Letter Graphs and Well-Quasi-Order by Induced Subgraphs

Marko Petkovšek
Faculty of Mathematics and Physics
Department of Mathematics and Mechanics
University of Ljubljana, Slovenia

Abstract

Given a word w over a finite alphabet and a set of ordered pairs of letters which define adjacencies, we construct a graph which we call the *letter graph* of w. The *lettericity* of a graph G is the least size of alphabet permitting to obtain G as a letter graph. The set of 2-letter graphs consists of threshold graphs, unbounded-interval graphs, and their complements. We determine the lettericity of cycles and bound the lettericity of paths to an interval of length one. We show that the class of k-letter graphs is well-quasi-ordered by the induced subgraph relation, and that it has a finite set of minimal forbidden induced subgraphs. As a consequence, k-letter graphs can be recognized in polynomial time for any fixed k.

1 Introduction

In graph theory, a reflexive and transitive relation is called a *quasi-order*. A quasi-order $\leq$ on X is a *well-quasi-order* if for any infinite sequence $a_1, a_2, \ldots \in X$ there are indices $i < j$ such that $a_i \leq a_j$. Equivalently, X contains no infinite strictly decreasing sequences and no infinite antichains. Yet another equivalent characterization of well-quasi-orders is that every nonempty subset of X has a nonzero finite number of minimal elements (cf. [9, 12]).

By the famous Graph Minor Theorem of N. Robertson and P. D. Seymour, the *graph minor* relation is a well-quasi-order on the class of all graphs. This, however, is not true for the more restrictive relations such as the *topological minor* (or *homeomorphic embeddability*), the *subgraph*, and the *induced subgraph* relations. It is therefore of interest to identify restricted classes of graphs which *are* well-quasi-ordered by these relations. For example, the class of all trees is well-quasi-ordered by the topological minor relation, according to a well-known theorem of J. B. Kruskal [11]. G. Ding has proved that a subgraph ideal (i.e., a class of graphs closed under taking subgraphs) is well-quasi-ordered by the subgraph relation if and only if it contains at most finitely many graphs C_n and F_n (C_n being the cycle on n vertices, and F_n the path on n vertices with two pendant edges attached to each of its endpoints).

Concerning the induced subgraph relation $\leq_i$ that we shall consider here, the following is known. P. Damaschke [3] has proved that P_4-reducible graphs (i.e., graphs in which all induced paths on four vertices are vertex-disjoint) are well-quasi-ordered by $\leq_i$. G. Ding has proved that the following classes of graphs are well-quasi-ordered by $\leq_i$:

- $\mathcal{G}_r$, the class of graphs G such that for some $R \subseteq V(G)$ with $|R| \leq r$, the graph $G - R$ has matroidal number at most three [6],
- any subgraph ideal which is well-quasi-ordered by the subgraph relation [5].

In [3] and [5], several further classes of graphs defined by excluding a finite set of forbidden induced subgraphs have been shown well-quasi-ordered by $\leq_i$.

In this paper we present another family of induced-subgraph ideals which are well-quasi-ordered by $\leq_i$. Given a word w over a finite alphabet and a set of ordered pairs of letters which define adjacencies, we construct a graph which we call the *letter graph* of w. The *lettericity* of a graph G is the least size of alphabet permitting to obtain G as a letter graph. In Section 3 we state some basic properties of k-letter graphs. The class of 2-letter graphs is described completely in Section 4: it is composed of threshold graphs, unbounded-interval graphs, and their complements. In Section 5 we determine the lettericity of cycles and paths (the latter only to within an interval of length one) and show that for large n there are n-vertex graphs whose lettericity exceeds $0.707\,n$. In Section 6 we show that the class of k-letter graphs is well-quasi-ordered by $\leq_i$ and has a finite set of minimal forbidden induced subgraphs. As a consequence, for any fixed k the class of k-letter graphs can be recognized in polynomial time.

2 Definitions and notation

Our graphs are undirected and simple. We write $x \sim_G y$ if x and y are adjacent vertices of G. As a set of pairs, the adjacency relation in $V(G)$ is denoted by Adj_G. The complement of a graph G is denoted by $\overline{G}$. If $\mathcal{A}$ is a set of graphs we write $\overline{\mathcal{A}}$ for the set $\{\overline{G};\ G \in \mathcal{A}\}$. The disjoint union of G_1 and G_2 is denoted by $G_1 + G_2$, and the disjoint union of n copies of G is denoted by nG. As usual, K_n denotes the complete graph on n vertices, $K_{p,q}$ the complete bipartite graph on $p+q$ vertices, P_n the path on n vertices, and C_n the cycle of length n. The vertex set of P_n is $\{1, 2, \ldots, n\}$, with $i \sim_{P_n} (i+1)$ for $i = 1, 2, \ldots, n-1$. The vertex set of C_n is $\{0, 1, \ldots, n-1\}$, with $i \sim_{C_n} (i+1) \bmod n$ for $i = 0, 1, \ldots, n-1$. If $\mathcal{A}$ is a set of graphs closed under taking induced subgraphs we denote by $Obs(\mathcal{A})$ the set of *obstructions* or *minimal forbidden induced subgraphs* for $\mathcal{A}$ (i.e., the minimal elements of the complement of $\mathcal{A}$ quasi-ordered by the induced subgraph relation). The isomorphism relation among graphs is denoted by $\cong$. By $z(G)$ we denote the *cochromatic number* of G, which is the minimum cardinality of a partition of $V(G)$ into subsets that are either a clique or an independent set.

Let Σ be a finite alphabet and Σ^* the set of all words over Σ (i.e., the free monoid generated by Σ under concatenation). For a word $w = s_1 s_2 \ldots s_n \in \Sigma^*$ where $s_i \in \Sigma$, let $w^R = s_n s_{n-1} \ldots s_1$ denote its reverse. If $\mathcal{A}$ is a set of words we write $\mathcal{A}^R$ for $\{w^R;\ w \in \mathcal{A}\}$.

Let $\mathcal{P} \subseteq \Sigma^2$ be a fixed set of ordered pairs of symbols from Σ. To each word $w = w_1 w_2 \dots w_n$ where $w_i \in \Sigma$ we assign its *letter graph* $G(\mathcal{P}, w)$ in the following way:

$$\begin{aligned} V(G(\mathcal{P}, w)) &= \{1, 2, \dots, n\}, \\ E(G(\mathcal{P}, w)) &= \{\{i, j\}\ ;\ w_{\min(i,j)} w_{\max(i,j)} \in \mathcal{P}\}. \end{aligned}$$

The vertices of $G(\mathcal{P}, w)$ are naturally labelled with the symbols of w.

Example 1 *Take $\Sigma = \{a, b, c\}$, $\mathcal{P} = \{ac, ba, cb, bb\}$, and $w = abcabc$. The corresponding letter graph $G(\mathcal{P}, w)$ is shown in Fig. 1 where vertex i is labelled with s_i. In this case, $G(\mathcal{P}, w)$ is the 6-cycle C_6.*

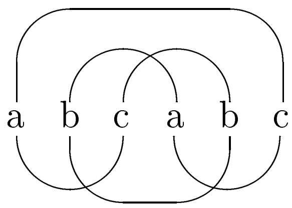

Figure 1: C_6 as a 3-letter graph ($\mathcal{P} = \{ac, ba, cb, bb\}$).

Denote

$$\mathcal{G}_\Sigma(\mathcal{P}) = \{G(\mathcal{P}, w);\ w \in \Sigma^*\},$$

$$\mathcal{G}_k = \bigcup_{|\Sigma|=k,\, \mathcal{P} \subseteq \Sigma^2} \mathcal{G}_\Sigma(\mathcal{P}),$$

$$l(G) = \min\{k;\ G \in \mathcal{G}_k\}.$$

Thus $\mathcal{G}_k$ is the set of all graphs that are letter graphs over some alphabet of size k, and $l(G)$ is the least alphabet size that suffices to represent G as a letter graph. The graphs from $\mathcal{G}_k$ will be called *k-letter graphs*, and $l(G)$ the *lettericity* of G. Example 1 shows that $l(C_6) \leq 3$.

3 Some properties of k-letter graphs

First we restate the definition of k-letter graphs in purely graph-theoretic terms.

Proposition 1 *A graph G is a k-letter graph if and only if*

1. *there is a partition $V_1, V_2, \dots, V_p$ of $V(G)$ with $p \leq k$ such that each V_i is either a clique or an independent set in G, and*

2. *there is a linear ordering L of $V(G)$ such that for each pair of indices $1 \leq i, j \leq p$, $i \neq j$, the intersection of Adj_G with $V_i \times V_j$ is one of*

(a) $L \cap (V_i \times V_j)$,

(b) $L^{-1} \cap (V_i \times V_j)$,

(c) $V_i \times V_j$, *or*

(d) $\emptyset$.

Proof: If G is a k-letter graph then $G = G(\mathcal{P}, w)$ for some $\mathcal{P} \subseteq \Sigma^2$ and $w \in \Sigma^*$ where $|\Sigma| = k$. Let $a_1, a_2, \ldots, a_p$ be the different symbols from Σ that actually appear in w. Define

$$V_i = \{v \in V(G);\ w_v = a_i\} \quad (1 \leq i \leq p).$$

If $a_i a_i \in \mathcal{P}$ then V_i is a clique, otherwise it is an independent set. Let L be the order induced on the vertices of $V(G)$ by the linear ordering of their labels in w, and $1 \leq i \neq j \leq p$. We distinguish four cases:

(a) $a_i a_j \in \mathcal{P}$, $a_j a_i \notin \mathcal{P}$: If $x \in V_i$ and $y \in V_j$ then $x \sim y$ if and only if xLy, so $Adj_G \cap (V_i \times V_j) = L \cap (V_i \times V_j)$.

(b) $a_i a_j \notin \mathcal{P}$, $a_j a_i \in \mathcal{P}$: If $x \in V_i$ and $y \in V_j$ then $x \sim y$ if and only if yLx, so $Adj_G \cap (V_i \times V_j) = L^{-1} \cap (V_i \times V_j)$.

(c) $a_i a_j \in \mathcal{P}$, $a_j a_i \in \mathcal{P}$: In this case $x \sim y$ for all $x \in V_i$ and $y \in V_j$, so $Adj_G \cap (V_i \times V_j) = V_i \times V_j$.

(d) $a_i a_j \notin \mathcal{P}$, $a_j a_i \notin \mathcal{P}$: In this case $x \not\sim y$ for all $x \in V_i$ and $y \in V_j$, so $Adj_G \cap (V_i \times V_j) = \emptyset$.

Conversely, let G be a graph on n vertices which satisfies conditions 1 and 2. Take $\Sigma = \{a_1, a_2, \ldots, a_p\}$ and

$$\begin{aligned} \mathcal{P} &= \{a_i a_i;\ V_i \text{ clique}\} \\ &\cup\ \{a_i a_j;\ i \neq j,\ Adj_G \cap (V_i \times V_j) = L \cap (V_i \times V_j)\} \\ &\cup\ \{a_j a_i;\ i \neq j,\ Adj_G \cap (V_i \times V_j) = L^{-1} \cap (V_i \times V_j)\} \\ &\cup\ \{a_i a_j, a_j a_i;\ i \neq j,\ Adj_G \cap (V_i \times V_j) = V_i \times V_j\}. \end{aligned}$$

Number the vertices of G so that $v_1 \, L \, v_2 \, L \, \ldots \, L \, v_n$, and define $w = s(v_1)s(v_2)\ldots s(v_n)$ where $s(x) = a_i$ if $x \in V_i$. We claim that the mapping $v_i \mapsto i$ is an isomorphism from G to $H = G(\mathcal{P}, w)$.

Let $x = v_l \in V_i$ and $y = v_m \in V_j$. First assume that $x \sim_G y$. If $i = j$ then $V_i = V_j$ must be a clique in G, so $a_i a_j = a_i a_i \in \mathcal{P}$ and hence $l \sim_H m$. If $i \neq j$ we distinguish four cases corresponding to those in condition 2:

(a) $Adj_G \cap (V_i \times V_j) = L \cap (V_i \times V_j)$: In this case $a_i a_j \in \mathcal{P}$. As $x \, L \, y$, we have $l < m$ and hence $l \sim_H m$.

(b) $Adj_G \cap (V_i \times V_j) = L^{-1} \cap (V_i \times V_j)$: In this case $a_j a_i \in \mathcal{P}$. As $x \, L^{-1} \, y$, we have $l > m$ and hence $l \sim_H m$.

(c) $Adj_G \cap (V_i \times V_j) = V_i \times V_j$: In this case $a_i a_j \in \mathcal{P}$ and $a_j a_i \in \mathcal{P}$, hence $l \sim_H m$.

(d) $Adj_G \cap (V_i \times V_j) = \emptyset$: This case is impossible because by assumption, $(x, y) \in Adj_G \cap (V_i \times V_j)$.

Now assume that $l \sim_H m$ and, w.l.o.g., that $l < m$. Then $a_i a_j \in \mathcal{P}$ and $x \, L \, y$. If $i = j$ then $V_i = V_j$ is a clique in G, so $x \sim_G y$. If $i \neq j$ we distinguish three cases corresponding to those in the definition of $\mathcal{P}$:

(a) $Adj_G \cap (V_i \times V_j) = L \cap (V_i \times V_j)$: As $x \, L \, y$, it follows that $x \sim_G y$.

(b) $Adj_G \cap (V_j \times V_i) = L^{-1} \cap (V_j \times V_i)$: As $y \, L^{-1} \, x$, it follows that $x \sim_G y$.

(c) $Adj_G \cap (V_i \times V_j) = V_i \times V_j$: In this case $x \sim_G y$ for all $x \in V_i$ and $y \in V_j$. □

Corollary 1 *Let G be a k-letter graph. Then $V(G)$ can be partitioned into $p \leq k$ sets $V_1, V_2, \ldots, V_p$ each of which is either a clique or an independent set in G, such that for each pair of indices $1 \leq i, j \leq p$, $i \neq j$, the family of neighborhoods $N_j(x) = \{y \in V_j;\ x \sim_G y\}$ of all $x \in V_i$ forms a chain of subsets of V_j.*

Proof: Let L be the linear order on $V(G)$ described in Proposition 1. Pick $x, y \in V_i$ such that $x \, L \, y$. If $Adj_G \cap (V_i \times V_j) = L \cap (V_i \times V_j)$ and $y \sim_G z \in V_j$ then $y \, L \, z$, so $x \, L \, z$ and $x \sim_G z$, hence $N_j(y) \subseteq N_j(x)$. If $Adj_G \cap (V_i \times V_j) = L^{-1} \cap (V_i \times V_j)$ and $x \sim_G z \in V_j$ then $z \, L \, x$, so $z \, L \, y$ and $y \sim_G z$, hence $N_j(x) \subseteq N_j(y)$. If $Adj_G \cap (V_i \times V_j) = V_i \times V_j$ then $N_j(x) = N_j(y) = V_j$. If $Adj_G \cap (V_i \times V_j) = \emptyset$ then $N_j(x) = N_j(y) = \emptyset$. In all four cases, one of $N_j(x)$, $N_j(y)$ is a subset of the other. □

Next we list some simple observations without proof. Let $f : \Sigma_1 \to \Sigma_2$ be a bijection, extended to Σ_1^* as a homomorphism.

Proposition 2 (i) $G(f(\mathcal{P}), f(w)) = G(\mathcal{P}, w)$.

(ii) $G(\mathcal{P}^R, w^R) = G(\mathcal{P}, w)$.

(iii) $G(\Sigma^2 \setminus \mathcal{P}, w) = \overline{G(\mathcal{P}, w)}$.

Corollary 2 (i) $\mathcal{G}_{\Sigma_2}(f(\mathcal{P})) = \mathcal{G}_{\Sigma_1}(\mathcal{P})$.

(ii) $\mathcal{G}_\Sigma(\mathcal{P}^R) = \mathcal{G}_\Sigma(\mathcal{P})$.

(iii) $\mathcal{G}_\Sigma(\Sigma^2 \setminus \mathcal{P}) = \overline{\mathcal{G}_\Sigma(\mathcal{P})}$.

Proposition 3 (i) *If $n = |V(G)| \geq 2$ then $z(G) \leq l(G) \leq n - 1$.*

(ii) *$l(G) = 1$ if and only if $G \cong K_n$ or $G \cong \overline{K_n}$.*

(iii) *$l(\overline{G}) = l(G)$.*

If z is a (not necessarily contiguous) subword of w then $G(\mathcal{P}, z)$ is an induced subgraph of $G(\mathcal{P}, w)$. Hence the set $\mathcal{G}_\Sigma(\mathcal{P})$ is closed under taking induced subgraphs, and therefore has a characterization with forbidden induced subgraphs. The same is true for $\mathcal{G}_k$. Thus lettericity is a monotone parameter w.r.t. the induced subgraph relation.

4 2-letter graphs

By Proposition 1, 2-letter graphs are bipartite, split, or cobipartite graphs. In this section we characterize cobipartite 2-letter graphs as unbounded-interval graphs, and split 2-letter graphs as threshold graphs. We also show how our representation helps enumerate the pairwise nonisomorphic n-vertex graphs in these classes. For a fixed set of pairs $\mathcal{P}$ write $w_1 \sim w_2$ whenever $G(\mathcal{P}, w_1) \cong G(\mathcal{P}, w_2)$. Clearly, this is an equivalence relation in the set Σ^n of words of length n over Σ.

4.1 Unbounded-interval graphs

An *unbounded-interval graph* is the intersection graph of a family of intervals of infinite length on the real line. We denote the set of unbounded-interval graphs by $\mathcal{U}$. Unbounded-interval graphs are studied in [10]. Complements of unbounded-interval graphs are studied in [4].

Example 2 *Let $I_1 = (-\infty, 0]$, $I_2 = [1, \infty)$, $I_3 = (-\infty, 2]$, and $I_4 = [3, \infty)$ (see Fig. 2). The intersection graph of these four intervals is the path P_4, which is therefore an unbounded-interval graph.*

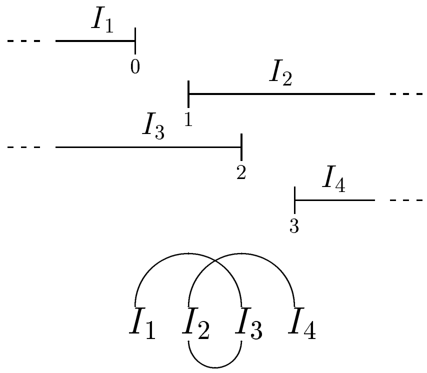

Figure 2: A family of unbounded intervals whose intersection graph is P_4.

The following characterization of unbounded-interval graphs can be found in [10]:

Theorem 1 *For a graph G, the following assertions are equivalent:*

(i) $G \in \mathcal{U}$,

(ii) G *is triangulated and* $\overline{G}$ *is bipartite,*

(iii) G *has no induced subgraphs isomorphic to* $\overline{K_3}$, C_4, *or* C_5,

(iv) $G \in \mathcal{G}_\Sigma(\mathcal{P})$ *where* $\Sigma = \{L, R\}$ *and* $\mathcal{P} = \{LL, RR, RL\}$.

In (iv), vertices corresponding to intervals unbounded on the left (resp. right) are labelled a (resp. b). Fig. 2 shows the example $G(\mathcal{P}, LRLR) \cong P_4$.

Let $f(w) = w^R|_{L\leftrightarrow R}$ be the word obtained by reversing w and swapping L's and R's. Let $\to$ be a rewrite relation defined by

$$\begin{aligned} w &\to f(w), \\ wL &\leftrightarrow Rw. \end{aligned}$$

It turns out that the reflexive-transitive closure of $\to$ in Σ^n coincides with the equivalence relation $\sim$ defined at the beginning of the section. This fact is used in [10] to show that the number of nonisomorphic n-vertex unbounded-interval graphs is $2^{n-2} + 2^{\lfloor n/2 \rfloor -1}$.

4.2 Threshold graphs

A graph G is called *threshold* if there is a labelling f of its vertices by nonnegative integers, and an integer threshold t such that a set $X \subseteq V(G)$ is independent if and only if $\sum_{v\in X} f(v) \leq t$. We denote the set of threshold graphs by $\mathcal{T}$.

Threshold graphs were introduced by Chvátal and Hammer in [1] where the following theorem is proved (see also [2], [7]):

Theorem 2 *For a graph G, the following assertions are equivalent:*

(i) $G \in \mathcal{T}$,

(ii) *G has no induced subgraphs isomorphic to $\overline{C_4}$, C_4, or P_4,*

(iii) *if $\delta_1 < \delta_2 < \cdots < \delta_m$ are the degrees of the nonisolated vertices of G, $\delta_0 = 0$, D_i is the set of all vertices of degree δ_i, $x \in D_i$, $y \in D_j$, and $x \neq y$, then x is adjacent to y iff $i + j > m$.*

Here we characterize threshold graphs as 2-letter graphs.

Theorem 3 *$\mathcal{T} = \mathcal{G}_\Sigma(\mathcal{P})$ where $\Sigma = \{C, S\}$ and $\mathcal{P} = \{CC, CS\}$.*

Proof: Consider a word $w \in \Sigma^*$, partitioned into blocks of successive C's and S's:

$$w = S^{p_0}C^{q_0}S^{p_1}C^{q_1}\cdots S^{p_k}C^{q_k},$$

where $p_0, q_k \geq 0$, $p_1, \ldots, p_k, q_0, \ldots, q_{k-1} > 0$. Let $G = G(\mathcal{P}, w)$. By changing the last letter of w if necessary, we can assume that the last nonempty block of w has length at least two. As both C and S have identical sets of left neighbors in P, such change does not affect G. Let D_i be the set of vertices of G corresponding to the i-th block of S's in w, and D_{m-i} the set of vertices corresponding to the i-th block of C's where m is the total number of nonempty blocks in the subword $C^{q_0}S^{p_1}C^{q_1}\cdots S^{p_k}C^{q_k}$. It is straightforward to verify that: 1) vertices within D_i have identical degree, say δ_i, 2) $0 = \delta_0 < \delta_1 < \cdots < \delta_m$, and 3) distinct vertices $x \in D_i$ and $y \in D_j$ are adjacent iff $i + j > m$. By Theorem 2(iii), G is a threshold graph.

Conversely, let G be a threshold graph. Partition $V(G)$ into D_0, D_1, …, D_m as described in Theorem 2(iii), and let $d_i = |D_i|$, $x = S$ if m is even, $x = C$ if m is odd, and $w = S^{d_0}C^{d_m}S^{d_1}C^{d_{m-1}}\cdots x^{d_{\lceil m/2\rceil}}$. It is straightforward to verify that $G \cong G(\mathcal{P}, w)$. $\square$

Let $\to$ be a rewrite relation defined by

$$wC \quad \leftrightarrow \quad wS.$$

It is easy to see that the reflexive-transitive closure of $\to$ in Σ^n coincides with the equivalence relation $\sim$ defined at the beginning of the section. From this it follows immediately that the number of nonisomorphic n-vertex threshold graphs is 2^{n-1}.

4.3 An overview of 2-letter graphs

Theorem 4 $\mathcal{G}_2 = \mathcal{T} \cup \mathcal{U} \cup \overline{\mathcal{U}}$.

Proof: Table 1 gives an overview of the 16 possible classes of 2-letter graphs over $\Sigma = \{a, b\}$, their minimal forbidden induced subgraphs, and their census. As $K_p, \overline{K_p}, K_p + qK_1, \overline{K_p + qK_1} \in \mathcal{T}$, $K_p + K_q \in \mathcal{U}$, and $K_{p,q} \in \overline{\mathcal{U}}$, the theorem follows. $\square$

elements of $\mathcal{P}$	$\mathcal{G}_\Sigma(\mathcal{P})$	elements of $Obs(\mathcal{G}_\Sigma(\mathcal{P}))$	number of pairwise nonisomorphic n-vertex graphs in $\mathcal{G}_\Sigma(\mathcal{P})$
–	$\{\overline{K_p}\}$	K_2	1
aa	$\{K_p + qK_1\}$	$P_3, \overline{C_4}$	n
bb	$\{K_p + qK_1\}$	$P_3, \overline{C_4}$	n
ab	$\overline{\mathcal{U}}$	$K_3, \overline{C_4}, C_5$	$2^{n-2} + 2^{\lfloor n/2\rfloor - 1}$
ba	$\overline{\mathcal{U}}$	$K_3, \overline{C_4}, C_5$	$2^{n-2} + 2^{\lfloor n/2\rfloor - 1}$
aa, bb	$\{K_p + K_q\}$	$P_3, \overline{K_3}$	$\lfloor n/2\rfloor + 1$
aa, ab	$\mathcal{T}$	$\overline{C_4}, P_4, C_4$	2^{n-1}
aa, ba	$\mathcal{T}$	$\overline{C_4}, P_4, C_4$	2^{n-1}
bb, ab	$\mathcal{T}$	$\overline{C_4}, P_4, C_4$	2^{n-1}
bb, ba	$\mathcal{T}$	$\overline{C_4}, P_4, C_4$	2^{n-1}
ab, ba	$\{K_{p,q}\}$	$\overline{P_3}, K_3$	$\lfloor n/2\rfloor + 1$
aa, bb, ab	$\mathcal{U}$	$\overline{K_3}, C_4, C_5$	$2^{n-2} + 2^{\lfloor n/2\rfloor - 1}$
aa, bb, ba	$\mathcal{U}$	$\overline{K_3}, C_4, C_5$	$2^{n-2} + 2^{\lfloor n/2\rfloor - 1}$
aa, ab, ba	$\overline{K_p + qK_1}$	$\overline{P_3}, C_4$	n
bb, ab, ba	$\overline{K_p + qK_1}$	$\overline{P_3}, C_4$	n
aa, bb, ab, ba	$\{K_p\}$	$\overline{K_2}$	1

Table 1: 2-letter graphs (p and q denote nonnegative integers).

Corollary 3 *All graphs on four or fewer vertices are 2-letter graphs.*

Proof: According to Theorem 2(ii), all graphs on four or fewer vertices except $\overline{C_4}, P_4$, and C_4 are threshold graphs. As $\overline{C_4} \in \mathcal{U}$, $C_4 \in \overline{\mathcal{U}}$, and $P_4 \in \mathcal{U} \cap \overline{\mathcal{U}}$, the claim follows from Theorem 4. □

Corollary 4 *$Obs(\mathcal{G}_2)$ is finite.*

Proof: From Theorem 4 and Table 1 it follows that the graphs not in $\mathcal{G}_2$ have at least one induced subgraph in each of the sets $\{\overline{C_4}, P_4, C_4\}$, $\{\overline{K_3}, C_4, C_5\}$, and $\{K_3, \overline{C_4}, C_5\}$. Checking all 27 combinations and discarding redundant ones we see that such graphs contain at least one of the following seven sets of induced subgraphs: $\{\overline{C_4}, C_4\}$, $\{\overline{K_3}, \overline{C_4}\}$, $\{\overline{C_4}, C_5\}$, $\{\overline{K_3}, K_3, P_4\}$, $\{P_4, C_5\}$, $\{K_3, C_4\}$, $\{C_4, C_5\}$. Thus a minimal forbidden induced subgraph for $\mathcal{G}_2$ can have at most $3 + 3 + 4 = 10$ vertices. □

Corollary 5 *2-letter graphs can be recognized in polynomial time.*

Proof: This follows from Theorem 4 because each of the classes $\mathcal{T}, \mathcal{U}, \overline{\mathcal{U}}$ has a polynomial-time recognition algorithm. □

5 Lettericity of some n-vertex graphs

In this section we consider the lettericity of cycles, paths, and perfect matchings. By a counting argument we show that for large n there are n-vertex graphs whose lettericity exceeds $0.707\,n$.

5.1 Cycles

Call an independent set S in C_n *tight* if $S = \{k, (k+2) \bmod n, \ldots, (k+2m) \bmod n\}$ for some $k \in \{0, 1, \ldots, n-1\}$ and $m \geq 0$.

Lemma 1 *Let $G(\mathcal{P}, w) \cong C_n$. If $a \in \Sigma$ gives rise to an independent set S of size three or more in $G(\mathcal{P}, w)$ then:*

(i) *S is tight,*

(ii) *$|S| = 3$,*

(iii) *the labels of the two vertices of $G(\mathcal{P}, w)$ which have both neighbors in S are distinct.*

Proof: (i) Let R be a maximal run of consecutive vertices of C_n which are not in S. If R has two or more vertices then the labels of the two vertices of S adjacent to one of the endpoints of R must be the leftmost and the rightmost a's in w. Hence there is at most one such run, meaning that S is tight.

(ii) If S contains more than three vertices, it is tight by (i). W.l.g. assume that $0, 2, 4, 6 \in S$. Then in w, the label of 1 (which is adjacent to 0 and 2, but not adjacent to 4 or 6) must be between the labels of $0, 2$ and $4, 6$, while the label of 3 must be between the labels of $2, 4$ and $0, 6$. As this is impossible, $|S| = 3$.

(iii) By (i) and (ii), S is tight and has three vertices. W.l.g. assume that $S = \{0, 2, 4\}$. If the vertices 1 and 3 are labelled the same, say b, these five vertices correspond to a subword $ababa$ of w where the left b is the label of 3 and forces $ba \in \mathcal{P}$, while the right b is the label of 1 and forces $ab \in \mathcal{P}$. But then 1 and 3 would have degree three or more. It follows that vertices 1 and 3 must be labelled differently. □

Theorem 5 *Let $n \geq 4$. Then $l(C_n) = \lfloor \frac{n+4}{3} \rfloor$.*

Proof: First we prove that at least $\lfloor \frac{n+4}{3} \rfloor$ letters are needed to obtain C_n. Let $C_n \cong G(\mathcal{P}, w)$ where w contains m different letters. As $n \geq 4$, the largest clique in C_n is of size 2. From Lemma 1(ii) it follows that each letter appears at most three times in w. Therefore $n \leq 3m$ and $m \geq \lceil n/3 \rceil$. If $n = 3k+1$ then $\lfloor \frac{n+4}{3} \rfloor = k+1 = \lceil n/3 \rceil$, so the assertion is proved. If $n = 3k$ or $n = 3k-1$ then $\lfloor \frac{n+4}{3} \rfloor = k+1 = \lceil n/3 \rceil + 1$. It remains to show that in the latter two cases $k = \lceil n/3 \rceil$ letters do not suffice.

a) $n = 3k$

Assume that w is a word consisting of k different letters whose letter graph is C_{3k}. By Lemma 1(ii), each letter gives rise to an independent set of size three. By Lemma 1(i) and (iii), the vertices of C_{3k} must be (cyclically) labelled $a_1^1 a_k^3 a_1^2 a_2^1 a_1^3 a_2^2 a_3^1 a_2^3 a_3^2 \ldots a_k^2 a_1^1$ where superscripts distinguish the three occurrences of each letter. It remains to see how these symbols could be arranged linearly in w.

As a_k^3 is adjacent to a_1^1 and a_1^2, while a_2^1 is adjacent to a_1^2 and a_1^3, it follows that a_1^2 must be between a_1^1 and a_1^3 in w. W.l.g. assume that the arrangement of these symbols in w is $a_1^1 a_1^2 a_1^3$. By induction on i it can be shown that a_i^1 precedes a_i^2 which precedes a_i^3 in w, and also that a_{i-1}^1 precedes a_i^1, for $i = 2, 3, \ldots, k$. Hence a_1^1 precedes all three occurrences of a_k in w. However, being adjacent to exactly two of the corresponding vertices this is impossible.

b) $n = 3k-1$

As before, assume that w is a word consisting of k different letters whose letter graph is C_{3k-1}. This is only possible if $k-1$ of the letters give rise to an independent set of size three, and the remaining letter, say a_1, gives rise to either a clique or an independent set of size two. In case of a clique, an independent set bordering on it must have the intervening two vertices labelled the same, contrary to Lemma 1(iii). So a_1 gives rise to an independent set of size two.

By Lemma 1(i) and (iii), the only possible way to label (cyclically) the vertices of C_{3k-1} is $a_1^1 a_k^3 a_2^1 a_1^2 a_2^2 a_3^1 a_2^3 a_3^2 \ldots a_k^2 a_1^1$ where superscripts distinguish different occurrences of each letter. It remains to see how these symbols could be arranged linearly in w. Similarly as in the case a) we can establish that a_i^1 precedes a_i^2 which precedes a_i^3 in w, for $i = 2, 3, \ldots, k$, and also that a_{i-1}^1 precedes a_i^1, for $i = 3, 4, \ldots, k$. Hence a_2^1 precedes all three occurrences of a_k in w. However, being adjacent to exactly one of the corresponding vertices this is impossible.

It remains to construct C_n using no more than $\lfloor \frac{n+4}{3} \rfloor$ letters. We distinguish three cases w.r.t. n mod 3. In all three cases, the alphabet is $\Sigma = \{a_0, a_1, \ldots, a_k\}$ where $k = \lfloor \frac{n+1}{3} \rfloor$. Let $\mathcal{P}_c = \{a_i a_{i-1 \pmod{k+1}}\, ;\ 0 \leq i \leq k\}$.

a) $n = 3k+1$

Take $\mathcal{P} = \mathcal{P}_c$ and $w = a_0^1 a_1^1 \cdots a_k^1 a_0^2 a_1^2 \cdots a_k^2 a_0^3 a_1^3 \cdots a_{k-2}^3$ where superscripts are added for easier reference. Write $t_i = a_i^1 a_{i-1}^2 a_{i-2}^3$. Then it is easy to check that $G(\mathcal{P}, w)$ is the cycle $t_k t_{k-1} \cdots t_2 a_1^1 a_0^2 a_k^2 a_0^1 a_k^1$ of length $3k+1$.

b) $n = 3k$

Take $\mathcal{P} = \mathcal{P}_c \cup \{a_{k-1} a_{k-1}\}$ and $w = a_0^1 a_1^1 \cdots a_k^1 a_0^2 a_1^2 \cdots a_k^2 a_0^3 a_1^3 \cdots a_{k-3}^3$. As before, write $t_i = a_i^1 a_{i-1}^2 a_{i-2}^3$. Then it is easy to check that $G(\mathcal{P}, w)$ is the cycle $t_{k-1} t_{k-2} \cdots t_2 a_1^1 a_0^2 a_k^2 a_0^1 a_k^1 a_{k-1}^2 a_{k-1}^1$ of length $3k$. For $n = 6$ this construction is shown in Fig. 1 (with $a_0 = a, a_1 = b, a_2 = c$).

c) $n = 3k - 1$

If $k = 2$ take $\mathcal{P} = \{a_0 a_2, a_1 a_0, a_2 a_1, a_1 a_1, a_0 a_0\}$ and $w = a_0 a_1 a_2 a_0 a_1$. Then $G(\mathcal{P}, w) \cong C_5$. If $k \geq 3$ let $\mathcal{P} = \mathcal{P}_c \cup \{a_{k-1} a_{k-1}, a_{k-2} a_{k-2}\}$ and $w = a_0^1 a_1^1 \cdots a_k^1 a_0^2 a_1^2 \cdots a_k^2 a_0^3 a_1^3 \cdots a_{k-4}^3$. Write again $t_i = a_i^1 a_{i-1}^2 a_{i-2}^3$. Then it is easy to check that $G(\mathcal{P}, w)$ is the cycle $t_{k-2} t_{k-3} \cdots t_2 a_1^1 a_0^2 a_k^2 a_0^1 a_k^1 a_{k-1}^2 a_{k-1}^1 a_{k-2}^2 a_{k-2}^1$ of length $3k - 1$. $\square$

5.2 Paths

Lemma 2 *Let $G(\mathcal{P}, w) \cong P_n$. If $a \in \Sigma$ gives rise to an independent set S of size three or more in $G(\mathcal{P}, w)$ then S is of one of the following types:*

(a) $\{1, 3, n-2, n\}$,

(b) $\{1, 3, i\}$, *where* $6 \leq i \leq n$,

(c) $\{i, n-2, n\}$, *where* $1 \leq i \leq n-5$,

(d) $\{i, i+2, i+4\}$, *where* $2 \leq i \leq n-5$.

Proof: Similar to that of Lemma 1. $\square$

Theorem 6 $\lfloor \frac{n+1}{3} \rfloor \leq l(P_n) \leq \lfloor \frac{n+4}{3} \rfloor$.

Proof: For the upper bound, we show how to construct P_n using no more than $\lfloor \frac{n+4}{3} \rfloor$ letters. We distinguish two cases w.r.t. n mod 3.

a) $n = 3k + 1$

Let $\Sigma = \{a_0, a_1, \ldots, a_k\}$, $\mathcal{P} = \{a_i a_{i-1 \pmod{k+1}};\ 0 \leq i \leq k-1\}$, and $w = a_0^1 a_1^1 \cdots a_k^1 a_0^2 a_1^2 \cdots a_k^2 a_0^3 a_1^3 \cdots a_{k-2}^3$ where superscripts are added for easier reference. Write $t_i = a_i^2 a_{i-1}^3 a_i^1$. Then it is easy to check that $G(\mathcal{P}, w)$ is the path $t_{k-1} t_{k-2} \cdots t_1 a_0^2 a_k^2 a_0^1 a_k^1$ of length $3k + 1$.

b) $n = 3k$ or $n = 3k - 1$

By Theorem 5, C_{n+1} can be constructed using $k+1$ letters. The same then goes for P_n as it is an induced subgraph of C_{n+1}.

For the lower bound, let $P_n \cong G(\mathcal{P}, w)$ where w contains m different letters. Lemma 2 implies that at most one letter can appear four times in w, while the rest can appear three times at most. Therefore $n \leq 4 + 3(m-1)$, so $m \geq \lceil (n-1)/3 \rceil = \lfloor (n+1)/3 \rfloor$. $\square$

Conjecture: If $n \geq 3$ then $l(P_n) = \lfloor \frac{n+4}{3} \rfloor$.

5.3 Maximum lettericity of n-vertex graphs

Let $l(n)$ denote the maximum lettericity of an n-vertex graph. Clearly, $l(1) = l(2) = 1$ and $l(3) = l(4) = 2$. As $l(G) \geq z(G)$, the maximum cochromatic number of an n-vertex graph (which is known to be of order $n/\log n$ [8]) constitutes a lower bound for $l(n)$. But this is a poor bound: we have seen that the lettericity of paths and cycles on n vertices is about $n/3$ which is much larger than $n/\log n$ when n is large. It is also easy to see that $l(nK_2) = n$ and $l(nK_2 + K_1) = n + 1$, so, in fact, $l(n) \geq n/2$. By a counting argument we now improve this bound to $l(n) > 0.707\,n$, provided that n is large enough.

Theorem 7 *For each $\alpha < \frac{\sqrt{2}}{2}$ there is an N such that for all $n > N$ there are n-vertex graphs G with $l(G) > \alpha n$.*

Proof: Assume that $l(G) \leq \alpha\, n$ for all graphs G on n vertices. Write $k = \lfloor \alpha\, n \rfloor$; then, by our assumption, all graphs on n vertices are k-letter graphs. There are $2^{\binom{n}{2}}$ labelled graphs on n vertices. Over a k-letter alphabet, there are k^2 pairs of letters, 2^{k^2} sets of pairs of letters, k^n words of length n, and at most $n!$ possible labellings of a graph on n vertices, hence there are no more than $n!\, k^n 2^{k^2}$ labelled k-letter graphs on n vertices. Therefore

$$2^{\binom{n}{2}} \;\leq\; n!\, k^n 2^{k^2} \;\leq\; n^n (\alpha\, n)^n 2^{(\alpha\, n)^2}.$$

Taking base 2 logarithms we have

$$\left(\frac{1}{2} - \alpha^2\right) n^2 \;\leq\; 2\, n \lg n + \left(\frac{1}{2} + \lg \alpha\right) n.$$

Since $1/2 > \alpha^2$ this is impossible when n is large. □

As for a simple upper bound, Proposition 3(i) implies that $l(n) \leq n - 1$ when $n \geq 2$. It is also not difficult to see that $l(n) \leq n - 2$ when $n \geq 4$.

6 k-letter graphs and well-quasi-order

By deleting a vertex the lettericity of a graph can decrease by more than one: for example, $l(C_5 + K_1) = 4$ but $l(P_4 + K_1) = 2$. We need an upper bound on the extent of this decrease.

Lemma 3 *$l(G) \leq 2\, l(G - v) + 1$ for all $v \in V(G)$.*

Proof: Let $l(G - v) = k$. Then $G - v = G(\mathcal{P}, w)$ for some $\mathcal{P} \subseteq \Sigma^2$ and $w \in \Sigma^*$ where $\Sigma = \{a_1, \ldots, a_k\}$. Let $a_{i_1}, \ldots, a_{i_r}$ be the labels of the neighbors of v in w. Take $\Sigma' = \Sigma \cup \{a'_1, \ldots, a'_k, b\}$ where $a'_1, \ldots, a'_k, b$ are new symbols, and $\mathcal{P}' = \mathcal{P} \cup \{a'_j a'_l, a'_j a_l, a_j a'_l;\ a_j a_l \in \mathcal{P}\} \cup \{a'_{i_j} b;\ 1 \leq j \leq r\}$. Denote by w' the word obtained from w by replacing the labels a_i of the neighbors of v by a'_i. Then $G = G(\mathcal{P}', w'b)$. Hence $l(G) \leq |\Sigma'| = 2k + 1$ as claimed. □

Theorem 8 *The class $\mathcal{G}_k$ of k-letter graphs is well-quasi-ordered by the induced subgraph relation.*

Proof: Fix an alphabet Σ of cardinality k and a set of pairs $\mathcal{P} \subseteq \Sigma^2$. By Higman's Lemma [9, Thm. 4.4], Σ^* is well-quasi-ordered by the (not necessarily contiguous) subword relation. Clearly, z is a subword of w if and only if $G(\mathcal{P}, z)$ is an induced subgraph of $G(\mathcal{P}, w)$, hence $\mathcal{G}_\Sigma(\mathcal{P})$ is well-quasi-ordered by the induced subgraph relation. As $\mathcal{G}_k$ is a union of finitely many sets of the form $\mathcal{G}_\Sigma(\mathcal{P})$ (one for each of the 2^{k^2} possible $\mathcal{P}$'s) the conclusion follows. □

Theorem 9 *The sets of obstructions $Obs(\mathcal{G}_\Sigma(\mathcal{P}))$ and $Obs(\mathcal{G}_k)$ are finite.*

Proof: If $G \in Obs(\mathcal{G}_\Sigma(\mathcal{P}))$ then $G - v \in \mathcal{G}_\Sigma(\mathcal{P})$ for each $v \in V(G)$. Let $k = |\Sigma|$. By Lemma 3, $l(G) \leq 2k+1$, hence $Obs(\mathcal{G}_\Sigma(\mathcal{P})) \subseteq \mathcal{G}_{2k+1}$. As $Obs(\mathcal{G}_\Sigma(\mathcal{P}))$ is an antichain, Theorem 8 implies that it is finite.

Finiteness of $Obs(\mathcal{G}_k)$ is proved in the same way. □

Corollary 6 *The graphs from $\mathcal{G}_\Sigma(\mathcal{P})$ and $\mathcal{G}_k$ are recognizable in polynomial time.*

Proof: The relation $H \leq_i G$ is decidable in time $O(n^m)$ where $n = |V(G)|$ and $m = |V(H)|$. For fixed H this is polynomial in n. Thus by Theorem 9, checking that $H \not\leq_i G$ for all $H \in Obs(\mathcal{G}_\Sigma(\mathcal{P}))$ (resp. $H \in Obs(\mathcal{G}_k)$) where G is given is a polynomial-time recognition algorithm for $\mathcal{G}_\Sigma(\mathcal{P})$ (resp. $\mathcal{G}_k$). □

Note that the proof of Corollary 6 is nonconstructive as the specification of the algorithm given there is incomplete: the finite sets of obstructions for $\mathcal{G}_\Sigma(\mathcal{P})$ and $\mathcal{G}_k$ that are used by the algorithm are, in general, unknown.

7 Conclusion

We conclude by listing some open problems.

Problem 1. Design efficient algorithms to recognize k-letter graphs for small fixed values of k.

Problem 2. What is the time complexity of finding the lettericity of a given graph?

Problem 3. Find the maximal possible lettericity of an n-vertex graph, and the corresponding extremal graphs.

Acknowledgements

The author is indebted to Bojan Mohar and Tomaž Pisanski for helping out with this paper (in particular, Tomaž suggested that k-letter graphs should be recognizable in polynomial time, and Bojan pointed out Theorem 7). He also wishes to thank the referees for their careful reading of the paper and valuable suggestions.

References

[1] V. Chvátal, P. Hammer: *Set-packing and threshold graphs*, Univ. Waterloo Res. Report, CORR 73-21.

[2] V. Chvátal, P. Hammer: Aggregation of inequalities in integer programming, in: *Studies in integer programming* (Proc. Workshop, Bonn, 1975), *Ann. of Discrete Math.*, Vol. 1, North-Holland, Amsterdam 1977, 145–162.

[3] P. Damaschke: Induced subgraphs and well-quasi-ordering, *J. Graph Theory* **14** (1990) 427–435.

[4] G. Ding: Covering the edges with consecutive sets, *J. Graph Theory* **15** (1991) 559–562.

[5] G. Ding: Subgraphs and well-quasi-ordering, *J. Graph Theory* **16** (1992) 489–502.

[6] G. Ding: Stable sets versus independent sets, *Discrete Math.* **117** (1993) 73–87.

[7] M. C. Golumbic: *Algorithmic Graph Theory and Perfect Graphs*, Academic Press, 1980.

[8] P. Erdös, J. Gimbel, D. Kratsch: Some extremal results in cochromatic and dichromatic theory, *J. Graph Theory* **15** (1991) 579–585.

[9] G. Higman: Ordering by divisibility in abstract algebras, *Proc. London Math. Soc.* **2** (1952) 326–336.

[10] S. Klavžar, M. Petkovšek: Intersection graphs of halflines and halfplanes, *Discrete Math.* **66** (1987) 133–137.

[11] J. B. Kruskal: Well-quasi-ordering, the tree theorem, and Vazsonyi's conecture, *Trans. Amer. Math. Soc.* **95** (1960) 210–225.

[12] J. B. Kruskal: The theory of well-quasi-ordering: a frequently discovered concept, *J. Comb. Theory (A)* **13** (1972) 297–305.