v$ outside $M_1(M_2)$ either has no neighbours in $M_2(M_1)$ or it is connected to all vertices from $M_2(M_1)$ or it has exactly one neighbour in $M_2(M_1)$. Indeed, if say $v \in V_1$ is adjacent to $y_i, y_j \in M_2$ and is not adjacent to $y_k \in M_2$, $i, j, k \in \{1, \ldots, s\}$, then $x_i, y_i, v, y_j, x_j, x_k, y_k$ induce a forbidden $P_5 + K_2$.

According to this observation, we denote by $A_1(A_2)$ the set of vertices which are adjacent to every vertex in $M_2(M_1)$ and by $B_1(B_2)$ the set of vertices with exactly one neighbor in $M_2$ $(M_1)$. Let $C_1 = V_1 \setminus (A_1 \cup M_1 \cup B_1)$ and $C_2 = V_2 \setminus (A_2 \cup M_2 \cup B_2)$. Let $X_i = N_{B_2}(x_i)$ and $Y_i = N_{B_1}(y_i)$ for $i = 1, \ldots, s$. From the definition it follows that $B_1(B_2)$ is the union of disjoint sets $Y_i$ $(X_i)$, $i = 1, \ldots, s$.




(1) *Let $i, j \in \{1, \ldots, s\}$ and $i \neq j$. Then no vertex in $X_i$ has a neighbour in $Y_j$.*

Assume by contradiction that $v \in X_i$ is adjacent to $u \in Y_j$. But then vertices $y_i$, $x_i$, $v$, $u$, $y_j$, $x_k$, $y_k$ with $k \in \{1, \ldots, s\}$, $k \neq i$ and $k \neq j$ induce $P_5 + K_2$.

(2) *Every vertex from $A_1(A_2)$ is adjacent to every vertex from $B_2(B_1)$.*

For the sake of definiteness, suppose that $v \in A_1$ is not adjacent to $u \in X_i \subseteq B_2$, $i \in \{1, \ldots, s\}$. But then $x_j, y_j, v, y_k, x_k, x_i, u$, where $j \neq i$, $k \neq i$ and $j, k \in \{1, \ldots, s\}$, would induce forbidden $P_5 + K_2$.

(3) *Every vertex from $C_1(C_2)$ has neighbours in at most one of the sets $X_i(Y_i)$, $i = 1, \ldots, s$.*

Assume for the contradiction that $v \in C_1$ is adjacent to $u_1 \in X_i$ and to $u_2 \in X_j$, $i \neq j$, but then $y_i, x_i, u_1, v, u_2, x_j, y_j$ would induce a forbidden $P_7$.

From (3) it follows that $C_1(C_2)$ is a union of disjoint sets $R, R_1, \ldots, R_s$ $(Q, Q_1, \ldots, Q_s)$, where $R(Q)$ is the set of vertices which have no neighbours in $B_2(B_1)$ and $R_i(Q_i)$, $i \in \{1, \ldots, s\}$, is the set of vertices which have at least one neighbour in $X_i(Y_i)$ and have no neighbour in $X_j(Y_j)$, $j \neq i$ and $j \in \{1, \ldots, s\}$.

(4) *Every vertex from $A_1(A_2)$ is adjacent to every vertex in $C_2 \setminus Q$ $(C_1 \setminus R)$*

Indeed, if say $v \in A_1$ is not adjacent to $u \in Q_i \subseteq C_2 \setminus Q$ and $w$ is a neighbour of $u$ in $Y_i$, then $x_j, y_j, v, y_k, x_k, w, u$, where $j \neq i$, $k \neq i$ and $i, j, k \in \{1, \ldots, s\}$, would induce a forbidden $P_5 + K_2$.

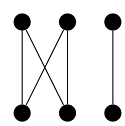
Note that for every $i \in \{1, \ldots, s\}$, $G[X_i \cup C_2 \cup Y_i \cup C_1]$ is $2K_2$-free, because otherwise $M$ would not be maximum. For the same reason there are no edges between $C_1$ and $C_2$.

We may assume that there exists $i \in \{1, \ldots, s\}$ such that $|X_i| \geq 2$ and $|Y_i| \geq 2$, otherwise there are at least $\lceil s/2 \rceil$ vertices in one of the parts $M_1, M_2$ which have at most one neighbour in $B_1 \cup B_2$ and hence for any two of these vertices $a, b$ we have $|N(a)\Delta N(b)| \leq 4$. Consider graph $G[X_i \cup Q_i \cup Y_i \cup R_i]$. As it is $2K_2$-free then by Lemma 41 it has two vertices $v, u$ which are in the same part, say $Y_i \cup R_i$, such that $|N_{X_i \cup Q_i}(v)\Delta N_{X_i \cup Q_i}(u)| \leq 1$. Note that $|N_{M_2}(v)\Delta N_{M_2}(u)| \leq 1$. Also from (2) and (4) it follows that $|N_{A_2}(v)\Delta N_{A_2}(u)| = 0$. Together with (1) it implies that $|N(v)\Delta N(u)| \leq 2$. □

This lemma together with Corollaries 34, 42 and remarks of Section 6.1.4 imply the following conclusion.

**Theorem 46.** *The class of $(P_7, P_5 + K_2)$-free bipartite graphs is factorial.*

**$(P_7, C_4 + K_2)$-free bipartite graphs**



$C_4 + K_2$

**Lemma 43.** *Let $H = (A, B, E)$ be a $(2K_2, C_4)$-free bipartite graph. Then in each part at most one vertex has degree more then 1 and all vertices with degree 1 have the same neighborhood.*

*Proof.* We prove the statement for the part $A$. For the part $B$ the same arguments are true. Let $x, y$ be some vertices from $A$. If degree of each of these vertices more then 0, then $N(x) \subseteq N(y)$ or $N(y) \subseteq N(x)$, otherwise forbidden $2K_2$ would arise. From this in particular follows that all vertices with degree 1 have the same neighborhood. Also, for the same reason, there is at most one vertex with degree more then 1 in $A$, otherwise forbidden $C_4$ would arise. □

**Lemma 44.** *Let $G = (V_1, V_2, E)$ be a $(P_7, C_4 + K_2)$-free bipartite graph. Then $G$ either is $3K_2$-free or has two vertices $a, b$ such that $|N(a)\Delta N(b)| \leq 8$.*

*Proof.* We suppose that $G$ contains $3K_2$ as an induced subgraph and show that it has two vertices $a, b$ such that $|N(a) \Delta N(b)| \leq 8$.

Let $\{(x_1, y_1), (x_2, y_2), \ldots, (x_s, y_s)\} \subseteq E, s \geq 3$ be a maximum induced matching in $G$ such that $M_1 = \{x_1, \ldots, x_s\} \subseteq V_1$ and $M_2 = \{y_1, \ldots, y_s\} \subseteq V_2$. Denote by $A_1$ ($A_2$) the set of vertices which are adjacent to every vertex in $M_2$ ($M_1$) and by $B_1$ ($B_2$) all other vertices which have neighbors in $M_2$ ($M_1$). Let $C_1 = V_1 \setminus (A_1 \cup M_1 \cup B_1)$ and $C_2 = V_2 \setminus (A_2 \cup M_2 \cup B_2)$. Then

(1) *for any two vertices $v, u \in B_1$ ($B_2$), either $N_{M_2}(v) \cap N_{M_2}(u) = \emptyset$ or $N_{M_2}(v) \subseteq N_{M_2}(u)$ or $N_{M_2}(u) \subseteq N_{M_2}(v)$ (either $N_{M_1}(v) \cap N_{M_1}(u) = \emptyset$ or $N_{M_1}(v) \subseteq N_{M_1}(u)$ or $N_{M_1}(u) \subseteq N_{M_1}(v)$).*

Suppose for the contradiction that there are $v, u$ in $B_1$ and $y_i, y_j, y_k$ in $M_2$ such that $(v, y_i) \in E$, $(v, y_k) \in E$, $(v, y_j) \notin E$, $(u, y_j) \in E$, $(u, y_k) \in E$ and $(u, y_i) \notin E$. But then $x_i, y_i, v, y_k, u, y_j, x_j$ induce a forbidden $P_7$.

(2) *for any two vertices $y_i, y_j \in M_2$, $|N_{B_1}(y_i) \cap N_{B_1}(y_j)| \leq 1$ (for any two vertices $x_i, x_j \in M_1$, $|N_{B_2}(x_i) \cap N_{B_2}(x_j)| \leq 1$).*

Suppose for the contradiction that there are $y_i, y_j$ in $M_2$ such that there are two different vertices $v$ and $u$ in $N_{B_1}(y_i) \cap N_{B_1}(y_j)$. From (1) it follows that $N_{M_2}(v) \subseteq N_{M_2}(u)$ or $N_{M_2}(u) \subseteq N_{M_2}(v)$. Without loss of generality, assume that $N_{M_2}(v) \subseteq N_{M_2}(u)$. By definition of $B_1$, there is vertex $y_k$ in $M_2$ which is not adjacent to $u$ and hence is not adjacent to $v$. Therefore $x_k, y_k, v, u, y_i, y_j$ induce a forbidden $C_4 + K_2$.

(3) *for any vertex $y_i \in M_2$, at most one vertex from $N_{B_1}(y_i)$ has neighbor in $M_2$ different from $y_i$.*

Suppose that there are two vertices $v$ and $u$ in $N_{B_1}(y_i)$ such that there are $y_j$ and $y_k$ in $M_2$ different from $y_i$ and $(v, y_j) \in E$ and $(u, y_k) \in E$. We have $(v, y_k) \notin E$ and $(u, y_j) \notin E$, otherwise $y_i$ and $y_k$ or $y_i$ and $y_j$ have more then one common neighbor in $B_1$, which contradicts (2). But then $x_j, y_j, v, y_i, u, y_k, x_k$ induce a forbidden $P_7$.

(4) *Let $v \in B_1$ be adjacent to exactly one vertex from $M_2$, say $y_i$. Then $v$ is adjacent to at most two vertices from $B_2$.*

Suppose by contradiction that $v$ has three neighbors $c, d, e$ in $B_2$. Assume $c$ is adjacent to $x_i$. Remember that $c$ must also have a non-neighbour $x_j \in M_1$. But then vertices $c, x_i, y_i, v$ together with $x_j, y_j$ induce a forbidden $C_4+K_2$. This contradiction shows that $c$ is not adjacent to $x_i$. Similarly, $d$ and $e$ are not adjacent to $x_i$.

Let $L_1 = M_1 \setminus \{x_i\}$. We claim that the graph $G[L_1 \cup \{c, d, e\}]$ is $(2K_2, C_4)$-free. Indeed, if, say, $c, x_j, d, x_k$ induce a $2K_2$, then $y_j, x_j, c, v, d, x_k, y_k$ induce a forbidden $P_7$, and if, say, $c, x_j, d, x_k$ induce a $C_4$, then $c, x_j, d, x_k$ together with $x_i, y_i$ induce a forbidden $C_4 + K_2$.

From Lemma 43 it follows that there are two vertices in $\{c, d, e\}$ say $c, d$ which have the same neighborhood in $L_1$ consisting of exactly one vertex, say $x_k$. But then $v, c, d, x_k, x_j, y_j$, where $x_j \in L_1$ and $x_j \neq x_k$, induce a forbidden $C_4 + K_2$.

(5) *Let $v \in B_1$ be adjacent to exactly one vertex from $M_2$, say $y_i$. Then $v$ is adjacent to all but at most one vertex from $A_2$.*

Suppose for the contradiction that $u, w \in A_2$ are not adjacent to $v$. But then $y_i, v, u, w, x_j, x_k$, where $j, k$ are different from $i$, induce forbidden $C_4 + K_2$.

(6) *Let $R_i \subseteq B_1$ be the set of vertices which are adjacent only to $y_i$ in $M_2$. Then $G[R_i \cup C_2]$ is a $(2K_2, C_4)$-free graph.*

If $G[R_i \cup C_2]$ contains an induced $2K_2$, then this contradicts to the maximality of the matching $\{(x_1, y_1), (x_2, y_2), \ldots, (x_s, y_s)\}$. If $G[R_i \cup C_2]$ contains a $C_4$ then $(x_k, y_k)$, where $k \neq i$, together with the $C_4$ constitute a forbidden $C_4 + K_2$.

(7) *$G$ has two vertices $a, b$ such that $|N(a) \Delta N(b)| \leq 8$.*

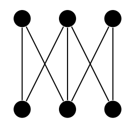
If $M_2$ has no vertices with more than 3 neighbors in $B_1$, then we can take as $a$ and $b$ any two vertices from $M_2$. Otherwise, if there is vertex $y_i$ in $M_2$ which has at least 4 neighbors in $B_1$, then by (3) $N_{B_1}(y_i)$ contains three vertices, say $c, d, e$, which are adjacent only to $y_i$ in $M_2$. From (6) and Lemma 43 it follows that for two of these vertices, say $c, d$, $|N_{C_2}(c) \Delta N_{C_2}(d)| \leq 1$. From (5) it follows that $|N_{A_2}(c) \Delta N_{A_2}(d)| \leq 2$ and from (4) it follows that $|N_{B_2}(c) \Delta N_{B_2}(d)| \leq 4$. Therefore $|N(c) \Delta N(d)| \leq 7$ and we can take $c$ and $d$ as $a$ and $b$.

$\square$

This lemma together with Corollaries 34, 42 and remarks of Section 6.1.4 imply the following conclusion.

**Theorem 47.** *The class of* $(P_7, C_4 + K_2)$*-free bipartite graphs is factorial.*

**($P_7$, *domino*)-free bipartite graphs**



*domino*

**Theorem 48.** *The class of ($P_7$, domino)-free bipartite graphs is factorial.*

*Proof.* This class contains all $2K_2$-free bipartite graphs (one of the three minimal factorial classes of bipartite graphs) and hence is at least factorial. Now let us show an upper bound.

Since $(P_7, C_4)$-free bipartite graphs is at most factorial (Lemma 37), we consider a ($P_7$, *domino*)-free bipartite graph containing a $C_4$. We extend this $C_4$ to a maximal biclique in $G$ with parts denoted by $A$ and $B$ of size at least 2. Then we define $C = N(B)\backslash A$, $D = N(A)\backslash B$, $E = N(D)\backslash(A \cup C)$, $F = N(C)\backslash(B \cup D)$, $I = N(F)\backslash(A \cup C \cup E)$, $J = N(E)\backslash(B \cup D \cup F)$. Now we prove a series of claims.

(1) *The set* $C \cup D$ *is independent.* Suppose, by contradiction, there is an edge $cd$ with $c \in C$ and $d \in D$. By definition $c$ must have a neighbour $b_1$ in $B$. Also, as $G[A \cup B]$ is a maximal biclique, $c$ has a non-neighbour $b_0$ in $B$. Similarly $d$ has a neighbour $a_1$ and a non-neighbour $a_0$ in $A$. But then $G$ contains a *domino* induced by $c, d, a_0, a_1, b_0, b_1$.

(2) *The subgraph induced by* $A \cup D$ *does not contain one-sided copy of a* $P_5$ *with 3 vertices in* $A$*, and hence it is* $(P_6, C_6)$*-free.* Assume $G[A \cup D]$ contains a one-sided copy of a $P_5$ with 3 vertices in $A$. Then this copy together with any vertex $b \in B$ induces a *domino*, a contradiction.

(3) By symmetry, the subgraph induced by $B \cup C$ is $(P_6, C_6)$-free.

(4) *The set $F \cup E$ is independent.* Suppose, by contradiction, there is an edge $ef$ with $e \in E$ and $f \in F$. Then pick a neighbour $c \in C$ of $f$ and a neighbour $d \in D$ of $e$, which must exist by definition of $E$ and $F$. Let $b \in B$ be a non-neighbour of $c$, and let $a_1 \in A$ and $a_2 \in A$ be a neighbour and a non-neighbour of $d$. Then vertices $a_1, b, a_2, d, e, f, c$ induce a $P_7$, a contradiction.

(5) *The subgraph induced by $F \cup C$ is $C_6$-free.* Suppose by contradiction there is a $C_6$ induced by $c_1, f_1, c_2, f_2, c_3, f_3$ with $c_1, c_2, c_3 \in C$ and $f_1, f_2, f_3 \in F$. If there is a vertex $b \in B$ adjacent to all $c_1, c_2, c_3$, then $G$ contains a *domino* induced by $b, c_1, f_1, c_2, f_2, c_3$. Also, if there is a vertex $b \in B$ adjacent to exactly one of $c_1, c_2, c_3$, say $c_1$, then together with any vertex $a \in A$ we have a $P_7$ induced by $a, b, c_1, f_1, c_2, f_2, c_3$. If there is a vertex $b_1 \in B$ not adjacent to any of $c_1, c_2, c_3$, then take a vertex $b_2 \in B$ adjacent to 2 of them, say $c_1$ and $c_3$, and together with any vertex $a \in A$ form a $P_7$ induced by $b_1, a, b_2, c_1, f_1, c_2, f_2$. Therefore, each vertex of $B$ is adjacent to exactly two of $c_1, c_2, c_3$ and since all vertices in $C$ have a non-neighbour in $B$, each pair must appear. So, pick vertex $b_1$ adjacent to $c_1$ and $c_2$, pick vertex $b_2$ adjacent to $c_2$ and $c_3$ and pick vertex $b_3$ adjacent to $c_3$ and $c_1$. Now $G[\{c_1, b_1, c_2, b_2, c_3, b_3\}]$ is a $C_6$, contradicting our claim 3. This contradiction shows that the graph $G[F \cup C]$ is $C_6$-free.

(6) By symmetry, the subgraph induced by $E \cup D$ is $C_6$-free.

(7) *The set $I \cup J$ is independent.* If not, assume $ij$ is an edge with $i \in I$ and $j \in J$. Then take a neighbour $e \in E$ of $j$, a neighbour $d \in D$ of $e$, a neighbour $a_1$ and a non-neighbour $a_2$ of $d$ in $A$ and an arbitrary vertex $b$ in $B$. Then $i, j, e, d, a_1, b, a_2$ induce a $P_7$, a contradiction.

(8) *The subgraph induced by $J \cup E$ is $C_6$-free.* Suppose by contradiction there is a $C_6$ induced by $e_1, j_1, e_2, j_2, e_3, j_3$ with $e_1, e_2, e_3 \in E$ and $j_1, j_2, j_3 \in J$. Now if there is a vertex $d \in D$ joined to all $e_1, e_2, e_3$, then $G$ contains a *domino* induced by $d, e_1, j_1, e_2, j_2, e_3$. Otherwise, there is a vertex $d \in D$ having a neighbour and a non-neighbour in the set $\{e_1, e_2, e_3\}$, say $d$ is non-adjacent to $e_1$ and adjacent to $e_2$. Then pick a neighbour $a_1 \in A$ of $d$ and a non-neighbour $a_2 \in A$ of $d$. Pick arbitrary

$b \in B$. Then $e_1, j_1, e_2, d, a_1, b, a_2$ induce a $P_7$ in $G$, a contradiction.

(9) By symmetry, the subgraph induced by $I \cup F$ is $C_6$-free.

(10) *If $G$ is connected, then $V(G) = A \cup B \cup C \cup D \cup E \cup F \cup I \cup J$.* Suppose that if $N(J) \backslash E \neq \emptyset$ and take an edge $jk$ with $j \in J$ and $k \in N(J) \backslash E$. Then take a neighbour $e \in E$ of $j$, a neighbour $d \in D$ of $e$, a neighbour $a_1$ and a non-neighbour $a_2$ of $d$ in $A$ and an arbitrary vertex $b$ in $B$. Then $i, j, e, d, a_1, b, a_2$ induce a $P_7$. This contradiction shows that $N(J) \backslash E = \emptyset$. By symmetry, $N(I) \backslash F = \emptyset$. Hence the claim.

This series of claim shows that every connected $(P_7, domino)$-free bipartite graph can be covered by finitely many graphs each coming from a class which is at most factorial. By Lemma 31 this implies that the class of $(P_7, domino)$-free bipartite graphs is at most factorial. □

**$(P_7, K_{3,3}$-$e)$-free bipartite graphs**

The graph $K_{3,3}$-$e$ is obtained from $K_{3,3}$ by deleting an edge.

**Theorem 49.** *The class of $(P_7, K_{3,3}$-$e)$-free bipartite graphs is factorial.*

*Proof.* The class of $(P_7, K_{3,3}$-$e)$-free bipartite graphs contains all graphs of degree at most one (one of the three minimal factorial classes of bipartite graphs) and hence is at least factorial. Now we show an upper bound. In the proof we follow the structure and notation of the proof of Theorem 48. In particular, we assume that a connected $(P_7, K_{3,3}$-e)-free bipartite graph $G$ contains a $K_{4,4}$ and denote by $A, B, C, D, E, F, I, J$ the subsets defined in the proof of Theorem 48. From this theorem we know that these subsets partition the vertex set of $G$ and that $I \cup J$ is an independent set, since otherwise an induced $P_7$ arises. Also, by definition, the subgraph of $G$ induced by $A \cup B$ is complete bipartite with at least 4 vertices in each part. Now we derive a number of claims as follows.

(1) *Every vertex outside of $A \cup B$ has at most one neighbour in $A \cup B$.* To show this, consider a vertex $x \notin A \cup B$ which has at least two neighbours in $A$. By definition, $G[A \cup B]$ is a maximal biclique and hence $x$ also has a non-neighbour in $A$. But then

a non-neighbour and two neighbours of $x$ in $A$ together with $x$ and any two vertices of $B$ induce a $K_{3,3}$-$e$.

(2) *The subgraph induced by $C \cup D \cup F \cup J$ is domino-free.* To prove this, assume by contradiction that this subgraph contains a *domino* induced by vertices $x_1, x_2, x_3 \in C$ and $y_1, y_2, y_3$ with $x_2$ and $y_2$ being the vertices of degree 3 in the induced *domino*. By definition, $x_1$ has a neighbour $z$ in $B$. Also, since vertices $x_1, x_2$ and $x_3$ have collectively at most 3 neighbours in $B$ and the size of $B$ is at least 4, there must exist a vertex $b \in B$ adjacent to none of $x_1, x_2, x_3$. For the same reason, one can find a vertex $a$ in $A$ which is adjacent to none of $y_1, y_2, y_3$. Now if $z$ is not adjacent to $x_2$, then vertices $b, a, z, x_1, y_1, x_2, y_3$ induce a $P_7$ in $G$, and if $z$ is not adjacent to $x_3$, then vertices $b, a, z, x_1, y_2, x_3, y_3$ induce a $P_7$ in $G$, and if $z$ is adjacent to both $x_2$ and $x_3$, then vertices $z, x_1, x_2, x_3, y_2, y_3$ induce a $K_{3,3}$-$e$ in $G$. A contradiction in all possible cases proves the claim.

(3) *The subgraph induced by $E \cup F \cup J$ is domino-free.* This can be proved by analogy with (2). We assume by contradiction that this subgraph contains a *domino* induced by vertices $x_1, x_2, x_3 \in E$ and $y_1, y_2, y_3$ and consider a neighbour $z$ of $x_1$ in $D$, a neighbour $a$ of $z$ in $A$ and an arbitrary vertex $b$ in $B$. Then the very same arguments as in (2) lead to a contradiction.

By symmetry we conclude that the subgraphs of $G$ induced by $D \cup C \cup E \cup I$ and by $F \cup E \cup I$ are *domino*-free. Therefore, $G$ can be covered by finitely many *domino*-free graphs. Together with Lemma 31 and Theorem 48 this completes the proof. $\square$

**Bipartite complements of $P_7$-free bipartite graphs**

Since the bipartite complement of $P_7$ is again $P_7$, from the preceding sections we derive the following conclusion.

**Theorem 50.** *Let $H$ be the bipartite complement of any of the following graphs: $Q(p)$, $L(s,p) + O_{0,1}$, $M(p)$, $N(p)$, $\mathcal{A}$, $S_{p,p}$, $K_{p,p} + O_{p,p}$, $K_{1,2} + 2K_2$, $P_5 + K_2$, $C_4 + K_2$, domino, $K_{3,3}$-$e$. The class of $(P_7, H)$-free bipartite graphs is at most factorial.*

This theorem together with the results of the preceding sections implies, in particular, that for any graph $H$ with at most 6 vertices the class of $(P_7, H)$-free bipartite graphs is at most factorial.

# Chapter 7

# Hereditary properties of graphs of high speed

According to the Alekseev-Bollobás-Thomason Theorem, in the family of hereditary classes there are precisely three minimal classes with non-zero entropy: bipartite, co-bipartite and split graphs. Therefore, by Theorem 13, a hereditary class $X$ has zero entropy if and only if VC-dimension is bounded for graphs in $X$. In this chapter, we study some hereditary properties of graphs with non-zero entropy.

## 7.1 Estimating the entropy of some hereditary graph classes

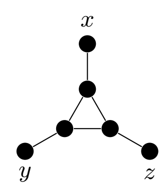
In our effort to estimate the entropy of some hereditary graph classes, we will start with a few simpler ones and then end the section with a discussion on word-representable graphs.

### 7.1.1 Perfect graphs

According to the Strong Perfect Graph Theorem [19] the class of perfect graphs is precisely the class of graphs containing no odd cycles of length at least 5 and no complements of these cycles. Since perfect graphs contain all bipartite graphs, the index of perfect graphs is at least 2. For the upper bound, we observe that $C_5$ belongs to all four minimal classes of index 3, i.e. to $\mathcal{E}_{3,0}$, $\mathcal{E}_{2,1}$, $\mathcal{E}_{1,2}$ and $\mathcal{E}_{0,3}$. Therefore, perfect graphs contain none of these classes and hence the index of perfect graphs is precisely 2.

### 7.1.2 AT-free graphs

Three vertices in a graph form an asteroidal triple if every two of them are connected by a path which avoids the neighbourhood of the third. A graph is called AT-free if it does not contain an asteroidal triple. It is easy to see that the set of AT-free graphs forms a hereditary graph class. The index of the class of AT-free graphs is at least 2 because it contains the class $\mathcal{E}_{0,2}$, the class of complements of bipartite graphs. To see this, note that given a graph in $\mathcal{E}_{0,2}$ any choice of three vertices will have at least two vertices in the same part and hence they belong in the neighbourhood of each other. To obtain an upper bound on the index consider the graph $G$ pictured below. This graph contains an asteroidal triple, indicated by the vertices $x$, $y$ and $z$, so it does not belong to the class of AT-free graphs. Now $G$ belongs to all four minimal classes of index 3, therefore similar to perfect graphs AT-free graphs contain none of these classes and hence has index precisely 2.




$G$

### 7.1.3 $K_n$-free graphs

The class $\mathcal{E}_{i,0}$ is a subclass of $K_n$-free graphs for each $1 \leq i \leq n-1$. However, the graph $K_n$ belongs to each class $\mathcal{E}_{i,j}$ with $i+j=n$ by taking each vertex to be either an independent set or clique of size one. Therefore the class of $K_n$-free graphs has index precisely $n-1$.

### 7.1.4 $C_n$-free graphs

The class $\mathcal{E}_{0,\lceil \frac{n}{2} \rceil -1}$ is a subclass of $C_n$-free graphs for any $n > 3$. This can easily be seen by noting that if $C_n \in \mathcal{E}_{0,\lceil \frac{n}{2} \rceil -1}$ then each part contains at most 2 vertices as $C_n$ is triangle free, which leads to a contradiction. After a little thought, we can check that the graph $C_n$ belongs to each class $\mathcal{E}_{i,j}$ with $i+j=\lceil \frac{n}{2} \rceil$. Therefore the class of $C_n$-free graphs has index precisely $\lceil \frac{n}{2} \rceil - 1$.

## 7.2 Word-representable graphs

We say $G$ is *word-representable* if there exists a word $w$ over the alphabet $V$ such that letters $x$ and $y$ alternate in $w$ if and only if $(x, y) \in E$ for each $x \neq y$.

The notion of word-representable graphs has its roots in the study of the celebrated Perkins semigroup [39, 63]. These graphs possess many attractive properties (e.g. a maximum clique in such graphs can be found in polynomial time), and they provide a common generalization of several important graph families, such as circle graphs, comparability graphs, 3-colorable graphs, graphs of vertex degree at most 3 (see [17] for definitions of these families).

Recently, a number of fundamental results on word-representable graphs were obtained in the literature [35, 36, 40, 41, 42]. In particular, Halldórsson et al. [36] have shown that a graph is word-representable if and only if it admits a *semi-transitive orientation*. However, our knowledge on these graphs is still very limited and many important questions remain open. For example, how hard is it to decide whether a given graph is word-representable or not? What is the minimum length of a word that represents a given graph? How many word-representable graphs on $n$ vertices are there? Does this family include all graphs of vertex degree at most 4?

The last question was originally asked in [36]. In this section we answer this question negatively by exhibiting a graph of vertex degree at most 4 which is not word-representable. This result allows us to obtain an upper bound on the asymptotic growth of the number of $n$-vertex word-representable graphs. Combining this result with a lower bound that follows from some previously known facts, we conclude that the number of $n$-vertex word-representable graphs is $2^{\frac{n^2}{3}+o(n^2)}$.

All preliminary information related to the notion of word-representable graphs can be found in Section 7.2.1. In Section 7.2.2, we prove our negative result about graphs of degree at most 4 and in Section 7.2.3, we derive the asymptotic formula on the number of word-representable graphs.

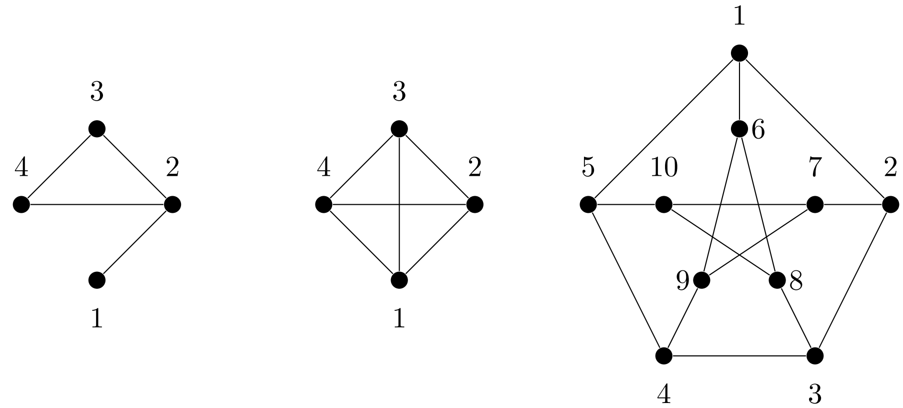



Figure 7.1: Three word-representable graphs $M$ (left), the complete graph $K_4$ (middle), and the Petersen graph (right)

### 7.2.1 Word-representable graphs: definition, examples and related results

Distinct letters $x$ and $y$ *alternate* in a word $w$ if the deletion of all other letters from the word results in either $xyxy\cdots$ or $yxyx\cdots$. A graph $G = (V, E)$ is *word-representable* if there exists a word $w$ over the alphabet $V$ such that letters $x$ and $y$ alternate in $w$ if and only if $(x, y) \in E$ for each $x \neq y$. For example, the graph $M$ in Figure 7.1 is word-representable, because the word $w = 1213423$ has the right alternating properties, i.e. the only non-alternating pairs in this word are 1,3 and 1,4 that correspond to the only non-adjacent pairs of vertices in the graph.

If a graph is word-representable, then there are infinitely many words representing it. For instance, the complete graph $K_4$ in Figure 7.1 can be represented by words 1234, 3142, 123412, 12341234, 432143214321, etc. In general, to represent a complete graph on $n$ vertices, one can start with writing up any permutation of length $n$ and adjoining from either side any number of copies of this permutation.

If each letter appears exactly $k$ times in a word representing a graph, the graph is said to be *$k$-word-representable*. It is known [40] that any word-representable graph is $k$-word-representable for some $k$. For example, a 3-representation of the Petersen graph shown in Figure 7.1 is

$$1387296(10)7493541283(10)7685(10)194562.$$

It is not difficult to see that a graph is 1-representable if and only if it is complete. Also, with a bit of work one can show that a graph is 2-representable if and only if it is a circle graph, i.e. the intersection graph of chords in a circle. Thus, word-representable graphs generalize both complete graphs and circle graphs. They also generalize two other important graph families, comparability graphs and 3-colorable graphs. This can be shown through the notion of semi-transitive orientation.

A directed graph (digraph) $G = (V, E)$ is *semi-transitive* if it has no directed cycles and for any directed path $v_1v_2 \cdots v_k$ with $k \geq 4$ and $v_i \in V$, either $v_1v_k \notin E$ or $v_iv_j \in E$ for all $1 \leq i < j \leq k$. In the second case, when $v_1v_k \in E$, we say that $v_1v_k$ is a shortcut. The importance of this notion is due to the following result proved in [36].

**Theorem 51** ([36]). *A graph is word-representable if and only if it admits a semi-transitive orientation.*

From this theorem and the definition of semi-transitivity it follows that all comparability (i.e. transitively orientable) graphs are word-representable. Moreover, the theorem implies two more important corollaries.

**Theorem 52** ([36]). *All* 3*-colorable graphs are word-representable.*

*Proof.* Partitioning a 3-colorable graph in three independent sets, say I, II and III, and orienting all edges in the graph so that they are oriented from I to II and III, and from II to III, we obtain a semi-transitive orientation. □

**Theorem 53** ([36]). *All graphs of vertex degree at most* 3 *are word-representable.*

*Proof.* By Brooks' Theorem, every connected graph of vertex degree at most 3, except for the complete graph $K_4$, is 3-colorable, and hence word-representable by Theorem 52. Also, as we observed earlier, all complete graphs are word-representable. Therefore, all connected graphs of degree at most 3 and hence all graphs of degree at most 3 are word-representable. □

Whether all graphs of degree at most 4 are word-representable is a natural question following from Theorem 53, which was originally asked in [36]. In the next section, we settle this question negatively.

### 7.2.2 A non-representable graph of vertex degree at most 4

The main result of this section is that the graph $A$ represented in Figure 7.2 is not word-representable. To prove this, we will show that this graph does not admit a semi-transitive orientation. This is a sufficient condition due to Theorem 51.

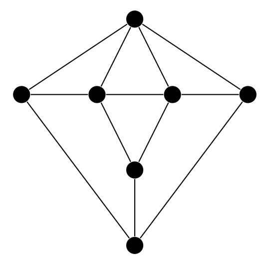


Figure 7.2: The graph $A$

Our proof is a case analysis and the following lemma will be used frequently in the proof.

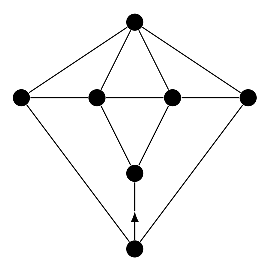
**Lemma 45.** *Let $D$ be a $K_4$-free graph admitting a semi-transitive orientation. Then no cycle of length 4 in this orientation has three consecutively oriented edges.*

*Proof.* If a semi-transitive orientation of $D$ contains a cycle of length four with three consecutively oriented edges, then the fourth edge has to be oriented in the opposite direction to avoid an oriented cycle. However, the fourth edge now creates a shortcut. Hence the cycle must contain both chords to make it transitive, which is impossible because $D$ is $K_4$-free. □

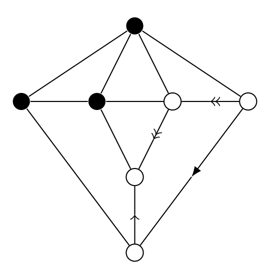
**Theorem 54.** *The graph $A$ does not admit a semi-transitive orientation.*

*Proof.* In order to prove that $A$ does not admit a semi-transitive orientation, we will explore *all* orientations of this graph and will show that each choice leads to a contradiction. At each step of the proof we choose a vertex and split the analysis into two cases depending on the orientation of the chosen edge. The chosen edge and its orientation will be shown by a solid arrow (⟶). This choice of orientation may lead to other edges having an orientation assigned to them to satisfy Lemma 45. The orientations forced by the solid arrow through an application of Lemma 45 will be shown by means of double arrows (⟶»). When we use Lemma 45 to derive a double arrow, we always apply it with respect to a particular cycle of length 4. This cycle will be indicated by four white vertices. Since the graph $A$ has many cycles of length 4, repeated use of Lemma 45 applied to different cycles may lead to a contradiction, where one more cycle of length 4 has three consecutively oriented edges. We will show that in all possible cases a contradiction of this type arises. The proof is illustrated by diagrams.

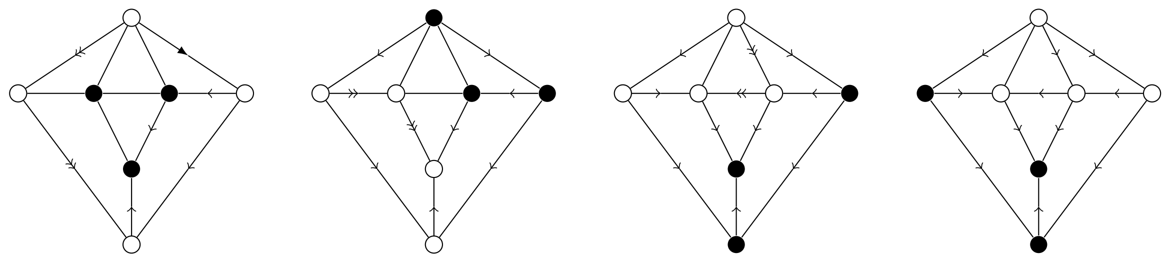
*Case* 1: We start by choosing the orientation for the edge indicated in the diagram below.



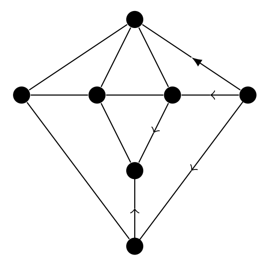
*Case* 1.1: Now we orient one more edge (solid arrow), which leads to two more orientations being assigned due to Lemma 1 (double arrows).



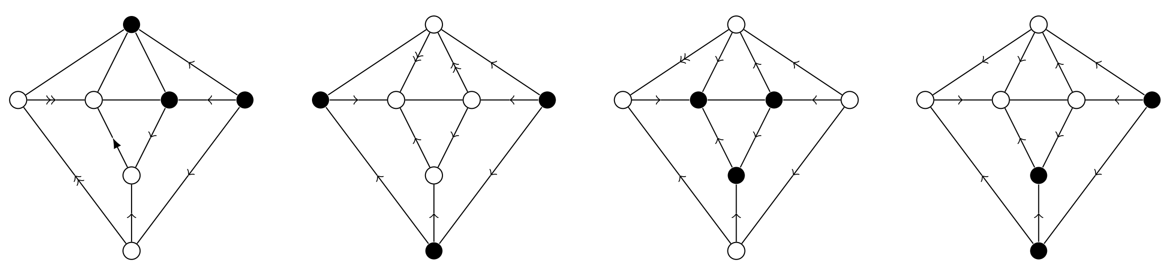
*Case* 1.1.1: One more edge is oriented (solid arrow) and this choice leads to a contradiction through repeated use of Lemma 1 (a cycle of four white vertices with three consecutively oriented edges in the final of the four diagrams below).



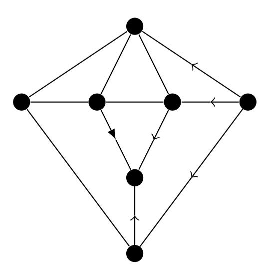
*Case* 1.1.2: In this case the orientation of the edge chosen in Case 1.1.1 is reversed.



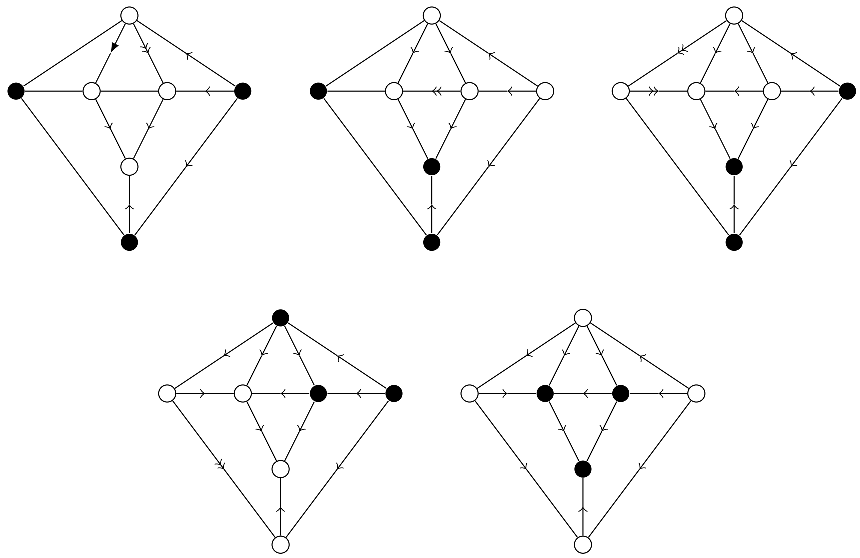
*Case* 1.1.2.1: By orienting one more edge (solid arrow) we obtain a contradiction through repeated use of Lemma 1 (a cycle of four white vertices with three consecutively oriented edges in the final diagram).



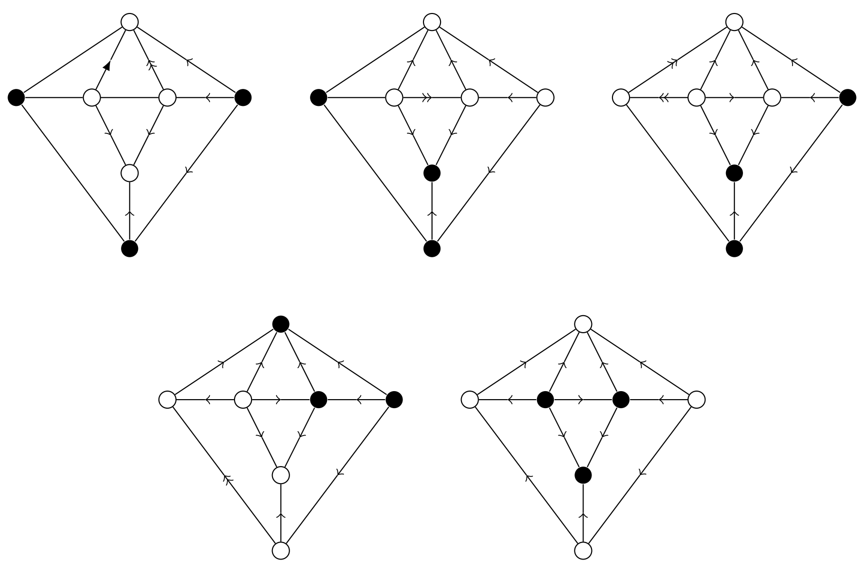
*Case* 1.1.2.2: Now we reverse the orientation of the edge chosen in Case 1.1.2.1.



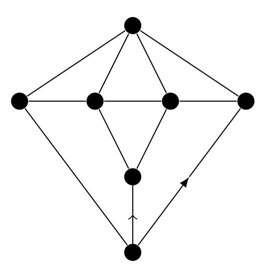
*Case* 1.1.2.2.1: By orienting one more edge (solid arrow) we obtain a contradiction through repeated use of Lemma 1.



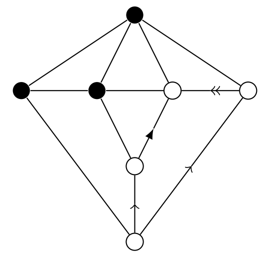
*Case* 1.1.2.2.2: By reversing the orientation of the edge chosen in Case 1.1.2.2.1 we obtain a contradiction again. This completes the analysis of Case 1.1.



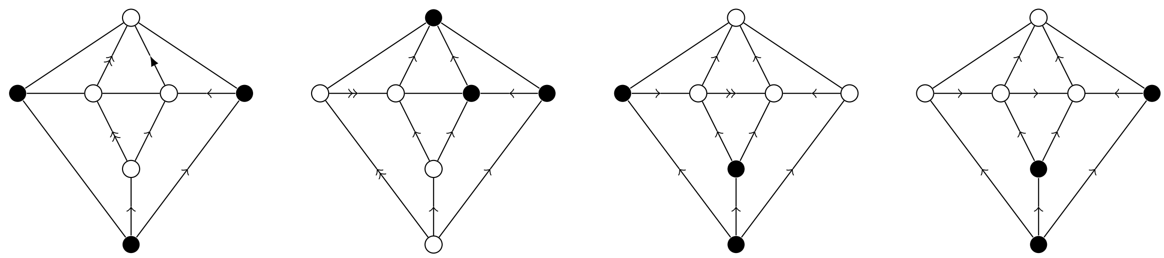
*Case* 1.2: The orientation of the edge chosen in Case 1.1 is reversed.



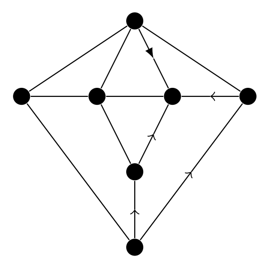
*Case* 1.2.1: We orient one more edge (solid arrow) and derive one consequence (double arrow).



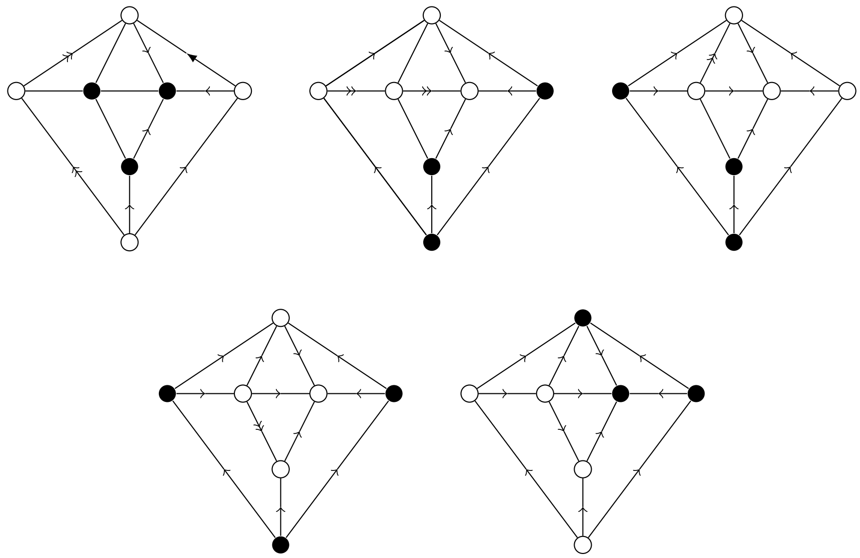
*Case* 1.2.1.1: One more edge is oriented (solid arrow) leading to a contradiction.



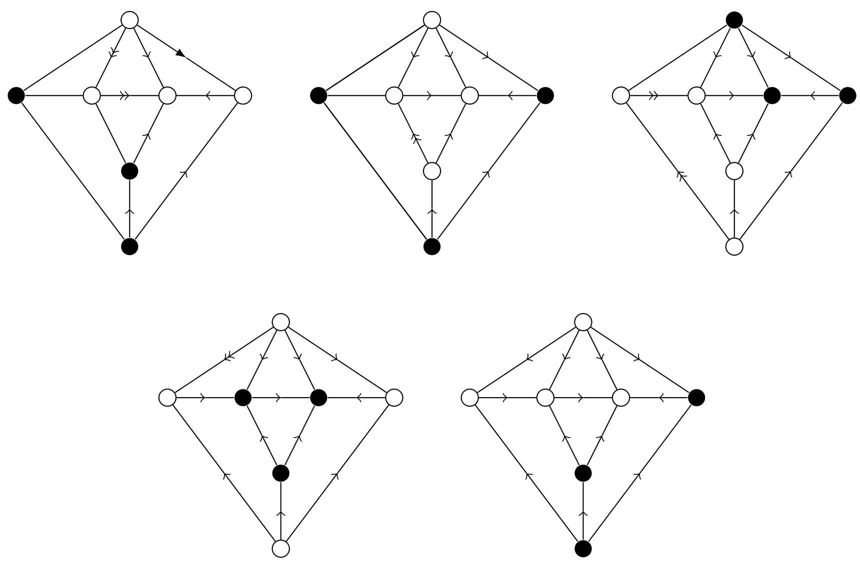
*Case* 1.2.1.2: The orientation of the edge chosen in Case 1.2.1.1 is reversed.



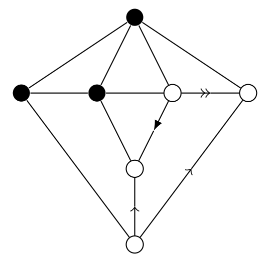
*Case* 1.2.1.2.1: One more edge is oriented (solid arrow) leading to a contradiction.



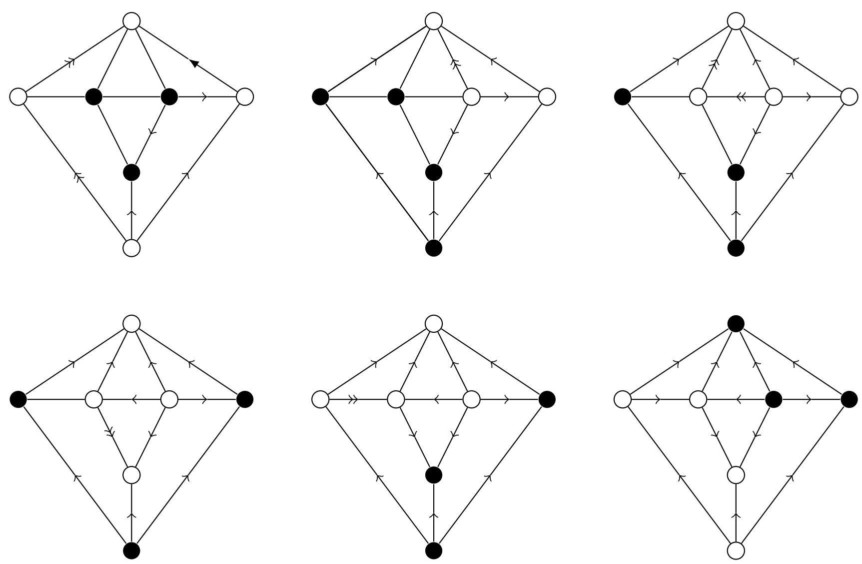
*Case* 1.2.1.2.2: The orientation of the edge chosen in Case 1.2.1.2.1 is reversed leading a contradiction again. This completes the analysis of Case 1.2.1.



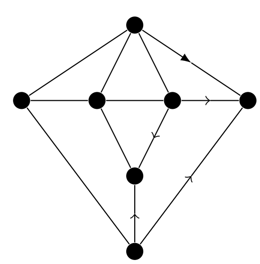
*Case* 1.2.2: The orientation of the edge chosen in Case 1.2.1 is reversed (solid arrow) and one consequence is derived (double arrow).



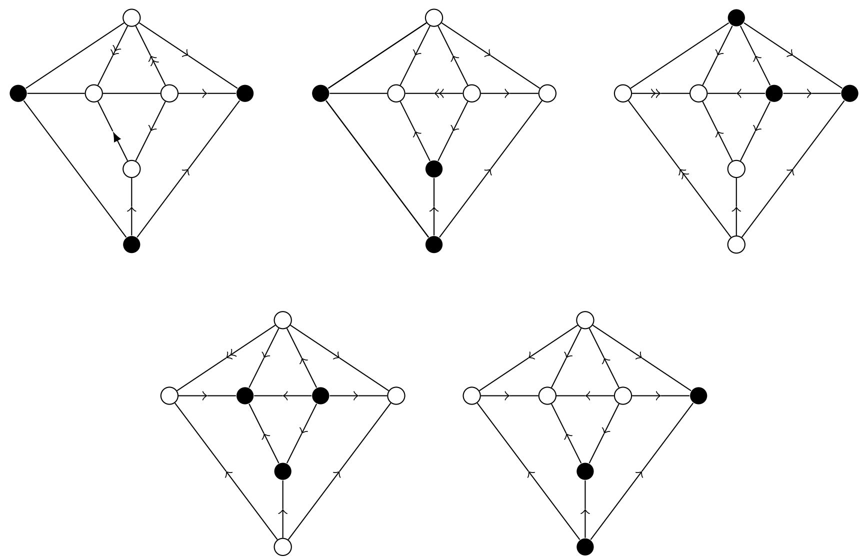
*Case* 1.2.2.1: One more edge is oriented (solid arrow) leading to a contradiction.



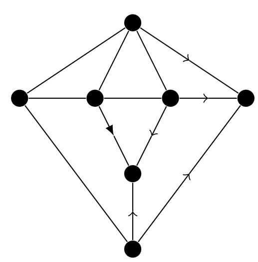
*Case* 1.2.2.2: The orientation of the edge chosen in Case 1.2.2.1 is reversed.



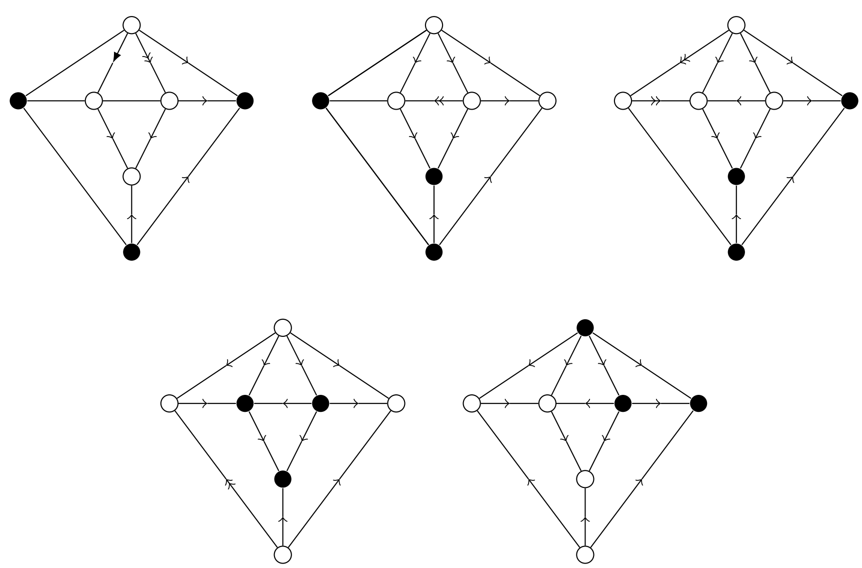
*Case* 1.2.2.2.1: One more edge is oriented (solid arrow) leading to a contradiction.



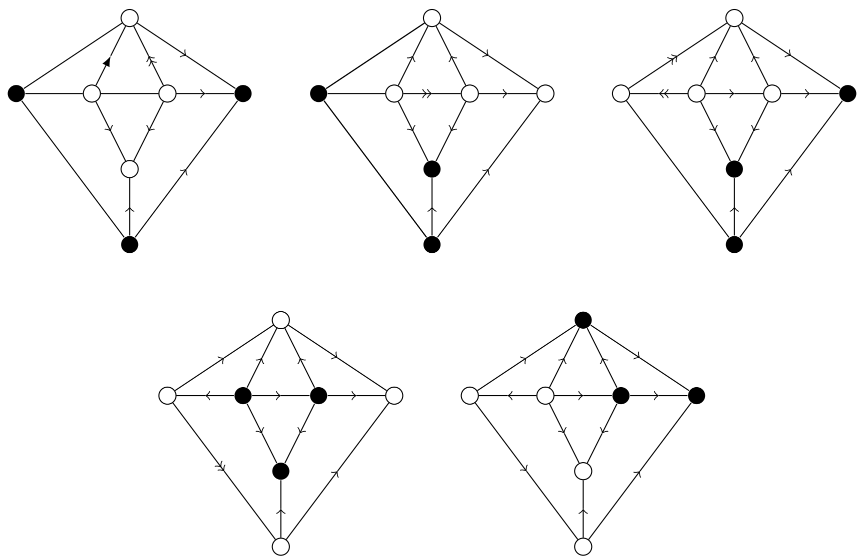
*Case* 1.2.2.2.2: The orientation of the edge chosen in Case 1.2.2.2.1 is reversed.



*Case* 1.2.2.2.2.1: One more edge is oriented (solid arrow) leading to a contradiction.



*Case* 1.2.2.2.2.2: The orientation of the edge chosen in Case 1.2.2.2.2.1 is reversed, which leads to a contradiction again. This completes the analysis of Case 1.



*Case* 2. In this case, we reverse the orientation of the edge chosen in Case 1 and complete the proof by symmetry, i.e. by reversing the orientations obtained in Case 1. □

### 7.2.3 Asymptotic enumeration of word-representable graphs

Clearly, if $G$ is a word-representable graph and $w$ is a word representing $G$, then for any vertex $x \in V(G)$ the word obtained from $w$ by deleting all appearances of $x$ represents $G - x$. This observation leads to the following obvious conclusion.

**Proposition 55.** *The class of word-representable graphs is hereditary.*

We now apply the Alekseev-Bollobás-Thomason Theorem in order to derive an asymptotic formula for the number of word-representable graphs. We start with the number of $n$-vertex *labelled* graphs in this class, which we denote by $b_n$.

**Theorem 56.**

$$\lim_{n\to\infty} \frac{\log_2 b_n}{\binom{n}{2}} = \frac{2}{3}.$$

*Proof.* By Theorem 52, $\mathcal{E}_{3,0}$ is a subclass of the class of word-representable graphs and hence its index is at least 3. In order to show that the index does not exceed 3, we observe that the graph $A$ represented in Figure 7.2 belongs to all minimal classes of index

4. Indeed, the set of vertices of $A$ can be partitioned into two subsets, one inducing a path on 4 vertices $P_4$ (the four vertices in the middle of the graph) and the other an induced subgraph of a $P_4$. It is easy to check that $P_4$ (and hence any of its induced subgraphs) is bipartite (belongs to $\mathcal{E}_{2,0}$), co-bipartite (belongs to $\mathcal{E}_{0,2}$) and split (belongs to $\mathcal{E}_{1,1}$). Therefore, $A$ belongs to $\mathcal{E}_{4,0}$, $\mathcal{E}_{3,1}$, $\mathcal{E}_{2,2}$, $\mathcal{E}_{1,3}$ and $\mathcal{E}_{0,4}$. Since $A$ is not word-representable (Theorems 54), the family of word-representable graphs does not contain any of these minimal classes. Therefore, the index of the class of word-representable graphs is precisely 3. $\square$

We now proceed to the number of unlabelled $n$-vertex word-representable graphs, which we denote by $a_n$.

**Theorem 57.**

$$\lim_{n\to\infty} \frac{\log_2 a_n}{\binom{n}{2}} = \frac{2}{3}.$$

*Proof.* Clearly, $b_n \le n!a_n$ and $\log_2 n! \le \log_2 n^n = n \log_2 n$. Therefore,

$$\lim_{n\to\infty} \frac{\log_2 b_n}{\binom{n}{2}} \le \lim_{n\to\infty} \frac{\log_2(n!a_n)}{\binom{n}{2}} = \lim_{n\to\infty} \frac{\log_2 n! + \log_2 a_n}{\binom{n}{2}} \le \lim_{n\to\infty} \frac{n\log_2 n + \log_2 a_n}{\binom{n}{2}} = \lim_{n\to\infty} \frac{\log_2 a_n}{\binom{n}{2}}.$$

On the other hand, obviously $b_n \ge a_n$ and hence $\lim_{n\to\infty} \frac{\log_2 b_n}{\binom{n}{2}} \ge \lim_{n\to\infty} \frac{\log_2 a_n}{\binom{n}{2}}$. Combining, we obtain $\lim_{n\to\infty} \frac{\log_2 b_n}{\binom{n}{2}} = \lim_{n\to\infty} \frac{\log_2 a_n}{\binom{n}{2}}$. Together with Theorem 56, this proves the result. $\square$

**Corollary 2.**

$$a_n = 2^{\frac{n^2}{3} + o(n^2)}.$$

# Bibliography


[1] M. Aigner, G.M. Ziegler, Proofs from THE BOOK, 4th ed. (2009) Berlin, New York: Springer-Verlag.

[2] V. E. Alekseev, On lower layers of a lattice of hereditary classes of graphs. *Diskretn. Anal. Issled. Oper. Ser. 1*, 4(1):3–12, 1997. In Russian.

[3] V.E. Alekseev, Coding of graphs and hereditary classes, *Probl. Cybernet.* 39 (1982) 151–164 (in Russian).

[4] V.E. Alekseev, Range of values of entropy of hereditary classes of graphs, *Diskret. Mat.* 4 (1992), no. 2, 148–157 (in Russian; translation in *Discrete Mathematics and Applications*, 3 (1993), no. 2, 191–199).

[5] P. Allen, Forbidden induced bipartite graphs, *J. Graph Theory*, 60 (2009) 219–241.

[6] P. Allen, V. Lozin and M. Rao, Clique-width and the speed of hereditary properties, *Electronic Journal of Combinatorics*, 16 (2009) Research Paper 35.

[7] N. Alon, J. Balogh, B. Bollobás and R. Morris, The structure of almost all graphs in a hereditary property, *J. Combin. Theory* B 79 (2011) 131–156.

[8] N. Alon, G. Brightwell, H. Kierstead, A. Kostochka and P. Winkler, Dominating sets in $k$-majority tournaments. *Journal of Combinatorial Theory, Series B*, 96 (2006) 374–387.

[9] A. Atminas, A. Collins, J. Foniok, and V. Lozin, Deciding the Bell number for hereditary graph properties, *SIAM J. Discrete Mathematics*, 30 (2016) 1015–1031.

[10] A. Atminas, A. Collins, V. Lozin, and V. Zamaraev, Implicit representations and factorial properties of graphs, *Discrete Mathematics* 338 (2015) 164–179.

[11] J. Balogh, B. Bollobás and D. Weinreich, The speed of hereditary properties of graphs, *J. Combin. Theory* B 79 (2000) 131–156.

[12] J. Balogh, B. Bollobás, and D. Weinreich, The penultimate rate of growth for graph properties. *European J. Combin.*, 22(3):277–289, 2001.

[13] J. Balogh, B. Bollobás, and D. Weinreich, A jump to the Bell number for hereditary graph properties. *J. Combin. Theory Ser. B*, 95(1):29–48, 2005.

[14] B. Bollobás and A. Thomason, Projections of bodies and hereditary properties of hypergraphs, *Bull. London Math. Soc.*, 27 (1995) 417-424.

[15] B. Bollobás and A. Thomason, Hereditary and monotone properties of graphs, "The mathematics of Paul Erdős, II" (R.L. Graham and J. Nesetril, Editors), *Alg. and Combin.*, Vol. 14, Springer-Verlag, New York/Berlin (1997), 70-78.

[16] A. Brandstädt, and V.V. Lozin, On the linear structure and clique-width of bipartite permutation graphs, Ars Combinatoria, 67 (2003) 273–281.

[17] A. Brandstädt, V.B. Le, J. Spinrad, Graph classes: a survey. SIAM Monographs on Discrete Mathematics and Applications. SIAM, Philadelphia, PA, 1999. xii+304 pp. ISBN: 0-89871-432-X.

[18] R. Brignall, N. Korpelainen, V. Vatter, Linear Clique-Width for Hereditary Classes of Cographs, arXiv:1305.0636.

[19] M. Chudnovsky, N. Robertson, P. Seymour, R. Thomas, The strong perfect graph theorem, *Annals of Mathematics* 164 (1) (2006) 51-229.

[20] V. Chvátal and P. L. Hammer, Aggregation of inequalities in integer programming. In *Studies in integer programming (Proc. Workshop, Bonn, 1975)*, pages 145–162. Ann. of Discrete Math., Vol. 1. North-Holland, Amsterdam, 1977.

[21] A. Collins, J. Foniok, N. Korpelainen, V. Lozin, V. Zamaraev, Infinitely many minimal classes of graphs of unbounded clique-width, *Discrete Applied Mathematics* (2017).

[22] A. Collins, S. Kitaev and V.V. Lozin, New results on word-representable graphs, *Discrete Applied Mathematics* Vol. 216 (1) (2017) 136–141.

[23] B. Courcelle, J. Engelfriet and G. Rozenberg, Handle-rewriting hypergraph grammars, *Journal of Computer and System Sciences* 46 (1993) 218–270.

[24] B. Courcelle, J.A. Makowsky and U. Rotics, Linear time solvable optimization problems on graphs of bounded clique-width, *Theory Comput. Syst.* 33 (2000) 125–150.

[25] B. Courcelle and S. Olariu, Upper bounds to the clique-width of a graph, *Discrete Applied Math.* 101 (2000) 77–114.

[26] K. Dabrowski, M. Demange, V.V. Lozin, New Results on Maximum Induced Matchings in Bipartite Graphs and Beyond, *Theoretical Computer Science*, 478 (2013) 33–40.

[27] K. Dabrowski, V. Lozin and V. Zamaraev, On factorial properties of chordal bipartite graphs *Discrete Applied Mathematics*, 312 (2012) 2457-2465.

[28] J. Daligault, M. Rao and S. Thomassé, Well-Quasi-Order of Relabel Functions, *Order* 27 (2010) 301–315.

[29] G. Ding, On canonical antichains, *Discrete Mathematics* 309 (2009) 1123–1134.

[30] P. Erdős, P. Frankl, and V. Rödl, The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent. *Graphs Combin.*, 2(2):113–121, 1986.

[31] P. Erdős, D. J. Kleitman, and B. L. Rothschild, Asymptotic enumeration of $K_n$-free graphs. In *Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Tomo II*, number 17 in Atti dei Convegni Lincei, pages 19–27. Accad. Naz. Lincei, Rome, 1976.

[32] P. Erdős and J. Spencer, Probabilistic methods in combinatorics, *Probability and Mathematical Statistics*, Vol. 17. Academic Press, New York-London, 1974.

[33] S. Földes and P. L. Hammer, Split graphs. In *Proceedings of the Eighth Southeastern Conference on Combinatorics, Graph Theory and Computing (Louisiana State Univ., Baton Rouge, La., 1977)*, number XIX in Congressus Numerantium, pages 311–315. Utilitas Math., Winnipeg, Man., 1977.

[34] H. Galperin, A. Wigderson, Succinct representations of graphs, *Information and Control*, 56 (1983) 183–198.

[35] M. Halldórsson, S. Kitaev and A. Pyatkin, Graphs capturing alternations in words, *Lect. Notes Comput. Sci.* 6224 (2010) 436–437. *Proc. 14th Conf. on Developments in Language Theory, DLT* 2010, London, Ontario, Canada, August 17–20, 2010.

[36] M. Halldórsson, S. Kitaev and A. Pyatkin, Alternation graphs, *Lect. Notes Comput. Sci.* 6986 (2011) 191–202. Proceedings of the *37th International Workshop on Graph-Theoretic Concepts in Computer Science*, *WG* 2011, Teplá Monastery, Czech Republic, June 21–24, 2011.

[37] M. Kaminski and V.V. Lozin, Coloring edges and vertices of graphs without short or long cycles, *Contributions to Discrete Mathematics*, 2 (2007) 61–66.

[38] S. Kannan, M. Naor, S. Rudich, Implicit representation of graphs, *SIAM Journal on Discrete Mathematics*, 5 (1992) 596-603.

[39] S. Kitaev and S. Seif, Word problem of the Perkins semigroup via directed acyclic graphs, *Order* 25 (2008) 3, 177–194.

[40] S. Kitaev and A. Pyatkin, On representable graphs, *J. Autom., Lang. and Combin.* 13 (2008) 1, 45–54.

[41] S. Kitaev, P. Salimov, C. Severs and H. Úlfarsson, On the representability of line graphs, *Lecture Notes in Computer Science* 6795 (2011) 478–479. *Proc. 15th*

*Conf. on Developments in Language Theory, DLT* 2011, University of Milano-Bicocca, Milan, Italy, July 19–22, 2011.

[42] S. Kitaev, P. Salimov, C. Severs and H. Úlfarsson, Word-representability and line graphs, *Open J. Discr. Math.* 1 (2011) 2, 96–101.

[43] Ph. G. Kolaitis, H. J. Prömel, and B. L. Rothschild, $K_{l+1}$-free graphs: asymptotic structure and a 0–1 law. *Trans. Amer. Math. Soc.*, 303(2):637–671, 1987.

[44] N. Korpelainen and V. Lozin, Two forbidden induced subgraphs and well-quasi-ordering. *Discrete Math.*, 311(16):1813–1822, 2011.

[45] D. Kühn and D. Osthus, Induced subdivisions in $K_{s,s}$-free graphs of large average degree, *Combinatorica*, 24 (2004) 287–304.

[46] V.V. Lozin, Minimal classes of graphs of unbounded clique-width, *Annals of Combinatorics*, 15 (2011) 707–722.

[47] V. Lozin, C. Mayhill and V. Zamaraev, A note on the speed of hereditary properties, *Electronic J. Combinatorics*, 18 (2011) Research paper 157.

[48] V. Lozin, C. Mayhill and V. Zamaraev, Locally bounded coverings and factorial properties of graphs, *European Journal of Combinatorics*, 33 (2012) 534–543.

[49] V. Lozin, I. Razgon, V. Zamaraev, Well-quasi-ordering does not imply bounded clique-width, arXiv:1503.00571.

[50] V. Lozin, V. Zamaraev, The structure and the number of $P_7$-free bipartite graphs, *European Journal of Combinatorics*, 65 (2017) 143-153.

[51] V. Lozin and V. Zamaraev, Boundary properties of factorial classes of graphs, *J. Graph Theory*, submitted.

[52] S. Norine, P. Seymour, R. Thomas and P. Wollan, Proper minor-closed families are small, *J. Combinatorial Theory* B 96 (2006) 754–757.

[53] S. Oum, Rank-width and vertex-minors, *Journal of Combinatorial Theory, Series B*, 95 (1) (2005) 79-100

[54] S. Oum, P. Seymour, Approximating clique-width and branch-width, *Journal of Combinatorial Theory, Series B*, 96 (4) (2006) 514-528

[55] H. J. Prömel and A. Steger, Excluding induced subgraphs: quadrilaterals. *Random Structures Algorithms*, 2(1):55–71, 1991.

[56] H. J. Prömel and A. Steger, Excluding induced subgraphs. II. Extremal graphs. *Discrete Appl. Math.*, 44(1–3):283–294, 1993.

[57] H. J. Prömel and A. Steger, Excluding induced subgraphs. III. A general asymptotic. *Random Structures Algorithms*, 3(1):19–31, 1992.

[58] H. Prüfer, Neuer Beweis eines Satzes über Permutationen, *Arch. Math. Phys.* 27 (1918) 742-744.

[59] F.P. Ramsey, On a problem of formal logic, *Proceedings of the London Mathematical Society* 30 (1930) 264–286.

[60] N. Robertson and P.D. Seymour, Graph minors. V. Excluding a planar graph, *J. Combinatorial Theory Ser. B*, 41 (1986) 92–114.

[61] N. Robertson and P.D. Seymour, Graph Minors. XX. Wagner's conjecture, *Journal of Combinatorial Theory Ser. B*, 92 (2004) 325–357.

[62] E.R. Scheinerman and J. Zito, On the size of hereditary classes of graphs, *J. Combin. Theory* B 61 (1994) 16–39.

[63] S. Seif, Monoids with sub-log-exponential free spectra, *J. Pure Appl. Algebra* 212 (2008) 5, 1162–1174.

[64] J.P. Spinrad, Nonredundant 1's in Γ-free matrices, *SIAM J. Discrete Math.* 8 (1995) 251–257.

[65] J.P. Spinrad, Efficient graph representations. Fields Institute Monographs, 19. American Mathematical Society, Providence, RI, 2003. xiii+342 pp.

[66] M. Yannakakis The complexity of the partial order dimension problem. *SIAM J. Algebraic Discrete Methods*, 3(3):351–358, 1982.