

# Uncountably many minimal hereditary classes of graphs of unbounded clique-width

R. Brignall     D. Cocks  
School of Mathematics and Statistics  
The Open University, UK

March 30, 2021

## Abstract

Given an infinite word over the alphabet  $\{0, 1, 2, 3\}$ , we define a class of bipartite hereditary graphs  $\mathcal{G}^\alpha$ , and show that  $\mathcal{G}^\alpha$  has unbounded clique-width unless  $\alpha$  contains at most finitely many non-zero letters.

We also show that  $\mathcal{G}^\alpha$  is minimal of unbounded clique-width if and only if  $\alpha$  belongs to a precisely defined collection of words  $\Gamma$ . The set  $\Gamma$  includes all almost periodic words containing at least one non-zero letter, which both enables us to exhibit uncountably many pairwise distinct minimal classes of unbounded clique width, and also proves one direction of a conjecture due to Collins, Foniok, Korpelainen, Lozin and Zamaraev. Finally, we show that the other direction of the conjecture is false, since  $\Gamma$  also contains words that are *not* almost periodic.

## 1 Introduction

Typically, when some graph parameter is bounded for a given graph or collection of graphs, then there exist efficient (polynomial time) algorithms for a range of problems that are in general intractable. To give two examples, Courcelle's Theorem [5] states that any graph property expressible in  $\text{MSO}_2$  logic can be decided in linear time on graphs with bounded treewidth, and Courcelle, Makowsky and Rotics [4] showed that any graph property expressible in  $\text{MSO}_1$  logic has a linear time algorithm on graphs with bounded clique-width.

While treewidth is the parameter of choice for minor-closed classes of graphs, clique-width has the property that it can remain bounded for graphs with a high edge density, and is thus of more use with *hereditary* classes of graphs. Indeed, if  $H$  is an induced subgraph of  $G$ , then the clique-width of  $H$  is at most the clique-width of  $G$ . (For formal definitions, see Section 2.)

In light of the algorithmic consequences, a natural goal is to characterise which classes of graphs are unbounded with respect to a given parameter. A standard approach is to identify the *minimal* classes: for example, planar graphs are the unique minimal minor-closed class of graphs

of unbounded treewidth (see Robertson and Seymour [17]), and circle graphs are the unique minimal vertex-minor-closed class of unbounded rank-width (or, equivalently, clique-width) – see Geelen, Kwon, McCarty and Wollan [11].

The situation for clique-width and hereditary classes is much more complicated, yet remains of significant interest (note that every vertex-minor-closed class is a hereditary class, but not vice-versa). First, there exist hereditary classes of graphs that have unbounded clique-width, but which contain *no* minimal class of unbounded clique-width: it is well-known that the class of square grid graphs has this property; a more recent example is due to Korpelainen [13], who also suggests possible ways to handle such classes.

However, there do also exist minimal hereditary classes of unbounded clique-width. We refer the reader to the excellent survey by Dabrowski, Johnson and Paulusma [8] for further details of the progress in recent years. Of particular relevance here is the work of Collins, Foniok, Korpelainen and Lozin [3], in which a countably infinite family of minimal hereditary classes of unbounded clique-width is given.

More precisely, the authors of [3] construct hereditary bipartite graph classes by taking the finite induced subgraphs of an infinite graph whose vertices form a two-dimensional array and whose edges are defined by an infinite word over the alphabet  $\{0, 1, 2\}$ . They show that classes defined by an infinite periodic word over the alphabet  $\{0, 1\}$  are minimal of unbounded clique-width, and conjecture that a class defined by a word over the alphabet  $\{0, 1, 2\}$  is minimal of unbounded clique-width if and only if the word is almost periodic, and not the all-zeros word.

In this paper, we go further by proving the following results.

**Theorem 1.1.** *Let  $\alpha$  be an infinite word over the alphabet  $\{0, 1, 2, 3\}$ , and let  $\mathcal{G}^\alpha$  be the corresponding hereditary graph class (as defined in Section 2).*

- (a) *If  $\alpha$  has an infinite number of non-zero letters, then  $\mathcal{G}^\alpha$  has unbounded clique-width (Theorem 3.10).*
- (b) *If  $\alpha$  is an almost periodic word with at least one non-zero letter then  $\mathcal{G}^\alpha$  is minimal of unbounded clique-width (Theorem 4.9).*
- (c) *The number of distinct minimal hereditary classes of graphs of unbounded clique-width is uncountably infinite (Theorem 4.11).*
- (d) *Let  $\Gamma$  denote the set of all recurrent words over  $\{0, 1, 2, 3\}$  with at least one non-zero letter for which the weight of the word between any two consecutive occurrences of any factor is bounded. Then  $\mathcal{G}^\alpha$  is minimal of unbounded clique-width if and only if  $\alpha \in \Gamma$  (Theorem 5.1).*

Part (b) of Theorem 1.1 establishes that one direction of the conjecture in [3] is true, although note that we go further by including the additional letter 3 in the alphabet (which represents a different type of inter-column connection). The other direction of their conjecture is false: this follows from part (d) of Theorem 1.1, which establishes precisely which of the classes, that are defined by words over  $\{0, 1, 2, 3\}$  and have unbounded clique-width, are minimal with this property. We given an explicit counterexample of a word in  $\Gamma$  that is not almost periodic.

The rest of this paper is organised as follows. Section 2 provides some background, and defines key definitions and concepts. Section 3 is devoted to providing that the classes defined over  $\{0, 1, 2, 3\}$  in general have unbounded clique-width. This is done from first principles for classes over the alphabet  $\{2, 3\}$ , and then extended to the full four-letter alphabet using rank-width techniques.

Section 4 provides the central proof that if  $\alpha$  is an infinite almost periodic word with at least one non-zero letter then  $\mathcal{G}^\alpha$  is a minimal hereditary class of graphs of unbounded clique-width. To do this we modify the notion of cluster graphs, first used by Lozin [14], and show how this can be used in conjunction with Menger's Theorem to provide an integrated proof of the minimality result. That there are uncountably many distinct minimal hereditary classes of graphs of unbounded clique width follows by considering Sturmian sequences.

Finally, in Section 5, we explore sequences that are recurrent but not almost periodic, and prove the precise characterisation between minimal and non-minimal hereditary classes of graphs of unbounded clique-width. We also give a specific example of a word that is recurrent but not almost periodic, which defines such a minimal class.

## 2 Preliminaries

A graph  $G$  is a pair of sets, vertices  $V(G)$  and edges  $E(G) \subseteq V(G) \times V(G)$ . Unless otherwise stated, all graphs in this paper are simple, i.e. undirected, without loops or multiple edges. We denote  $N(v)$  as the neighbourhood of a vertex  $v$ , that is, the set of vertices adjacent to  $v$ .

A set of vertices is *independent* if no two of its elements are adjacent. A graph is *bipartite* if its vertices can be partitioned into two independent sets.

Given a graph  $G(V, E)$ , a subset  $U \subseteq V$  and a vertex  $v \in V \setminus U$ , we say that  $v$  *distinguishes*  $U$  if  $v$  has both a neighbour and a non-neighbour in  $U$ . If  $U$  is indistinguishable by the vertices outside  $U$ , we call  $U$  a *module*. A module  $U$  is *trivial* if  $|U| = 1$  or  $U = V(G)$ . A graph, every module of which is trivial, is called *prime*. We denote the set of prime induced subgraphs of  $G$ ,  $\text{Prime}(G)$ .

We will use the notation  $H \leq_1 G$  to denote graph  $H$  is an *induced subgraph* of graph  $G$ , meaning  $H$  can be obtained from  $G$  by a sequence of vertex removals. If graph  $G$  does not contain the induced subgraph  $H$  we say that  $G$  is *H-free*.

A class of graphs  $\mathcal{C}$  is *hereditary* if it is closed under taking induced subgraphs, that is  $G \in \mathcal{C}$  implies  $H \in \mathcal{C}$  for every induced subgraph  $H$  of  $G$ . It is well known that for any hereditary class  $\mathcal{C}$  there exists a unique (but not necessarily finite) set of minimal forbidden graphs  $\{H_1, H_2, \dots\}$  such that  $\mathcal{C} = \text{Free}(H_1, H_2, \dots)$ .

If  $H$  is an induced subgraph of  $G$ , then this can be witnessed by one or more embeddings, where an *embedding* of  $H$  in  $G$  is an injective map  $\phi : V(H) \rightarrow V(G)$  such that the subgraph of  $G$  induced by the vertices  $\phi(V(H))$  is isomorphic to  $H$ . In other words,  $vw \in E(H)$  if and only if  $\phi(v)\phi(w) \in E(G)$ .

## 2.1 Bipartite hereditary graph classes defined by an infinite word

The graph classes we consider are all formed by taking the set of finite induced subgraphs of an infinite graph defined on a grid of vertices. We start by defining an infinite empty graph  $\mathcal{P}$  with vertices

$$V(\mathcal{P}) = \{v_{i,j} : i, j \in \mathbb{N}\}.$$

In general, we think of  $\mathcal{P}$  as an infinite two-dimensional array in which  $v_{i,j}$  represents the vertex in the  $i$ -th row (counting from the left) and  $j$ -th column (counting from the top). Hence vertex  $v_{1,1}$  is in the top left corner of the grid and the grid extends infinitely to the right and downwards. The  $j$ -th column of  $\mathcal{P}$  is the set  $C_j = \{v_{i,j} : i \in \mathbb{N}\}$ , and the  $i$ -th row of  $\mathcal{P}$  is the set  $R_i = \{v_{i,j} : j \in \mathbb{N}\}$ .

We will refer to a (finite or infinite) sequence of letters chosen from a finite alphabet as a *word*. We denote by  $\alpha_j$  the  $j$ -th letter of the word  $\alpha$  and we denote  $\alpha_j^k$  to be the concatenation of  $k$  copies of the letter  $\alpha_j$ . A *factor* of  $\alpha$  is a contiguous subword of  $\alpha$ . The *length* of a word  $\alpha$  is the number of letters the word contains, while the *weight* of  $\alpha$  is the number of non-zero letters it has, which we will denote  $|\alpha|_1$ .

An infinite word  $\alpha$  is *recurrent* if each of its factors occurs in it infinitely many times. We say that  $\alpha$  is *almost periodic* (sometimes called *uniformly recurrent* or *minimal*) if for each factor  $\beta$  of  $\alpha$  there exists a constant  $\mathcal{L}(\beta)$  such that every factor of  $\alpha$  of length at least  $\mathcal{L}(\beta)$  contains  $\beta$  as a factor. Finally,  $\alpha$  is *periodic* if there is a positive integer  $p$  such that  $\alpha_k = \alpha_{k+p}$  for all  $k$ . Clearly, every periodic word is almost periodic, and every almost periodic word is recurrent.

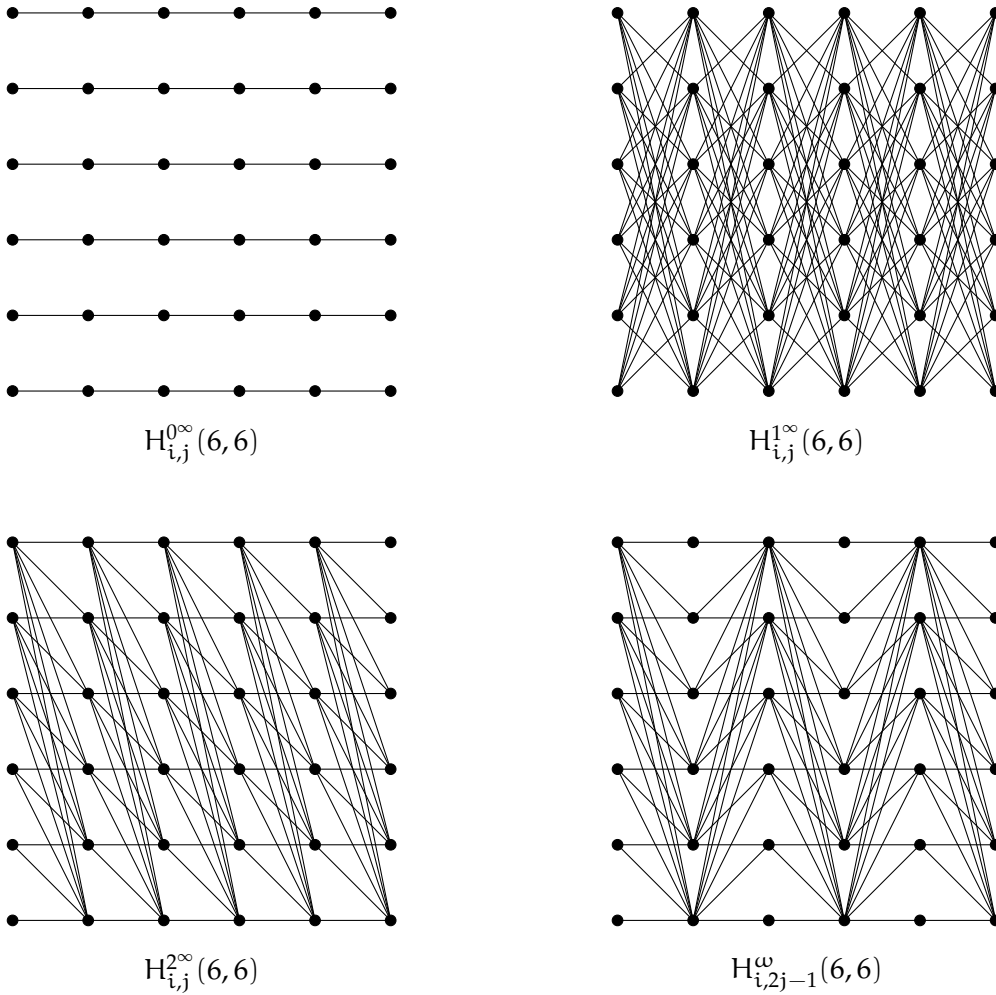
Let  $\alpha$  be an infinite word such that  $\alpha_j \in \{0, 1, 2, 3\}$  for each natural  $j$ . We define a family of infinite graphs  $\{\mathcal{P}^\alpha\}$  with vertices  $V(\mathcal{P})$ , and with edges between consecutive columns  $C_j, C_{j+1}$  of  $V(\mathcal{P}^\alpha)$ , such edges determined by the letters of the word  $\alpha$ .

- (i) If  $\alpha_j = 0$  then the edges between  $C_j$  and  $C_{j+1}$  are given by  $\{(v_{i,j}, v_{i,j+1}) : i \in \mathbb{N}\}$  (i.e. a matching).
- (ii) If  $\alpha_j = 1$  then the edges between  $C_j$  and  $C_{j+1}$  are given by  $\{(v_{i,j}, v_{k,j+1}) : i \neq k; i, k \in \mathbb{N}\}$  (i.e. the complement of a matching).
- (iii) If  $\alpha_j = 2$  then the edges between  $C_j$  and  $C_{j+1}$  are given by  $\{(v_{i,j}, v_{k,j+1}) : i \leq k; i, k \in \mathbb{N}\}$ .
- (iv) If  $\alpha_j = 3$  then the edges between  $C_j$  and  $C_{j+1}$  are given by  $\{(v_{i,j}, v_{k,j+1}) : i \geq k; i, k \in \mathbb{N}\}$ .

This notation matches and extends that used in [3].

For ease of reference we will denote  $H_{i,j}^\alpha(m, n)$  as the  $m \times n$  induced subgraph of  $\mathcal{P}^\alpha$  formed from the vertices  $\{v_{x,y} : x = i, i+1, \dots, i+m-1, y = j, j+1, \dots, j+n-1\}$ . See Figure 1.

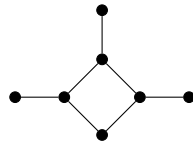
Let  $\mathcal{G}^\alpha$  denote the class of all finite induced subgraphs of  $\mathcal{P}^\alpha$ . By definition,  $\mathcal{G}^\alpha$  is a hereditary class, and any graph  $G \in \mathcal{G}^\alpha$  can be witnessed by an embedding into the infinite graph  $\mathcal{P}^\alpha$ . Given such an embedding of  $G$  into  $\mathcal{P}^\alpha$ , we will be especially interested in the induced subgraphs of  $G$  that occur in two adjacent columns: an  $\alpha_j$ -*link* is the induced subgraph of  $G$  on the vertices of  $G \cap (C_j \cup C_{j+1})$ , and will be denoted by  $G_j$ .



**Figure 1:** Bipartite graphs defined on a  $6 \times 6$  square grid.

Letting  $2^\infty$  stand for the infinite word of all 2s, we note that  $\mathcal{G}^{2^\infty}$  is the class of *bipartite permutation* graphs. Note, too, that  $\mathcal{G}^{3^\infty} = \mathcal{G}^{2^\infty}$  (this can be seen by considering a vertical reflection of the 2-dimensional array), so it is not necessarily the case that two words  $\alpha$  and  $\beta$  give rise to distinct hereditary classes.

On the other hand, the word with alternating 2s and 3s, which throughout we will denote by  $\omega = 232323 \dots$ , defines the class  $\mathcal{G}^\omega$  which is distinct from  $\mathcal{G}^{2^\infty}$ . Indeed, the graph,



embeds in  $\mathcal{P}^\omega$ , but is one of the minimal forbidden induced subgraphs of permutation graphs.

## 2.2 Clique-width and rank-width

Clique-width is a graph width parameter introduced by Courcelle, Engelfriet and Rozenberg in the 1990s [6]. A recent survey of clique-width for hereditary graph classes is [8]. The clique-width of a graph is denoted  $\text{cwd}(G)$  and is defined as the minimum number of labels needed to construct  $G$  by means of the following four graph operations:

- (a) creation of a new vertex  $v$  with label  $i$  (denoted  $i(v)$ ),
- (b) taking the disjoint union of two previously-constructed labelled graphs  $G$  and  $H$  (denoted  $G \oplus H$ ),
- (c) adding an edge between every vertex labelled  $i$  and every vertex labelled  $j$  for distinct  $i$  and  $j$  (denoted  $\eta_{i,j}$ ) and
- (d) giving all vertices labelled  $i$  the label  $j$  (denoted  $\rho_{i \rightarrow j}$ ).

Every graph can be defined by an algebraic expression  $\tau$  using the four operations above, which we will refer to as a *clique-width expression*. This expression is called a  $k$ -expression if it uses  $k$  different labels.

Alternatively, any clique-width expression  $\tau$  defining  $G$  can be represented as a rooted tree,  $\text{tree}(\tau)$ , whose leaves correspond to the operations of vertex creation, the internal nodes correspond to the  $\oplus$ -operation, and the root is associated with  $G$ . The operations  $\eta$  and  $\rho$  are assigned to the respective edges of  $\text{tree}(\tau)$ .

A related parameter is that of rank-width, which was introduced by Oum and Seymour in 2006 [16]. They showed that the measures are closely related through the inequality

$$\text{rwd}(G) \leq \text{cwd}(G) \leq 2^{\text{rwd}(G)+1} - 1$$

so that a graph class has bounded clique-width if and only if it has bounded 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. For a graph  $G$  and an edge  $vw$ , the graph obtained by *pivoting*  $vw$  is the graph obtained by applying local complementation at  $v$ , then at  $w$  and then at  $v$  again. When  $G$  is bipartite, Oum showed in [15] that this is equivalent to complementing the edges between  $N(v) \setminus w$  and  $N(w) \setminus v$ .

We will use the notation  $H \leq_v G$  to denote graph  $H$  is a *vertex-minor* of graph  $G$ , meaning  $H$  can be obtained from  $G$  by a sequence of vertex removals and local complementations. A useful result is the following.

**Lemma 2.1** (Oum [15]). *If  $H \leq_v G$  then  $\text{rwd}(H) \leq \text{rwd}(G)$ .*

### 3 Graph classes with unbounded clique-width

In this section we show that the graph classes  $\{\mathcal{G}^\alpha\}$ , where  $\alpha$  is an infinite word over  $\{0, 1, 2, 3\}$  that contains infinitely many letters from  $\{1, 2, 3\}$ , have unbounded clique-width. We will extend the methods of Golumbic and Rotics [12] and Brandstädt and Lozin [2], the latter of which proved that the clique-width of the class of bipartite permutation graphs (i.e.  $\mathcal{G}^{2^\infty}$ ) is unbounded. We deal with  $\{2, 3\}$  graphs first as these are susceptible to direct proof methods. We can then use rank-width/local complementation techniques following Collins et al [3] to show that the more complex  $\{0, 1, 2, 3\}$  graphs have a suitable  $\{2, 3\}$  vertex-minor to prove that these too have unbounded clique-width.

In the case of binary words, the following result covers what we require.

**Lemma 3.1** (Collins, Foniok, Korpelainen, Lozin and Zamaraev [3]). *If  $\alpha$  is an infinite binary word containing infinitely many 1s then the graph class  $\mathcal{G}^\alpha$  has unbounded clique-width.*

#### 3.1 $\{2, 3\}$ graph classes with unbounded clique-width

To assist with determining the clique-width of these graph classes, it is helpful to consider a sequence  $\{W_n^\alpha\}$  of graphs which are constructed as follows.

The vertices of  $W_n^\alpha$  are an  $n \times n$  array with vertex  $u_{i,j}$  in row  $i$  and column  $j$ . These vertices are partitioned according to the diagonal they are on, such that  $u_{i,j}$  is on diagonal  $D_{i+j-1}$ . Hence,  $D_1 = \{u_{1,1}\}$ ,  $D_2 = \{u_{2,1}, u_{1,2}\}$ , and so on.

$W_n^\alpha$  has an edge  $(u_{i,j}, u_{k,l})$  whenever  $u_{i,j} \in D_m, u_{k,l} \in D_{m+1}$  and either :

- (a)  $\alpha_m = 2$  and  $k \geq i$ , or
- (b)  $\alpha_m = 3$  and  $l \geq j$ .

Observe that these edges always create an  $n \times n$  square grid. Furthermore, the edges between two consecutive diagonals  $D_m$  and  $D_{m+1}$  of  $W_n^\alpha$  correspond to the edges between two consecutive columns  $C_m$  and  $C_{m+1}$  of  $\mathcal{P}^\alpha$ .

**Lemma 3.2.** *For all  $n \in \mathbb{N}$ ,  $W_n^\alpha$  can be embedded into  $\mathcal{P}^\alpha$ .*

*Proof.* For each vertex  $u_{x,y}$  of  $W_n^\alpha$  we will identify a vertex  $v_{i,j}$  of  $\mathcal{P}^\alpha$  so that we can define an injective map  $\phi : V(W_n^\alpha) \rightarrow V(\mathcal{P}^\alpha)$  such that  $\phi(u_{x,y}) = v_{i,j}$  as follows:

If  $(x, y) = (1, 1)$  then  $(i, j) = (n, 1)$ , otherwise  $(i, j)$  is given by:

$$i = n + x - 1 - \sum_{m=1}^{x+y-2} \mathbf{1}_m$$

$$j = x + y - 1$$



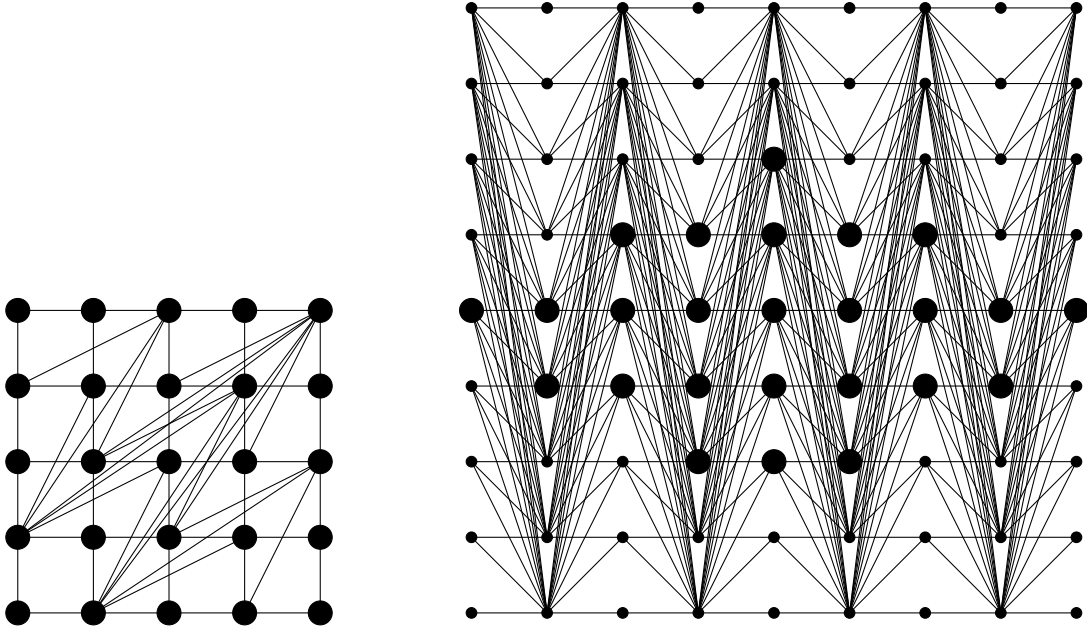


Figure 2: Embedding  $W_5^\omega$  in  $H_{1,1}^\omega(9,9)$  (shown by large vertices)

where

$$\mathbf{1}_m = \begin{cases} 1 & \text{if } \alpha_m = 3, \\ 0 & \text{if } \alpha_m = 2. \end{cases}$$

It can be seen that the subgraph of  $\mathcal{P}^\alpha$  induced by the vertices  $\phi(V(W_n^\alpha))$  is isomorphic to  $W_n^\alpha$ .  $\square$

An example of embedding  $W_5^\omega$  is shown in Figure 2.

We will now calculate a lower bound for the clique-width of  $W_n^\alpha$  by demonstrating a minimum number of labels needed to construct  $W_n^\alpha$  using the allowed four graph operations.

Let  $\tau$  be a clique-width expression defining  $W_n^\alpha$  and  $\text{tree}(\tau)$  the rooted tree representing  $\tau$ . The subtree of  $\text{tree}(\tau)$  rooted at a node  $x$  will be denoted  $\text{tree}(x, \tau)$ . This subtree corresponds to a subgraph of  $W_n^\alpha$  we will call  $W_x$ .

Let  $a$  be the lowest  $\oplus$  node in  $\text{tree}(\tau)$  such that  $W_a$  contains a full row and a full column of  $W_n^\alpha$ . We denote by  $r$  and  $b$  the two children of  $a$  in  $\text{tree}(\tau)$ . Let us colour the vertices of  $W_r$  and  $W_b$  red and blue, respectively, and all the other vertices in  $W_n^\alpha$  white. We let  $\text{colour}(v)$  denote the colour of a vertex  $v$  as described above, and  $\text{label}(v)$  denote the label of vertex  $v$  (if any) at node  $a$ .

We assume that there is neither a blue nor a red column in  $W_n^\alpha$ . For if  $W_n^\alpha$  contains a blue column then obviously it cannot contain a red row, and it cannot contain a blue row due to the choice of  $a$ . Likewise, if  $W_n^\alpha$  contains a red column. Hence, if it does include a blue or red column, as a consequence of the symmetry of  $W_n^\alpha$ , we can apply similar types of argument to those that follow to the rows instead of the columns to deliver the same result.



**Observation 3.3.** *Suppose  $u, v, w$  are three vertices such that  $u$  and  $v$  are non-white,  $uw \in E(W_n^\alpha)$  but  $vw \notin E(W_n^\alpha)$ , and  $\text{label}(w) \neq \text{label}(u)$ . Then  $u$  and  $v$  must have different labels at node  $a$  because the edge  $uw$  still needs to be created, whilst respecting the non-adjacency of  $v$  and  $w$ .*

Let us denote a row without white vertices by  $r$ . From the foregoing, we know that every column  $j$  in  $W_n^\alpha$  must have a vertex with different colour than that of  $v_{r,j}$ . We denote by  $x_j$  a nearest to  $v_{r,j}$  vertex in the same column with  $\text{colour}(x_j) \neq \text{colour}(v_{r,j})$ . We then let  $y_j$  denote  $v_{i+1,j}$  if  $i < r$  or  $v_{i-1,j}$  if  $i > r$ . Notice that  $y_j$  is non-white and  $x_j$  is adjacent to  $y_j$ . This creates two sets of  $n$  vertices,  $X = \{x_j : j = 1, \dots, n\}$  and  $Y = \{y_j : j = 1, \dots, n\}$  where  $x_j$  is adjacent to  $y_j$  for each  $j$ .

We examine first the case for the word  $\omega = 2323 \dots$ , the infinite word with alternating 2s and 3s, as this is relatively simple whilst demonstrating the technique required.

**Lemma 3.4.** *Let  $\omega$  be the infinite word with alternating 2s and 3s. Then*

$$\text{cwd}(W_n^\omega) \geq n/2.$$

*Proof.* We claim that for  $W_n^\omega$ , no three vertices in  $Y$  can have the same label. From this the lemma will follow since the  $n$  vertices in  $Y$  are labelled with at least  $n/2$  labels at node  $a$ . Suppose for a contradiction that vertices  $Y' = \{y_{j_1}, y_{j_2}, y_{j_3}\} \subseteq Y$  have the same label at  $a$ . Without loss of generality we assume that  $j_1 < j_2 < j_3$ . Let  $X' = \{x_{j_1}, x_{j_2}, x_{j_3}\} \subseteq X$  be the corresponding subset of  $X$ .

Assume a vertex  $y_{j_i}$  in  $Y'$  is not adjacent to a vertex  $x_{j_k}$  in  $X'$ . Clearly,  $i \neq k$  and hence vertices  $y_{j_i}, y_{j_k}$  and  $x_{j_k}$  form a triple as described in Observation 3.3, so  $y_{j_i}$  and  $y_{j_k}$  have different labels, a contradiction. Hence, each vertex in  $Y'$  must be adjacent to each vertex in  $X'$ .

**Observation 3.5.** (a) *if a vertex  $v_{i,j}$  lies on an odd diagonal  $D_{2m-1}$  then it has only one neighbour to its right and all its neighbours lie on the even diagonals  $D_{2m}$  and  $D_{2m+2}$ ;*

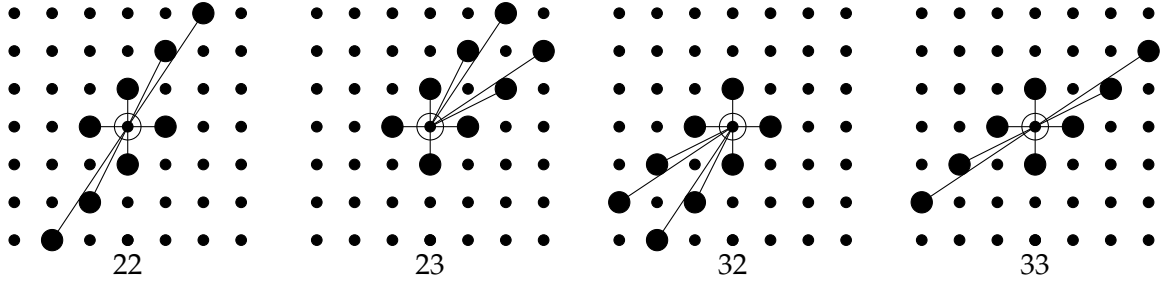
(b) *if a vertex  $v_{i,j}$  lies on an even diagonal  $D_{2m}$  then it has only one neighbour to its left and all its neighbours lie on the odd diagonals  $D_{2m-1}$  and  $D_{2m+1}$ .*

The leftmost vertex in  $X'$ ,  $x_{j_1}$ , must have two neighbours in  $Y'$  to its right, so from Observation 3.5,  $x_{j_1}$  must sit on an even diagonal and all the vertices in  $Y'$  must sit on odd diagonals. On the other hand, the rightmost vertex in  $X'$ ,  $x_{j_3}$ , must have two neighbours in  $Y'$  to its left, so from Observation 3.5,  $x_{j_3}$  must sit on an odd diagonal and all the vertices in  $Y'$  must sit on even diagonals. We have a contradiction and hence no three vertices in  $Y$  can have the same label, as required.  $\square$

The extension to consider arbitrary words over  $\{2, 3\}$  is similar, but the behaviour of the diagonals given in Observation 3.5 represent just two of several possible types of behaviour.

**Lemma 3.6.** *If  $\alpha$  is any infinite word over the alphabet  $\{2, 3\}$  then*

$$\text{cwd}(W_n^\alpha) \geq n/4.$$



**Figure 3:** Neighbourhood types (large vertices are neighbours of central vertex)

*Proof.* We claim that for  $W_n^\alpha$ , no five vertices in  $Y$  can have the same label. From this the lemma will follow since the  $n$  vertices in  $Y$  are labelled with at least  $n/4$  labels at node  $a$ .

Suppose for a contradiction that vertices  $Y' = \{y_{j_1}, y_{j_2}, y_{j_3}, y_{j_4}, y_{j_5}\} \subseteq Y$  have the same label at  $a$ . Without loss of generality we assume that  $j_1 < j_2 < j_3 < j_4 < j_5$ . Let  $X' = \{x_{j_1}, x_{j_2}, x_{j_3}, x_{j_4}, x_{j_5}\} \subseteq X$  be the corresponding subset of  $X$ . As in Lemma 3.4, we must have each vertex in  $Y'$  adjacent to each vertex in  $X'$ .

If a vertex lies on diagonal  $D_m$  we say it has *type*  $\alpha_{m-1}\alpha_m$ . Hence, we have 4 possible types of vertex, namely 22, 23, 32 and 33. The distinguishing feature of each is its neighbourhood. Each has two *branches* to its neighbourhood on diagonals  $D_{m-1}$  and  $D_{m+1}$  – see Figure 3.

- (a) a type 22 vertex neighbourhood has one branch ( $D_{m-1}$ ) extending up/right and one ( $D_{m+1}$ ) extending down/left,
- (b) a type 23 vertex neighbourhood has two branches ( $D_{m-1}$  and  $D_{m+1}$ ) extending up/right,
- (c) a type 32 vertex neighbourhood has two branches ( $D_{m-1}$  and  $D_{m+1}$ ) extending down/left,
- (d) a type 33 vertex neighbourhood has one branch ( $D_{m-1}$ ) extending down/left and one ( $D_{m+1}$ ) extending up/right.

We have 5 vertices in  $X'$  each of which could be one of 4 types, so potentially there are  $4^5 = 1,024$  different cases, but we can handle these in groups that depend on the type of the two outside vertices  $x_{j_1}$  and  $x_{j_5}$ .

*Case 1:*  $x_{j_1}$  and  $x_{j_5}$  are each of type 22 or 33.

Since  $x_{j_1}$  must be adjacent to 4 vertices in  $Y'$  to its right and  $x_{j_5}$  must be adjacent to 4 vertices in  $Y'$  to its left, then all the vertices in  $Y'$  must be on the same diagonal (as the neighbourhood of  $x_{j_1}$  has only one branch going rightwards and the neighbourhood of  $x_{j_5}$  has only one branch going leftwards). But now there is no possible neighbourhood type for  $x_{j_3}$  that could make it simultaneously adjacent to  $y_{j_1}$  and  $y_{j_5}$ , so we have a contradiction.

*Case 2:*  $x_{j_1}$  and  $x_{j_5}$  are each of type 23 or 32.

In Lemma 3.4 we dealt with the case where there are 3 vertices in  $X'$  that are either 23 or 32 types so we only need to consider the case where the middle 3 vertices in  $X'$  are type 22 or 33.

Furthermore, as  $x_{j_1}$  must be adjacent to 4 vertices in  $Y'$  to its right and  $x_{j_5}$  must be adjacent to 4 vertices in  $Y'$  to its left, then  $x_{j_1}$  can only be of type 23 and  $x_{j_5}$  type 32. Then considering vertex  $x_{j_3}$  we must have all three vertices  $x_{j_1}$ ,  $x_{j_3}$  and  $x_{j_5}$  on the same diagonal otherwise  $x_{j_3}$  cannot be adjacent to both  $y_{j_1}$  and  $y_{j_5}$ . Lastly,  $x_{j_2}$  and  $x_{j_4}$  must be of the same type as  $x_{j_3}$  (i.e. either type 22 or 33) and on the same diagonal as  $x_{j_3}$  in order to be adjacent to both  $y_{j_1}$  and  $y_{j_5}$ . But then they cannot both be adjacent to  $y_{j_3}$  and we have a contradiction.

*Case 3:* One of  $x_{j_1}$  and  $x_{j_5}$  is of type 22 or 33 and the other is of type 23 or 32.

Suppose  $x_{j_1}$  is type 22 and  $x_{j_5}$  is type 32, then  $y_{j_3}$ ,  $y_{j_4}$  and  $y_{j_5}$  must be on the same diagonal.  $x_{j_3}$  cannot be type 23 or 32 as it could not be adjacent to both  $y_{j_1}$  and  $y_{j_5}$ .

If  $x_{j_3}$  is type 33 then it must be on the same diagonal as  $x_{j_1}$  in order for both  $x_{j_1}$  to be adjacent to  $y_{j_3}$  and  $x_{j_3}$  to be adjacent to  $y_{j_1}$ . But then  $x_{j_2}$  cannot be any of the 4 types and still be simultaneously adjacent to  $y_{j_1}$ ,  $y_{j_3}$  and  $y_{j_5}$ , hence we have a contradiction.

Likewise if  $x_{j_3}$  is type 22 then it must be on the diagonal  $D_{m-2}$  (where  $x_{j_1}$  is on  $D_m$ ) in order for both  $x_{j_1}$  to be adjacent to  $y_{j_3}$  and  $x_{j_3}$  to be adjacent to  $y_{j_1}$ . But then, as before,  $x_{j_2}$  cannot be any of the 4 types and still be simultaneously adjacent to  $y_{j_1}$ ,  $y_{j_3}$  and  $y_{j_5}$ , hence we have a contradiction.

For the other Case 3 combinations (e.g.  $x_{j_1}$  is type 33 and  $x_{j_5}$  is type 32) we can use an analogous argument and each time reach a contradiction. We omit the details.

We have now considered all possible combinations of vertex type and each one leads to a contradiction. Hence, no five vertices in  $Y$  can have the same label.  $\square$

**Lemma 3.7.** *If  $\alpha$  is any infinite word over the alphabet  $\{2, 3\}$  then the graph class  $\mathcal{G}^\alpha$  is a hereditary class of graphs of unbounded clique-width.*

*Proof.* This follows immediately from Lemmas 3.4 and 3.6.  $\square$

### 3.2 $\{0, 1, 2, 3\}$ graph classes with unbounded clique-width

We now extend our results to graph classes  $\mathcal{G}^\alpha$  where  $\alpha$  is an infinite word over the alphabet  $\{0, 1, 2, 3\}$  containing an infinite number of letters from  $\{2, 3\}$ . For this we will use the rank-width parameter described in Section 2.2. From [3] we have a toolkit of graph operations which we extend to show that the graph class  $\mathcal{G}^\alpha$  contains a graph with a vertex-minor  $H_{1,1}^\gamma(q, q)$  for some  $q$  where  $\gamma$  is an infinite word from the alphabet  $\{2, 3\}$ . If we can make  $q$  as large as we like then combining Lemma 2.1 with Lemma 3.7 gives us the result that  $\mathcal{G}^\alpha$  has unbounded rank-width and therefore unbounded clique-width.

The following graph operations are demonstrated in [3] unless otherwise stated. Each operation takes a graph  $H_{1,1}^\alpha(m, n)$  and uses local complementation and vertex deletion to create a  $p \times q$  vertex-minor  $H_{1,1}^\gamma(p, q)$ , for some  $p \leq m$  and  $q \leq n$ , where  $\gamma$  is a  $q - 1$  letter word derived from  $\alpha$  with certain letters removed. We use the term *0 removal* where the removed letter(s) are 0s and likewise *1 removal* where the removed letter(s) are 1s.

## 0 Removal Operations

- (i) Removing a 0 from the factor 00 can be achieved by applying local complementation to each of the vertices in the middle column and then deleting the vertices in that column.
- (ii) Removing the 0 from the factor 01 can be achieved by applying local complementation to each of the vertices in the middle column sequentially from top to bottom and then deleting the vertices in that column. If the number of rows  $m$  is even this is equivalent to removing the 0 from the factor 01. If the number of rows is odd the same result is achieved by modifying the process so that the local complementation ends on row  $m - 1$  and deleting the last row of vertices. The factor 10 can be reduced to 1 in the same way.
- (iii) Removing the 0 from the factor 02 can be achieved by local complementation on the vertices of the middle column and then deleting the odd rows. Also factor 20 can be reduced to 2, 03 to 3 and 30 to 3 in the same way.

These operations allow us to create a vertex-minor  $H_{1,1}^\gamma(p, q)$  with the letters of  $\gamma$  from the alphabet  $\{1, 2, 3\}$ .

## 1 Removal Operations

- (i) Transforming the factor 211 into 2 can be achieved by using one pivot and deleting the first and last rows to give a 200 factor and then using 0 removal operations to reduce to 2. In the same way we can transform the factor 112 into 2, 311 into 3 and 113 into 3.
- (ii) Transforming the factor 212 into 22 can be achieved by using one pivot and deleting the first and last rows to give a 202 factor and then using 0 removal operations to reduce to 22. In the same way we can also reduce 313 to 33.
- (iii) Finally, we claim we can transform the factor 213 into 22. As this is not covered by [3] we give the proof here.

Let  $C_k, C_{k+1}, C_{k+2}, C_{k+3}$  be four consecutive columns of  $H_{i,j}(m, n)$  such that  $C_k \cup C_{k+1}$  induce a 2-link,  $C_{k+1} \cup C_{k+2}$  induce a 1-link and  $C_{k+2} \cup C_{k+3}$  induce a 3-link. Let  $x$  be the vertex in the first row of column  $C_{k+1}$  and  $y$  be the vertex in the last row of column  $C_{k+2}$ . It can be seen that by pivoting on the edge  $xy$  and then deleting the whole of the first row, the second row to the right of column  $C_{k+2}$  and deleting  $y$  and vertices on the last row to the right of  $y$ , we have transformed the factor 213 into 202. We can then use the zero removal operations to reduce to 22. In the same way we can also reduce 312 to 33.

**Observation 3.8.** *If  $r$  is the number of rows prior to the removal of a 0 or 1 by one of these operations then after the operation the number of rows left will be at least  $(r/2) - 2$ .*

Thus, by starting with a large enough number  $m$  in our choice of  $H_{i,j}(m, n)$ , we may remove a finite number of 0s and 1s and still ensure that there are enough rows left at the end of the process.

We now have a complete set of tools, using local complementation and vertex removal applied to  $H_{1,1}^\alpha(m, n)$ , to create a vertex-minor  $H_{1,1}^\gamma(p, q)$  with the letters of  $\gamma$  coming only from the alphabet  $\{2, 3\}$ .

**Lemma 3.9.** *Let  $\alpha$  be an infinite word over the alphabet  $\{0, 1, 2, 3\}$  which has an infinite number of letters from  $\{2, 3\}$ . Further, let  $\beta$  be a factor  $\alpha_k \alpha_{k+1} \cdots \alpha_{k+p-1}$  of length  $p$  which has  $q$  ( $0 < q \leq p$ ) letters from  $\{2, 3\}$  and  $p - q$  letters from  $\{0, 1\}$ .*

*Then  $H_{1,k}^\alpha((q + 4)2^{p-q}, p)$  has a vertex-minor isomorphic to  $H_{1,1}^\gamma(q, q)$ , for some word  $\gamma$  using only letters from the alphabet  $\{2, 3\}$ .*

*Proof.* This follows by applying the graph operations described above to remove the 0s and 1s. There are  $p - q$  such letters, and using Observation 3.8 it can be seen that by starting with at least  $(q + 4)2^{p-q}$  rows, there will be at least  $q$  rows remaining after executing the necessary operations to remove them.  $\square$

We now have the following theorem:

**Theorem 3.10.** *If  $\alpha$  is an infinite word from the alphabet  $\{0, 1, 2, 3\}$  with an infinite number of non-zero letters, then the graph class  $\mathcal{G}^\alpha$  is a hereditary class of graphs of unbounded clique-width.*

*Proof.* If there are no 2s or 3s or only a finite number of 2s and 3s in  $\alpha$ , we can use Lemma 3.1.

If there is an infinite number of 2s and 3s in  $\alpha$  then we can use Lemmas 3.7 and 3.9 as follows. For any  $q$  we can find a graph  $G$  in  $\mathcal{G}^\alpha$  that has a vertex-minor  $H_{1,1}^\gamma(q, q)$ , for some infinite word  $\gamma$  using only letters from the alphabet  $\{2, 3\}$ . In turn,  $H_{1,1}^\gamma(q, q)$  contains an induced subgraph  $W_{q/2}^\gamma$ , so using Lemma 2.1 we have

$$\text{rwd}(G) \geq \text{rwd}(H_{1,1}^\gamma(q, q)) \geq \text{rwd}(W_{q/2}^\gamma).$$

But from Lemma 3.6  $\text{cwd}(W_{q/2}^\gamma) \geq q/8 \rightarrow \infty$  as  $q \rightarrow \infty$  (and hence also  $\text{rwd}(W_{q/2}^\gamma) \rightarrow \infty$ ) so it follows that  $\mathcal{G}^\alpha$  is a graph class with unbounded rank-width and hence unbounded clique-width.  $\square$

## 4 Minimality of almost periodic graph classes

Let  $\alpha$  be an infinite almost periodic word from the alphabet  $\{0, 1, 2, 3\}$ , with at least one non-zero letter. In this section we will prove that the graph classes  $\mathcal{G}^\alpha$  are minimal of unbounded clique-width. To do this, we must show that any proper hereditary subclass has bounded clique-width. If  $\mathcal{C}$  is a proper hereditary subclass of  $\mathcal{G}^\alpha$  then there must exist a non-trivial finite forbidden graph  $F$  that is in  $\mathcal{G}^\alpha$  but not in  $\mathcal{C}$ . In turn, this graph  $F$  must be an induced subgraph of some  $H_{1,j}^\alpha(k, k)$  for some  $k \geq 2$ .

Consider a graph  $G \in \mathcal{C} \subseteq \text{Free}(H_{1,j}^\alpha(k, k))$ . If there exists an embedding  $\phi : V(G) \rightarrow V(\mathcal{P}^\alpha)$  straddling columns  $C_j, \dots, C_{j+k-1}$  of  $\mathcal{P}^\alpha$  then there must be limits on the vertices of  $\phi(V(G))$  in these columns to avoid creating an induced subgraph isomorphic to  $H_{1,j}^\alpha(k, k)$ .

Clearly the most obvious thing to avoid is any  $k$  complete horizontal rows which would automatically generate the forbidden graph. For graphs only involving letters 0 and 1 this is sufficient, and was dealt with in [3]. However, letters 2 and 3 are more complex. Lozin in [14] dealt with the class  $\mathcal{G}^{2^\infty}$  by introducing the concept of *clusters* and creating a cluster graph which provided a method of defining a partition of the vertices of  $G$  which gave the desired boundary on clique-width.

In considering  $\{0, 1, 2, 3\}$  graphs we use a modified version of the cluster graph method used in [14] which we describe in Section 4.1.

The following lemma will be used to place a bound on the clique-width of induced subgraphs of  $\mathcal{P}^\alpha$ . Given a graph  $G$  and a subset of vertices  $U \subseteq V(G)$ , 2 vertices of  $U$  will be called *U-similar* if they have the same neighbourhood outside  $U$ .  $U$ -similarity is an equivalence relation. The number of equivalence classes of  $U$  in  $G$  will be denoted  $\mu_G(U)$  (or  $\mu(U)$  when the context is clear). Also, by  $G[U]$  we will denote the subgraph of  $G$  induced by  $U$ . It follows that  $U$  is a module of  $G$  if and only if  $\mu(U) = 1$ .

**Lemma 4.1** (Lozin [14]). *If the vertices of a graph  $G$  can be partitioned into subsets  $U_1, U_2, \dots, U_n$  in such a way that for every  $i$*

- (a) *the clique-width of  $G[U_i]$  is at most  $k \geq 2$ , and*
- (b)  *$\mu(U_i) \leq l$  and  $\mu(U_1 \cup \dots \cup U_i) \leq l$ ,*

*then the clique-width of  $G$  is at most  $kl$ .*

**Corollary 4.2.** *Suppose for  $n \geq 2$  that the  $(n - 1)$  letter factor  $\alpha_j \alpha_{j+1} \dots \alpha_{j+n-2}$  consists of  $t \geq 1$  non-zeros and  $n - t - 1$  zeros. Then the clique-width of  $H_{i,j}^\alpha(m, n)$ , ( $m \geq 1$ ), is at most  $6t + 3$ .*

*Proof.* To build  $H_{i,j}^\alpha(m, n)$  we partition it into subsets  $U_1, U_2, \dots, U_m$  by including in  $U_k$  the vertices of the  $k$ -th row of  $H_{i,j}^\alpha(m, n)$ . This means  $G(U_k)$  is a disjoint union of paths so has clique-width at most 3. Also, only a vertex in  $U_k$  that is in a column  $s$  such that  $\alpha_{s-1}$  and/or  $\alpha_s$  is non-zero could have a neighbour outside  $U_k$ , so  $\mu(U_k) \leq 2t + 1$ . Also, it is not difficult to see that  $\mu(U_1 \cup \dots \cup U_k) \leq 2t + 1$ , as the vertices in each column have the same neighbourhood in  $H_{i,j}^\alpha(m, n)$  outside  $U_1 \cup \dots \cup U_k$ . Therefore, the result follows by applying Lemma 4.1.  $\square$

We will also make use of the following lemma.

**Lemma 4.3** (Courcelle and Olariu [7, Corollary 3.6]). *For every graph  $G$ ,*

$$\text{cwd}(G) = \max\{\text{cwd}(H) \mid H \in \text{Prime}(G)\}.$$

Thus we can assume that our arbitrary graph  $G$  is prime, and therefore connected, since if it were not so, we could any prime subgraph  $H$  which has the same clique-width.

## 4.1 Cluster graphs

Consider a connected graph  $G$  embedded in  $\mathcal{P}^\alpha$  such that its leftmost vertex is in column  $C_a$  and rightmost vertex in column  $C_{a+n-1}$ . Let a *left module* of  $G \cap C_j$  be a maximal set of vertices in that column that are indistinguishable by vertices in  $G \cap C_{j-1}$ . Similarly, a *right module* of  $G \cap C_j$  is a maximal set of vertices that are indistinguishable by vertices in  $G \cap C_{j+1}$ . Thus the vertices of every column of  $G$ , except the leftmost and rightmost columns, can be partitioned in two ways, as a set of left modules or as a set of right modules.

Now consider an  $\alpha_j$ -link,  $G_j$  (the subgraph of  $G$  induced by the vertices of  $G \cap C_j$  and  $G \cap C_{j+1}$ ). For convenience, we will say  $G_j$  is in *standard form* if, without changing the vertical order of the vertices, it is presented as an induced subgraph of  $H_{1,j}^\alpha(m, 2)$  with minimum  $m$  (i.e. taking out any superfluous gaps). Let  $G_j^s$  be the standard form of  $G_j$ , noting that the (left and right) modules of  $G_j$  and  $G_j^s$  are identical. Suppose  $R$  is the set of right modules of  $G_j^s \cap C_j$  and  $L$  the set of left modules of  $G_j^s \cap C_{j+1}$ . We will say that two modules  $A$  and  $B$  *overlap* if the set of rows containing vertices of  $A$  has non-zero intersection with the set of rows containing vertices of  $B$ . It can be seen that a right module in  $R$  can only overlap with a left module in  $L$  on at most one row, for if they overlapped on two or more rows they would no longer be (right/left) modules. Furthermore, a right module in  $R$  cannot overlap with more than one left module in  $L$  and vice-versa.

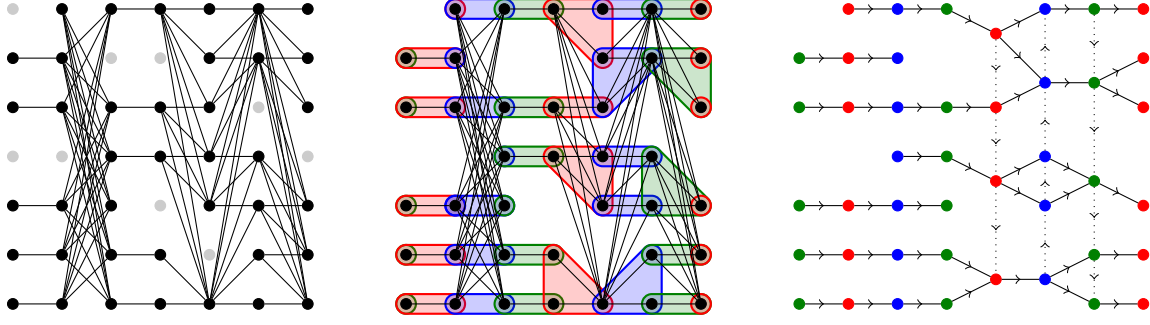
If a right module in  $R$  overlaps with a left module in  $L$ , the two modules can be paired to form a *cluster*. This pairing process is well-defined and matches all such modules except for at most one unmatched right module in  $G \cap C_j$  and one unmatched left module in  $G \cap C_{j+1}$ . These unmatched right/left modules have the characteristic they are indistinguishable to all vertices in the column to the right/left respectively. We will refer to them as *right/left boundary modules* and the vertices in them as *right/left boundary vertices* respectively. We put each boundary vertex in its own cluster in  $G_j$ .

Hence, the clusters of  $G_j$  form a partition of the vertices of the  $\alpha_j$ -link. If  $\alpha_j$  is 0 or 1 then every cluster in  $G_j$  is either a horizontal pair of vertices or a boundary vertex. If  $\alpha_j$  is 2 or 3 then each cluster is a complete bipartite induced subgraph of  $G_j$  or a boundary vertex. When there are no boundary vertices,  $G_j$  is isomorphic to  $H_{1,j}^\alpha(m, 2)$  consisting of  $m$  clusters, each containing two vertices of a same row.

At this stage, the vertices in the leftmost ( $G \cap C_a$ ) and rightmost ( $G \cap C_{a+n-1}$ ) columns are each only in one cluster as they only appear in one  $\alpha_j$ -link. We now add two additional columns of clusters, one to the left of  $G$  with a cluster for each vertex of  $G \cap C_a$ , and one to the right of  $G$  with a cluster for each vertex of column  $G \cap C_{a+n-1}$ . Thus the vertices of every column  $G \cap C_j$  of  $G$  are now in two clusters.

With any finite induced subgraph  $G$  of  $\mathcal{P}^\alpha$  we associate an oriented graph which we call the *cluster graph*  $B(G)$ , whose vertices are each associated with one of the clusters of  $G$ . The vertices of  $B(G)$  representing clusters of the same  $\alpha_j$ -link,  $G_j$ , we call a column of  $B(G)$ , and denote this by  $B(G_j)$ . Of the two additional cluster columns defined in the previous paragraph we will call the one on the left  $B(G_{a-1})$  and the one on the right  $B(G_{a+n})$ . For ease of exposition, we will always present the vertices of  $B(G_j)$  in the same vertical order as in  $G$ .





**Figure 4:** For  $\alpha = 010232 \dots$ , an illustration of the formation of a cluster graph. From left to right: an embedding of  $G^*$  in  $H_{1,1}(7,7)$ , the clusters associated with this embedding, and the cluster graph  $B^* = B(G^*)$ , with type B edges indicated by dotted lines.

In the following we denote the  $i$ -th cluster of  $G_j$ , counting from top to bottom, as  $K_{i,j}$ , with corresponding vertex,  $u_{i,j}$  in  $B(G_j)$ . The edges of  $B(G)$  are defined as follows.

**Type A** If  $K_{r,j}$  has a vertex of  $G$  in common with cluster  $K_{s,j+1}$  then  $B(G)$  has a directed edge  $(u_{r,j}, u_{s,j+1})$  (i.e an edge oriented from  $u_{r,j}$  to  $u_{s,j+1}$ ). Note that if there was more than one vertex of  $G$  in the intersection between two clusters of  $G$ , then these form a module of size greater than one and  $G$  is not prime, a contradiction. Hence, any two clusters of  $G$  have at most one vertex in the intersection. It follows that each Type A edge of  $B(G)$  corresponds to a vertex of  $G$ .

**Type B** Let  $u_{i,j}$  and  $u_{i+1,j}$  be two consecutive vertices in a column of  $B(G)$ . If  $\alpha_j = 2$  then  $B(G)$  has a directed edge  $(u_{i,j}, u_{i+1,j})$  and if  $\alpha_j = 3$  then  $B(G)$  has a directed edge  $(u_{i+1,j}, u_{i,j})$ .

Edges of type A are oriented from left to right and go between consecutive columns of  $B(G)$  whilst edges of type B are oriented down when  $\alpha_j = 2$  and up when  $\alpha_j = 3$ . Drawing  $B(G)$  by arranging the vertices in columns in the same order as the respective clusters of  $G$  it becomes clear that  $B(G)$  is a planar graph.

If we have a right boundary vertex in  $C_j$  then it will be in both a cluster in  $G_{j-1}$ , say  $K_{r,j-1}$ , and a singleton cluster in  $G_j$ , say  $K_{s,j}$ . The two vertices,  $u_{r,j-1}$  and  $u_{s,j}$  in  $B(G)$  associated with these clusters will be joined by a type A edge. However, there can be no type A edge to the immediate right of  $u_{s,j}$  as  $K_{s,j}$  contains no vertex from column  $C_{j+1}$ . Similarly for left boundary vertices, mutatis mutandis.

An example of a cluster graph is shown in Figure 4.

As we assume  $G$  is prime (from Lemma 4.3) and therefore connected, when it is embedded in  $\mathcal{P}^\alpha$  it must occupy vertices in consecutive columns (i.e. if it has one or more vertices in columns  $C_x$  and  $C_{x+2}$  then it must also have at least one vertex in column  $C_{x+1}$ ). Suppose  $G$  straddles a set of columns including the  $k$  columns  $C_j, \dots, C_{j+k-1}$  with defining factor  $\beta = \alpha_j \alpha_{j+1} \dots \alpha_{j+k-2}$ . Let us denote the subgraph of  $G$  induced by these columns  $G^*$ . The respective graph  $B(G^*)$  will be denoted by  $B^*$ ; it has  $k+1$  columns denoted by  $B_1, \dots, B_{k+1}$ . It can be

seen that if  $B^*$  has  $k$  directed disjoint paths from column  $B_1$  to column  $B_{k+1}$  then  $G^*$  contains the forbidden subgraph  $H_{1,1}^\alpha(k, k)$ , although the reverse is not necessarily true.

This leads to the following result:

**Lemma 4.4.** *Let  $\mathcal{C}$  be a proper subclass of  $\mathcal{G}^\alpha$  such that  $\mathcal{C} \subseteq \text{Free}(H_{1,1}^\alpha(k, k)) \subset \mathcal{G}^\alpha$ . Furthermore, let  $G$  be any graph in  $\mathcal{C}$  with induced subgraph  $G^*$  and associated cluster graph  $B^*$  defined as above. Then  $B^*$  can have at most  $k - 1$  directed paths from column  $B_1$  to column  $B_{k+1}$ .*

## 4.2 Applying Menger's Theorem

We will be using Menger's Theorem to help us define a partition of the vertices of  $G$  on which to apply Lemma 4.1. Menger's Theorem is one of the cornerstones of graph theory.

**Theorem 4.5** (Menger, see e.g. Diestel [9]). *Let  $G$  be a directed graph and  $A, B \subseteq V(G)$ . Then the minimum number  $n$  of vertices separating  $A$  from  $B$  in  $G$  is equal to the maximum number of disjoint directed  $A \rightarrow B$  paths in  $G$ .*

**Corollary 4.6.** *If  $S$  is a set of  $n$  vertices forming such a separator, then there exists a partition of  $V(G) \setminus S$  into two sets  $X$  and  $Y$  such that there are no edges directed from a vertex in  $X$  to a vertex in  $Y$ .*

*Proof.* Let  $X$  denote the set of vertices  $V(G) \setminus S$  that can be reached from  $A$  by following directed paths, and let  $Y$  denote the set of vertices of  $V(G) \setminus S$  from which there starts a directed path that ends in a vertex of  $B$ . Now,  $X$  and  $Y$  are disjoint (by Menger's theorem). If there are any vertices of  $V(G) \setminus S$  that lie in neither  $X$  nor  $Y$ , we can assign them to either arbitrarily. Now every edge with one endpoint in  $X$  and the other in  $Y$  must be oriented from  $Y$  to  $X$ , otherwise we find a directed path from  $A$  to  $B$ .  $\square$

We can apply this to the cluster graph  $B^*$  referred to in Lemma 4.4 with columns  $B_1$  and  $B_{k+1}$  connected to each other by a set  $P$  of at most  $k - 1$  disjoint paths. Denote  $s = |P|$ . The  $s$  paths of  $P$  cut  $B^*$  into  $s + 1$  horizontal *stripes*, that is, subgraphs induced by two consecutive paths in  $P$  and all the vertices between them ( $s - 1$  such stripes) and 2 further stripes for the subgraphs induced by the top path and all vertices above it, and the bottom path and all vertices below it.

From Menger's Theorem these two columns can be separated from each other by a set  $S$  of  $s \leq k - 1$  vertices, containing exactly one vertex in each of the paths, such that there are no paths from  $B_1$  to  $B_{k+1}$  that avoid this set  $S$ . From Corollary 4.6 we have a partition of the vertices of  $B^* \setminus S$  into two sets  $X_{V(B^*)}$  and  $Y_{V(B^*)}$  such that there are no directed edges from a vertex in  $X_{V(B^*)}$  to a vertex in  $Y_{V(B^*)}$ . As  $B^*$  is planar this means we can draw a curve  $\Omega$  between  $X_{V(B^*)}$  and  $Y_{V(B^*)}$  such that this curve crosses  $B^*$  at precisely the set  $S$  and such that there are no directed edges crossing  $\Omega$  from  $X_{V(B^*)}$  to  $Y_{V(B^*)}$ . It follows that we can partition the Type A edges of  $B^*$  into two sets  $X_{E(B^*)}$  and  $Y_{E(B^*)}$  either side of  $\Omega$ , and as these Type A edges correspond to the vertices of  $G^*$  then we also have a partition of these vertices into two sets  $X_{V(G^*)}$  and  $Y_{V(G^*)}$ .

### 4.3 Almost periodic $\{0, 1, 2, 3\}$ graph classes are minimal of unbounded clique-width

We now come to the key result of Section 4.

**Lemma 4.7.** *Let  $\alpha$  be an infinite almost periodic word from the alphabet  $\{0, 1, 2, 3\}$  which has at least one non-zero letter, and  $k$  a natural number at least 2. Further, let  $\beta = \alpha_j \alpha_{j+1} \cdots \alpha_{j+k-2}$  be a  $k-1$  letter factor of  $\alpha$  such that  $\beta$  appears in every factor of  $\alpha$  of length  $\mathcal{L}(\beta)$ , so that  $H_{1,j}^\alpha(k, k)$  is a graph in  $\mathcal{G}^\alpha$  whose edges correspond to the subword  $\beta$ . Then any graph  $G$  in  $\mathcal{G}^\alpha$  that is  $H_{1,j}^\alpha(k, k)$ -free has clique-width bounded by a constant  $c(k, \mathcal{L}(\beta))$  that depends only on  $k$  and  $\mathcal{L}(\beta)$ .*

*Proof.* Let  $G$  be a graph in  $\mathcal{G}^\alpha$  that is  $H_{1,j}^\alpha(k, k)$ -free. In the following we refer to vertex grid coordinates  $(x, y)$  of an embedding of  $G$  in  $\mathcal{P}^\alpha$  as described in Section 2.1. As before, we assume  $G$  is prime (from Lemma 4.3) and therefore connected.

We define a partition  $\{V_1, \dots, V_n\}$  of the vertices of  $G$  as follows. Let  $a$  be the first column of  $\mathcal{P}^\alpha$  in which a vertex of  $G$  is embedded. Denoting the set of vertices of  $G$  in a set of consecutive columns as a *bar*, let  $V_i$  be the bar of  $G$  in columns  $[a + (i-1)(\mathcal{L}(\beta) + 1)]$  through to  $[(a-1) + i(\mathcal{L}(\beta) + 1)]$ .

The corresponding subword for the graph induced by bar  $V_i$  is of length  $\mathcal{L}(\beta)$  so must contain a copy of  $\beta$  by definition. Let this copy of  $\beta$  correspond to columns  $C_y, \dots, C_{y+k-1}$  of  $\mathcal{P}^\alpha$ . Following the same notation as Section 4.1 we define  $G^*$  as the subgraph of  $G$  induced by the columns  $C_y, \dots, C_{y+k-1}$  and  $B^*$  its respective cluster graph, with columns  $B_1, \dots, B_{k+1}$ . We define  $P, S, s, \Omega, X_{E(B^*)}, Y_{E(B^*)}, X_{V(G^*)}$  and  $Y_{V(G^*)}$  as in Section 4.2.

We now show that the partition  $X_{V(G^*)}/Y_{V(G^*)}$  of the vertices of  $G^*$  defined in Section 4.2 gives us a number of equivalence classes,  $\mu_{G^*}(X_{V(G^*)})$  and  $\mu_{G^*}(Y_{V(G^*)})$ , bounded by a function of  $k$ . We consider this in 3 cases depending on the alphabet of  $\beta$ :

**Case 1**  $\beta$  is a subword from the alphabet  $\{0, 1\}$ .

A  $\{0, 1\}$  cluster graph  $B(G)$  contains only edges of type A and only horizontal paths. Every cluster is either a horizontal pair of vertices or a boundary vertex. Each row of  $B(G)$  is either a (left to right) directed path or a disjoint union of directed paths. If a row is a disjoint union of paths then the gaps between the paths have either no vertex or a boundary vertex immediately on either side.

It is easy to see that the curve  $\Omega$  must traverse each stripe of  $B^*$  by passing through a gap in each row between the paths at the top and bottom of the stripe.

From Section 4.2 the  $X_{E(B^*)}/Y_{E(B^*)}$  partition of the Type A edges of  $B^*$  defined by  $\Omega$  gives a corresponding  $X_{V(G^*)}/Y_{V(G^*)}$  partition of the vertices of  $G^*$ . We can partition the edges of  $B^*$  that correspond to vertices of a column  $C_j$  of  $G^*$ , into at most  $2s + 1$  subsets  $C_{1,j}, \dots, C_{2s+1,j}$ , as follows:

- (i) The edges forming the paths of  $P$  ( $s$  edges/subsets).
- (ii) The remaining edges in each stripe (at most  $s + 1$  subsets).

We claim that no vertex of  $Y_{V(G^*)}$  can distinguish the vertices of  $C_{i,j} \cap X_{V(G^*)}$ . Suppose to the contrary, that a vertex  $y \in Y_{V(G^*)}$  is not adjacent to  $x_1 \in C_{i,j} \cap X_{V(G^*)}$  but is adjacent to  $x_2 \in C_{i,j} \cap X_{V(G^*)}$ . Then  $x_1$  and  $x_2$  cannot be in the same cluster in  $G_j$ , because they are on different rows, but one of them must be in the same cluster as  $y$ . But this cluster is then not a boundary cluster as it contains two vertices and hence the cluster must be on a path of  $P$ . But  $x_1$  and  $x_2$  are in different clusters in  $G_j$  so cannot both be on the same path of  $P$  and so are in different subsets,  $C_{i,j} \cap X_{V(G^*)}$ , a contradiction.

So the maximum number of equivalence classes in  $C_j \cap X_{V(G^*)}$  is  $2s+1 \leq 2k-1$  and hence  $\mu_{G^*}(X_{V(G^*)})$  is at most the number of different  $C_{i,j}$ 's, which is at most  $k(2s+1) \leq 2k^2 - k$ . Symmetrically,  $\mu_{G^*}(Y_{V(G^*)}) \leq 2k^2 - k$ .

**Case 2**  $\beta$  is a subword from the alphabet  $\{2, 3\}$ .

Without loss of generality, we may assume that no  $\alpha$ -link  $G_j$ , where  $\alpha_j \in \{2, 3\}$ , contains a boundary vertex. For if such vertices exist, they will be positioned at one extreme (top or bottom) of a column. It is then possible to add an additional vertex in the opposite column to turn them into a cluster. Therefore, by adding at most two vertices to each column of  $G$ , we can extend it to a graph  $G'$  which has no boundary vertices, contains  $G$  as an induced subgraph and is  $H_{1,j}^\alpha(k+2, k)$ -free.

**Observation 4.8.** *The curve  $\Omega$  traverses each stripe of  $B^*$  monotonically in a horizontal direction, meaning that its  $x$ -coordinate changes within a stripe either non-increasingly or non-decreasingly.*

*Proof of Observation.* Suppose for a contradiction that  $\Omega$  had an unavoidable local maximum within a stripe, we would have a vertex  $v$  (to the left of the curve) that causes this maximum  $x$ -coordinate. Obviously,  $v$  does not belong to  $B_{k+1}$  (since otherwise  $B_{k+1}$  is not separated from  $B_1$ ), and  $v$  must have a neighbour to its right within the stripe (since there are no boundary vertices in  $G$ ). But then the Type A edge connecting  $v$  to that neighbour would cross  $\Omega$ , which contradicts Corollary 4.6.  $\square$

This observation allows us to conclude that whenever  $\Omega$  separates the Type A edges between two columns of  $B^*$  within a stripe, the result is two intervals, one above  $\Omega$  and one below it.

From Section 4.2 the  $X_{E(B^*)}/Y_{E(B^*)}$  partition of the Type A edges of  $B^*$  defined by  $\Omega$  gives a corresponding  $X_{V(G^*)}/Y_{V(G^*)}$  partition of the vertices of  $G^*$ . We can partition the Type A edges of  $B^*$  that correspond to vertices of a column  $C_j$  of  $G^*$ , into at most  $4s+1 \leq 4k-3$  subsets  $C_{1,j}, \dots, C_{4s+1,j}$ , as follow:

- (i) The Type A edges intersecting the paths of  $P$  ( $s$  edges/subsets).
- (ii) For each such edge  $e$ , the Type A edges that have a common vertex with  $e$ , at most 2 subsets in each stripe (up to  $2s$  subsets).
- (iii) The remaining Type A edges in each stripe (at most  $s+1$  subsets).

From Observation 4.8 the vertices of each  $C_{i,j}$  form an interval, i.e., they are consecutive in  $C_j$ . We claim that no vertex of  $Y_{V(G^*)}$  can distinguish the vertices of  $C_{i,j} \cap X_{V(G^*)}$ . Suppose to the contrary, that a vertex  $y \in Y_{V(G^*)}$  is not adjacent to  $x_1 \in C_{i,j} \cap X_{V(G^*)}$  but

is adjacent to  $x_2 \in C_{i,j} \cap X_{V(G^*)}$ . Without loss of generality we assume that  $\alpha_j = 2$ , as the case  $\alpha_j = 3$  follows by symmetry.

The vertices  $x_1$  and  $x_2$  cannot be in the same cluster in  $G_j$  because they are distinguished by  $y$ . Furthermore, as  $\alpha_j = 2$  then  $y$  must be in column  $C_{j+1}$  on a row above that of  $x_1$  but below or level with the row of  $x_2$ . We can assume that  $y$  is in the same  $G_j$  cluster as  $x_2$  because if it is not, then it must be in a cluster positioned between the cluster containing  $x_2$  and the cluster containing  $x_1$ . As  $C_{i,j}$  is an interval, this cluster must include some other vertex  $x_3 \in C_{i,j} \cap X_{V(G^*)}$ , and we can proceed using  $x_3$  instead of  $x_2$ . Let the cluster of  $G_j$  including  $y$  and  $x_2$  be denoted  $K_{r,j}$ .

Let  $u_{r,j}$  denote the vertex of  $B^*$  corresponding to  $K_{r,j}$ . Also, let  $e_{x_1}, e_{x_2}, e_y$  be the edges of  $B^*$  corresponding to vertices  $x_1, x_2$ , and  $y$  respectively. Since  $e_{x_2}$  and  $e_y$  are incident to  $u_{r,j}$  but separated by  $\Omega$ , vertex  $u_{r,j}$  lies on  $\Omega$  and hence belongs to the separator  $S$ . Therefore,  $u_{r,j}$  belongs to a path from  $P$ . But then  $C_{i,j}$  is of the second type and therefore  $e_{x_1}$  must also be incident to  $u_{r,j}$ . This contradicts the fact that  $x_1$  does not belong to  $K_{r,j}$ . This contradiction shows that any two vertices of  $C_{i,j} \cap X_{V(G^*)}$  have the same neighbourhood in  $Y$ .

So  $\mu_{G^*}(X_{V(G^*)})$  is at most the number of different  $C_{i,j}$ s, which is at most  $k(4s + 1) \leq 4k^2 - 3k$ . Symmetrically,  $\mu_{G^*}(Y_{V(G^*)}) \leq 4k^2 - 3k$ .

**Case 3**  $\beta$  is a subword from the alphabet  $\{0, 1, 2, 3\}$ .

Each column  $B_i$  of cluster graph  $B^*$  is associated with a letter of  $\beta$ . We can divide  $B^*$  into alternating  $\{0, 1\}$  bars and  $\{2, 3\}$  bars (reminder, a bar is a set of consecutive columns). Suppose we label these bars  $D_1, D_2, \dots, D_m$  of lengths  $k_1, k_2, \dots, k_m$  so that  $k_1 + k_2 + \dots + k_m = k - 1$ . Without loss of generality, we will say that if  $i$  is odd,  $D_i$  is a  $\{0, 1\}$  bar and if  $i$  is even  $D_i$  is a  $\{2, 3\}$  bar.

Define the partition curve  $\Omega$  as before. If  $\Omega$  stays in only one  $\{0, 1\}$  or  $\{2, 3\}$  bar then we can revert to Case 1 or Case 2. If  $\Omega$  straddles several  $\{0, 1\}$  and  $\{2, 3\}$  bars we can argue as follows.

It can be observed that, within each stripe,  $\Omega$  can only pass at most once through each bar in  $B^*$ . For if it passed twice through a column in a  $\{2, 3\}$  bar, in a given stripe, with at least one vertex between the two sections of  $\Omega$  then there must be a Type B edge passing from  $X_{V(B^*)}$  to  $Y_{V(B^*)}$  which contradicts Corollary 4.6 of Menger's Theorem.

Also, for the same reasons given in Case 2, within each stripe, the line  $\Omega$  must pass across  $\{2, 3\}$  bars monotonically in a left/right  $x$ -coordinate sense.

Using the same arguments as used in the  $\{0, 1\}$  and  $\{2, 3\}$  proofs we can partition the vertices of  $G^*$  into sets  $X_{V(G^*)}$  and  $Y_{V(G^*)}$  such that the maximum number of different equivalence classes for a column  $X_{V(G^*)} \cap C_j$  is  $(2k - 1) \leq (4k - 3)$  for a  $\{0, 1\}$  column and  $(4k - 3)$  for a  $\{2, 3\}$  column. So for  $k \geq 2$  we have  $\mu_{G^*}(X_{V(G^*)}) \leq 4k^2 - 3k$ . Symmetrically,  $\mu_{G^*}(Y_{V(G^*)}) \leq 4k^2 - 3k$ .

Using this  $X_{V(G^*)}/Y_{V(G^*)}$  partition of the vertices of  $G^*$  we can create a partition of  $V_i$ . All vertices in  $V_i$  in columns to the left of  $G^*$  are added to the vertices of  $X_{V(G^*)}$  and all the vertices in  $V_i$  in columns to the right of  $G^*$  are added to the vertices of  $Y_{V(G^*)}$  to produce a partition of

bar  $V_i$  into two parts  $X_i$  and  $Y_i$ . Let  $U_i = Y_{i-1} \cup X_i$ . Each  $G[U_i]$  has at most  $2(\mathcal{L}(\beta) + 1)$  columns. The subsets  $U_1, \dots, U_n$  form a partition of the vertices of  $G$ , such that for every  $i$ :

- (a) using Corollary 4.2 the clique-width of  $G[U_i]$  is at most  $6(2\mathcal{L}(\beta) + 2) + 3 = 12\mathcal{L}(\beta) + 15$ , and
- (b)  $\mu(U_i) \leq 2(4k^2 - 3k)$  and  $\mu(U_1 \cup \dots \cup U_i) \leq 2(4k^2 - 3k)$ .

Thus from Lemma 4.1 the clique-width of  $G$  is at most  $(12\mathcal{L}(\beta) + 15)(8k^2 - 6k)$ .

Hence the clique-width of  $G$  is bounded by a constant that depends only on  $k$  and  $\mathcal{L}(\beta)$ .  $\square$

**Theorem 4.9.** *Let  $\alpha$  be an infinite almost periodic word over the alphabet  $\{0, 1, 2, 3\}$  containing at least one non-zero letter. Then the class  $\mathcal{G}^\alpha$  is a minimal hereditary class of graphs of unbounded clique-width.*

*Proof.* If  $\mathcal{C}$  is a proper hereditary subclass of  $\mathcal{G}^\alpha$  then there must exist a non-trivial finite forbidden graph  $F$  that is in  $\mathcal{G}^\alpha$  but not in  $\mathcal{C}$ . But  $F$  must be an induced subgraph of some  $H_{1,j}^\alpha(k, k)$  so  $\mathcal{C} \subseteq \text{Free}(H_{1,j}^\alpha(k, k))$  and Lemma 4.7 gives us a bound on the clique-width. Hence,  $\mathcal{G}^\alpha$  is a minimal hereditary class of graphs of unbounded clique-width.  $\square$

#### 4.4 Uncountably many minimal graph classes with unbounded clique-width

We now proceed to show that there is an uncountably infinite number of such graph classes. To do this we will use the class of almost periodic sequences known as *Sturmian*. One definition of a Sturmian sequence is a binary sequence that has *complexity*  $p_\alpha(n) = n + 1$ , where the complexity function  $p_\alpha(n)$  is the number of different factors of length  $n$  in  $\alpha$  [10].

An alternative characterisation of Sturmian sequences is as rotation sequences defined by an irrational number, and hence it follows that the number of such sequences is uncountably infinite. We say that two sequences are *locally isomorphic* if they have the same factors. If two Sturmian sequences are locally isomorphic this means they have the same  $n + 1$  factors of length  $n$  out of a possible  $2^n$  such factors [18]. Hence the set of Sturmian sequences with a particular set of factors is countable in number and so it follows there is an uncountable number of such sets with different factors.

We denote  $\text{rev}(\beta)$  as the sequence  $\beta$  in reverse order (mirror image).

**Lemma 4.10.** *Let  $\alpha$  be an infinite binary word and  $\beta$  a finite binary word of length  $k - 1$  ( $k \geq 2$ ) with at least one 1. Further, let  $F_\beta$  be the graph  $H_{1,1}^\beta(3, k)$ .*

*Then  $F_\beta$  can be embedded in  $\mathcal{P}^\alpha$  if and only if  $\beta$  or  $\text{rev}(\beta)$  is a factor of  $\alpha$ .*

*Furthermore, if  $k \geq 3$ , such embedding is only possible in 3 rows and  $k$  consecutive columns  $s, \dots, s + k - 1$  when  $\beta$  (or  $\text{rev}(\beta)$ ) =  $\alpha_s \cdots \alpha_{s+k-1}$ .*



*Proof.* Clearly, by its definition,  $F_\beta$  can be embedded in  $\mathcal{P}^\alpha$  in the way described if  $\beta$  (or  $\text{rev}(\beta)$ ) is a factor of  $\alpha$ . [To avoid much repetition in what follows we will just refer to  $\beta$  to mean  $\beta$  or  $\text{rev}(\beta)$ .] We prove that  $F_\beta$  can only be embedded in  $\mathcal{P}^\alpha$  in the way described, and only if  $\beta$  is a factor of  $\alpha$ , by induction on  $k$ .

Firstly, if  $k = 2$  then  $\beta = 1$  and  $F_\beta = C_6$ , the cycle on 6 vertices. It is trivial to see that this can be embedded in  $\mathcal{P}^\alpha$  only if there is at least one 1 in  $\alpha$ . In fact,  $C_6$  can be embedded in two ways. Firstly, (Method 1) in the way described in the Lemma, with 6 vertices from 3 rows and 2 consecutive columns, or secondly, (Method 2) over 4 consecutive columns with 1 vertex from the first and last column and 2 from each of the middle columns.

If  $k = 3$  then  $\beta$  must be 10, 01 or 11.  $F_\beta$  still includes an induced subgraph  $C_6$ , but now the addition of 3 more vertices and corresponding edges means we can no longer use Method 2. Hence,  $F_\beta$  can only be embedded in  $\mathcal{P}^\alpha$  in the way described in the Lemma (Method 1).

Next using the strong induction hypothesis, we assume that the Lemma is true for all words of length less than  $k - 1$ . Thus if  $\beta$  contains a factor that is not a factor of  $\alpha$  then  $F_\beta$  cannot embed in  $\mathcal{P}^\alpha$ .

If  $F_\beta$  does embed in  $\mathcal{P}^\alpha$  then, if  $\beta^-$  is the word  $\beta$  without its last letter, we must have  $\beta^-$  a factor of  $\alpha$  where  $F_{\beta^-}$  can only embed in  $\mathcal{P}^\alpha$  by Method 1. Now it is straightforward to see that this cannot be extended to  $F_\beta$  if the next letter is not the same as the last letter of  $\beta$ , and that if it is the same, it can only be done by Method 1.  $\square$

**Theorem 4.11.** *There exists an uncountably infinite number of minimal hereditary classes of graphs of unbounded clique-width.*

*Proof.* There exists an uncountably infinite number of Sturmian binary sequences that are not locally isomorphic. Suppose we have Sturmian words  $\alpha_1$  and  $\alpha_2$  that have unique factors  $\beta_1$  and  $\beta_2$  respectively. Then using Lemma 4.10, the class  $\mathcal{G}^{\alpha_1}$  does not contain the graph  $F_{\beta_2}$  and the class  $\mathcal{G}^{\alpha_2}$  does not contain the graph  $F_{\beta_1}$ . So  $\mathcal{G}^{\alpha_1}$  and  $\mathcal{G}^{\alpha_2}$  are different graph classes. It follows from Theorem 4.9 each one defines a different minimal hereditary class of graphs of unbounded clique-width.  $\square$

## 5 Recurrent but not almost periodic words

We have seen that (with the exception of the all-zeros word) every almost periodic word  $\alpha$  over  $\{0, 1, 2, 3\}$  defines a minimal hereditary class  $\mathcal{G}^\alpha$  of unbounded clique width. At the other extreme, if  $\alpha$  is a word over  $\{0, 1, 2, 3\}$  that contains a factor  $\beta = \alpha_j \alpha_{j+1} \cdots \alpha_{j+k-2}$  that either does not repeat, or repeats only a finite number of times, then  $\mathcal{G}^\alpha$  cannot be a minimal class of unbounded clique-width, as forbidding the induced subgraph  $H_{1,j}^\alpha(k, k)$  would leave a proper subclass that by Theorem 3.10 still has unbounded clique-width.

Thus, to complete the delineation between minimality and non-minimality (with respect to having unbounded clique-width) of the classes  $\mathcal{G}^\alpha$ , it remains to consider words  $\alpha$  that are



recurrent but not almost periodic, i.e. words in which each factor occurs infinitely many times, but where the gap between consecutive occurrences of a factor may be arbitrarily large.

Fix a recurrent but not almost periodic word  $\alpha$  over  $\{0, 1, 2, 3\}$ . Since  $\alpha$  is recurrent, any factor  $\beta$  of  $\alpha$  must occur an infinite number of times, and we will call the factors between any consecutive pair of occurrences of  $\beta$  the  $\beta$ -gap factors. Since  $\alpha$  is not almost periodic, there exists a factor  $\beta$  of length  $k-1$ , say, such that the  $\beta$ -gap factors can be arbitrarily long. Denote the sequence of  $\beta$ -gap factors by  $\gamma_1, \gamma_2, \dots$ . If, amongst these gap factors, we find that for any integer  $m$  there exists some (indeed, infinitely many)  $\gamma_i$  which has at least  $m$  letters that are not 0, then by the analysis in Section 3 there exist graphs whose clique-width grows as a function of  $m$ . Thus, the proper subclass  $\mathcal{C} = \text{Free}(H_{1,j}^\alpha(k, k)) \cap \mathcal{G}^\alpha$  (where  $j$  denotes the start of the first occurrence of  $\beta$  in  $\alpha$ ) contains graphs of arbitrarily large clique-width, and thus  $\mathcal{G}^\alpha$  is not minimal.

Now let  $\Gamma$  denote the collection of all recurrent words  $\alpha$  over  $\{0, 1, 2, 3\}$  other than the all-zeros word, with the property that for any factor  $\beta$  of  $\alpha$ , the weight of every  $\beta$ -gap factor is bounded. We now show that it is precisely the words in  $\Gamma$  that define minimal classes of unbounded clique-width.

**Theorem 5.1.** *Let  $\gamma$  be an infinite sequence over  $\{0, 1, 2, 3\}$  other than the all-zero sequence. Then  $\mathcal{G}^\gamma$  is a minimal hereditary graph class of unbounded clique-width if and only if  $\gamma \in \Gamma$ .*

*Proof.* If  $\mathcal{G}^\gamma$  is a minimal hereditary graph class of unbounded clique-width, and  $\gamma$  is not almost periodic, then from the preamble to Section 5 we have already demonstrated that  $\gamma \in \Gamma$ .

To prove the converse, suppose  $\gamma \in \Gamma$ . In the case that  $\gamma$  is almost periodic, we may appeal directly to Lemma 4.7. For this more general setting, we may proceed in an almost identical manner.

If  $\mathcal{C}$  is a proper hereditary subclass of  $\mathcal{G}^\gamma$  then there must exist a non-trivial finite forbidden graph  $F$  that is in  $\mathcal{G}^\gamma$  but not in  $\mathcal{C}$ . In turn, this graph  $F$  must be an induced subgraph of some  $H_{1,j}^\gamma(k, k)$  for some  $k \in \mathbb{N}$ . Any graph  $G$  in  $\mathcal{C}$  must be  $\text{Free}(H_{1,j}^\alpha(k, k))$  for the fixed value of  $k \geq 2$ .

As before, let  $\beta = \gamma_j \gamma_{j+1} \dots \gamma_{j+k-2}$  and  $G^*$  denote the subgraph of  $G$  induced by the columns  $C_j \dots C_{j+k-1}$ . We can use the same cluster graph arguments to show that there is a partition  $X_{V_G}/Y_{V_G}$  of the vertices of  $G^*$  such that  $\mu_{G^*}(X_{V_G}) \leq 4k^2 - 3k$ . Symmetrically,  $\mu_{G^*}(Y_{V_G}) \leq 4k^2 - 3$ .

We know that the factor  $\beta$  appears an infinite number of times in  $\gamma$  and that the weight of the string between each copy of  $\beta$  is bounded by a constant, say,  $W(\beta)$ .

Suppose the  $i$ -th copy of  $\beta$  in  $\gamma$  generates the subgraph  $G_i^*$  of  $G$ , with corresponding partition  $X_i/Y_i$ , then we define  $U_i$  as the subgraph induced by the vertices of  $Y_{i-1}$ ,  $X_i$  and all the vertices of  $G$  in columns between these two sets.

This gives us a partition of  $G$  such that for every  $i$ :

- (a) by using Corollary 4.2 the clique-width of  $G(U_i)$  is at most  $6(2k + W(\beta)) + 3$ , and
- (b)  $\mu(U_i) \leq (8k^2 - 6k)$  and  $\mu(U_1 \cup \dots \cup U_i) \leq (8k^2 - 6k)$ ,

So from Lemma 4.1 the clique-width of  $G$  is at most  $(6(2k + W(\beta)) + 3) \times (8k^2 - 6k)$ .

But we know that  $k$  and  $W(\beta)$  are fixed dependent on the forbidden graph  $F$  and hence the graph class  $\mathcal{C}$  has bounded clique-width. Thus  $\mathcal{G}^\gamma$  is a minimal hereditary graph class of unbounded clique-width.  $\square$

While  $\Gamma$  includes every periodic and almost periodic word over  $\{0, 1, 2, 3\}$ , it does also contain other (recurrent) words. One simple way to generate such sequences is by substitution. We use [10] as our reference work on substitutions. For example, consider the infinite binary word  $\psi$  generated by an iterative substitution  $\sigma$ , beginning with 1 such that  $\sigma(1) = 1010$  and  $\sigma(0) = 0$ .

If we denote  $\sigma^n(1)$  as the  $n$ -th iteration beginning with  $\sigma^0(1) = 1$  then

$$\sigma^n(1) = \sigma^{n-1}(\sigma(1)) = \sigma^{n-1}(1) 0 \sigma^{n-1}(1) 0.$$

The first four iterates, and the start of  $\psi$ , are as follows.

$$\begin{aligned} \sigma^1(1) &= 1010 \\ \sigma^2(1) &= 1010010100 \\ \sigma^3(1) &= 1010010100010100101000 \\ \sigma^4(1) &= 1010010100010100101000010100101000101001010000 \\ \psi &= 1010010100010100101000010100101000101001010000010100101000\dots \end{aligned}$$

The word  $\psi$  has the following characteristics.

- (i) The number of ones doubles with each iteration and therefore  $\psi$  contains an infinite number of ones.
- (ii)  $\psi$  is a fixed point of  $\sigma$  (i.e.  $\sigma(\psi) = \psi$ ).
- (iii) By construction  $\psi$  is recurrent but is not almost periodic, because it contains arbitrarily long strings of zeros.

The following lemma shows that  $\psi \in \Gamma$ , and therefore provides us with the promised counterexample to the conjecture of Collins et al [3].

**Lemma 5.2.** *For any factor  $\beta$  of the word  $\psi$ , the weight of the  $\beta$ -gap factors is bounded, and thus  $\psi \in \Gamma$ .*

*Proof.* Suppose the longest subfactor of contiguous zeros in  $\beta$  is  $0^k$ . It can be observed that  $\sigma^n(1)$  ends with the factor  $0^n$ . Hence  $\beta$  must have appeared by the  $(k+1)$ -th iteration,  $\sigma^{k+1}(1)$  or it is not a factor of  $\psi$ . Since  $|\sigma^{k+1}(1)|_1 = 2^{k+1}$ , we have this as a bound on the weight between any consecutive occurrences.  $\square$

We can extend this idea to construct other recurrent but not almost periodic infinite binary sequences in  $\Gamma$ . Indeed, any iterative substitution  $\sigma_\gamma$  where  $\sigma_\gamma(1) = \delta$  and  $\sigma_\gamma(0) = 0$  such that  $\delta$  is a finite binary word whose first letter is 1, last letter is 0, and with  $|\delta|_1 \geq 2$  will define a

sequence  $\gamma$ . Now  $|\sigma_\gamma^n(1)|_1 = |\delta|_1^n$  and it follows using Lemma 5.2 that the weight of the set of  $\beta$ -gap factors for every factor  $\beta$  is bounded, and hence  $\gamma \in \Gamma$ .

Finally, notice that  $\Gamma$  does not comprise all recurrent binary sequences. Indeed, for any  $\gamma \in \Gamma$  that is recurrent but not almost periodic, then the sequence  $\bar{\gamma}$ , formed as the complement of  $\gamma$  (i.e. inverting the 1s and 0s), is a recurrent sequence that does not lie in  $\Gamma$ , and so  $\mathcal{G}^{\bar{\gamma}}$  is not a minimal hereditary graph class of unbounded clique-width.

## 6 Concluding remarks

**Linear clique-width** The *linear clique-width* of a graph  $G$  is defined as the minimum number of labels needed to construct  $G$  by means of the operations allowed for standard clique-width, except for the disjoint union operation. Our minimality of unbounded clique-width arguments rest on constructing partitions that satisfy the conditions in Lemma 4.1. In fact, there exists a ‘linear’ analogue of this, see [3, Lemma 3], and it is likely that this may be used in conjunction with our arguments above to show that  $G^\alpha$  for any  $\alpha \in \Gamma$  is also minimal of unbounded linear clique-width.

**Towards a characterisation of clique width for bipartite graphs** While the ultimate goal of characterising which hereditary graph classes have unbounded clique-width remains somewhat remote, a nearer goal is the restriction of this characterisation to cover classes of *bipartite* graphs.

The identification of the collection of words  $\Gamma$  represents a key step towards a fuller classification: even though we now have uncountably many minimal classes, the collection  $\Gamma$  is relatively easily stated, and gives us the precise delineation between minimal and non-minimal for the classes under consideration.

To extend our work to cover all bipartite graphs still faces a number of hurdles. First, there exist minimal classes of bipartite graphs that are not of the form  $\mathcal{G}^\alpha$  for any  $\alpha \in \Gamma$  (for example, the bichain graphs of Atminas, Brignall, Lozin and Stacho [1]), so the current four-letter alphabet  $\{0, 1, 2, 3\}$  is certainly not complete. Second, even with a more complete construction of classes, one must prove that such a list is complete, taking into account the pernicious issue of the class of square grids (which is bipartite and has unbounded clique-width yet contains no minimal class).

**Acknowledgements** We are grateful to Reem Yassawi for helpful discussions concerning recurrent and almost periodic sequences.

## References

- [1] A. Atminas, R. Brignall, V. Lozin, and J. Stacho. Minimal classes of graphs of unbounded clique-width defined by finitely many forbidden induced subgraphs. *Discrete Applied Mathematics*, 295:57–69, 2021.
- [2] A. Brandstädt and V. V. Lozin. On the linear structure and clique-width of bipartite permutation graphs. *Ars Combin.*, 67:273–281, 2003.
- [3] A. Collins, J. Foniok, N. Korpelainen, V. Lozin, and V. Zamaraev. Infinitely many minimal classes of graphs of unbounded clique-width. *Discrete Appl. Math.*, 248:145–152, 2018.
- [4] B. Courcelle, J. A. Makowsky, and U. Rotics. Linear time solvable optimization problems on graphs of bounded clique-width. *Theory Comput. Syst.*, 33(2):125–150, 2000.
- [5] Bruno Courcelle. The monadic second-order logic of graphs. I. Recognizable sets of finite graphs. *Inform. and Comput.*, 85(1):12–75, 1990.
- [6] Bruno Courcelle, Joost Engelfriet, and Grzegorz Rozenberg. Handle-rewriting hypergraph grammars. *J. Comput. System Sci.*, 46(2):218–270, 1993.
- [7] Bruno Courcelle and Stephan Olariu. Upper bounds to the clique width of graphs. *Discrete Appl. Math.*, 101(1-3):77–114, 2000.
- [8] Konrad K. Dabrowski, Matthew Johnson, and Daniël Paulusma. Clique-width for hereditary graph classes. In *Surveys in combinatorics 2019*, volume 456 of *London Math. Soc. Lecture Note Ser.*, pages 1–56. Cambridge Univ. Press, Cambridge, 2019.
- [9] Reinhard Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer, Berlin, fifth edition, 2017.
- [10] N. Pytheas Fogg. *Substitutions in dynamics, arithmetics and combinatorics*, volume 1794 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2002. Edited by V. Berthé, S. Ferenczi, C. Mauduit and A. Siegel.
- [11] Jim Geelen, O joungh Kwon, Rose McCarty, and Paul Wollan. The grid theorem for vertex-minors. *Journal of Combinatorial Theory, Series B*, 2020.
- [12] Martin Charles Golumbic and Udi Rotics. On the clique-width of some perfect graph classes. *Internat. J. Found. Comput. Sci.*, 11(3):423–443, 2000. Selected papers from the Workshop on Theoretical Aspects of Computer Science (WG 99), Part 1 (Ascona).
- [13] Nicholas Korpelainen. A new graph construction of unbounded clique-width. In *TCDM 2016—1st IMA Conference on Theoretical and Computational Discrete Mathematics, University of Derby*, volume 56 of *Electron. Notes Discrete Math.*, pages 31–36. Elsevier Sci. B. V., Amsterdam, 2016.
- [14] Vadim V. Lozin. Minimal classes of graphs of unbounded clique-width. *Ann. Comb.*, 15(4):707–722, 2011.
- [15] Sang-il Oum. Rank-width and vertex-minors. *J. Combin. Theory Ser. B*, 95(1):79–100, 2005.

- [16] Sang-il Oum and Paul Seymour. Approximating clique-width and branch-width. *J. Combin. Theory Ser. B*, 96(4):514–528, 2006.
- [17] Neil Robertson and P. D. Seymour. Graph minors. V. Excluding a planar graph. *J. Combin. Theory Ser. B*, 41(1):92–114, 1986.
- [18] Zhi Xiong Wen and Zhi Ying Wen. Local isomorphisms of invertible substitutions. *C. R. Acad. Sci. Paris Sér. I Math.*, 318(4):299–304, 1994.