New results on word-representable graphs

Andrew Collins[*]   Sergey Kitaev[†]   Vadim V. Lozin[‡]

Abstract

A graph $G = (V, E)$ is word-representable if there exists a word w over the alphabet V such that letters x and y alternate in w if and only if $(x, y) \in E$ for each $x \neq y$. The set of word-representable graphs generalizes several important and well-studied graph families, such as circle graphs, comparability graphs, 3-colorable graphs, graphs of vertex degree at most 3, etc. By answering an open question from [9], in the present paper we show that not all graphs of vertex degree at most 4 are word-representable. Combining this result with some previously known facts, we derive that the number of n-vertex word-representable graphs is $2^{\frac{n^2}{3} + o(n^2)}$.

Keywords: Semi-transitive orientation; Hereditary property of graphs; Speed of hereditary properties

1 Introduction

Let $G = (V, E)$ be a simple (i.e. undirected, without loops and multiple edges) graph with vertex set V and edge set E. We say G is *word-representable* if there exists a word w over the alphabet V such that letters x and y alternate in w if and only if $(x, y) \in E$ for each $x \neq y$.

The notion of word-representable graphs has its roots in the study of the celebrated Perkins semi-group [10, 14]. These graphs possess many attractive properties (e.g. a maximum clique in such graphs can be found in polynomial time), and they provide a common generalization of several important graph families, such as circle graphs, comparability graphs, 3-colorable graphs, graphs of vertex degree at most 3 (see [7] for definitions of these families).

Recently, a number of fundamental results on word-representable graphs were obtained in the literature [8, 9, 11, 12, 13]. In particular, Halldórsson et al. [9] have shown that a graph is word-representable if and only if it admits a *semi-transitive orientation*. However, our knowledge on these graphs is still very limited and many important questions remain open. For example, how hard is it to decide whether a given graph is word-representable or not? What is the minimum length of a word that represents a given graph? How many word-representable graphs on n vertices are there? Does this family include all graphs of vertex degree at most 4?

The last question was originally asked in [9]. In the present paper we answer this question negatively by exhibiting a graph of vertex degree at most 4 which is not word-representable. This result allows us to obtain an upper bound on the asymptotic growth of the number of n-vertex word-representable graphs. Combining this result with a lower bound that follows from some previously known facts, we conclude that the number of n-vertex word-representable graphs is $2^{\frac{n^2}{3} + o(n^2)}$.

---

[*]Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK. Email: A.Collins.2@warwick.ac.uk.

[†]Department of Computer and Information Sciences, University of Strathclyde, Glasgow, G1 1XH, UK. Email: sergey.kitaev@cis.strath.ac.uk

[‡]Mathematics Institute and DIMAP, University of Warwick, Coventry, CV4 7AL, UK. Email: V.Lozin@warwick.ac.uk.

All preliminary information related to the notion of word-representable graphs can be found in Section 2. In Section 3, we prove our negative result about graphs of degree at most 4 and in Section 4, we derive the asymptotic formula on the number of word-representable graphs.

2 Word-representable graphs: definition, examples and related results

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

If a graph is word-representable, then there are infinitely many words representing it. For instance, the complete graph K_4 in Figure 1 can be represented by words 1234, 3142, 123412, 12341234, 432143214321, etc. In general, to represent a complete graph on n vertices, one can start with writing up any permutation of length n and adjoining from either side any number of copies of this permutation.
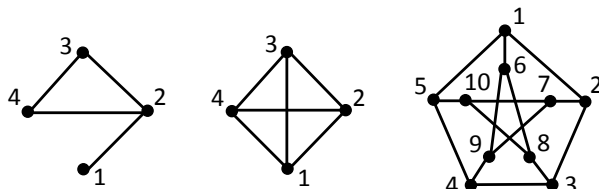

Figure 1: Three word-representable graphs M (left), the complete graph K_4 (middle), and the Petersen graph (right)

If each letter appears exactly k times in a word representing a graph, the graph is said to be *k-word-representable*. It is known [11] that any word-representable graph is k-word-representable for some k. For example, a 3-representation of the Petersen graph shown in Figure 1 is

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

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

A directed graph (digraph) $G = (V, E)$ is *semi-transitive* if it has no directed cycles and for any directed path $v_1 v_2 \cdots v_k$ with $k \geq 4$ and $v_i \in V$, either $v_1 v_k \notin E$ or $v_i v_j \in E$ for all $1 \leq i < j \leq k$. In the second case, when $v_1 v_k \in E$, we say that $v_1 v_k$ is a shortcut. The importance of this notion is due to the following result proved in [9].

Theorem 1 ([9]). *A graph is word-representable if and only if it admits a semi-transitive orientation.*

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

Theorem 2 ([9]). *All 3-colorable graphs are word-representable.*

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

Theorem 3 ([9]). *All graphs of vertex degree at most 3 are word-representable.*

Proof. By Brooks' Theorem, every connected graph of vertex degree at most 3, except for the complete graph K_4, is 3-colorable, and hence word-representable by Theorem 2. Also, as we observed earlier, all complete graphs are word-representable. Therefore, all connected graphs of degree at most 3 and hence all graphs of degree at most 3 are word-representable. □

Whether all graphs of degree at most 4 are word-representable is a natural question following from Theorem 3, which was originally asked in [9]. In the next section, we settle this question negatively.

3   A non-representable graph of vertex degree at most 4

The main result of this section is that the graph A represented in Figure 2 is not word-representable. To prove this, we will show that this graph does not admit a semi-transitive orientation.
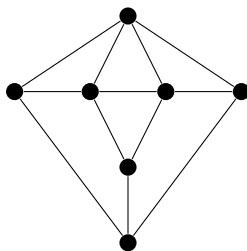
Figure 2: The graph A

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

Lemma 1. *Let D be a K_4-free graph admitting a semi-transitive orientation. Then no cycle of length 4 in this orientation has three consecutively oriented edges.*

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

**Theorem 4.** *The graph A does not admit a semi-transitive orientation.*

Proof. In order to prove that A does not admit a semi-transitive orientation, we will explore *all* orientations of this graph and will show that each choice leads to a contradiction. At each step of the proof we choose a vertex and split the analysis into two cases depending on the orientation of the chosen edge. The chosen edge and its orientation will be shown by a solid arrow ($\longrightarrow$). This choice of orientation may lead to other edges having an orientation assigned to them to satisfy Lemma 1. The orientations forced by the solid arrow through an application of Lemma 1 will be shown by means of double arrows ($\longrightarrow\!\!\!\!\rightarrow$). When we use Lemma 1 to derive a double arrow, we always apply it with respect to a particular cycle of length 4. This cycle will be indicated by four white vertices. Since the graph A has many cycles of length 4, repeated use of Lemma 1 applied to different cycles may lead to a

contradiction, where one more cycle of length 4 has three consecutively oriented edges. We will show that in all possible cases a contradiction of this type arises. The proof is illustrated by diagrams.

Case 1: We start by choosing the orientation for the edge indicated in the diagram below.
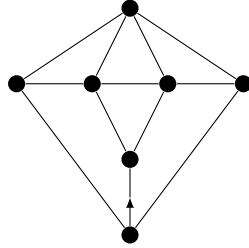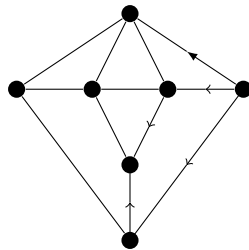
Case 1.1: Now we orient one more edge (solid arrow), which leads to two more orientations being assigned due to Lemma 1 (double arrows).
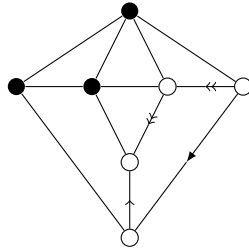
Case 1.1.1: One more edge is oriented (solid arrow) and this choice leads to a contradiction through repeated use of Lemma 1 (a cycle of four white vertices with three consecutively oriented edges in the final of the four diagrams below).
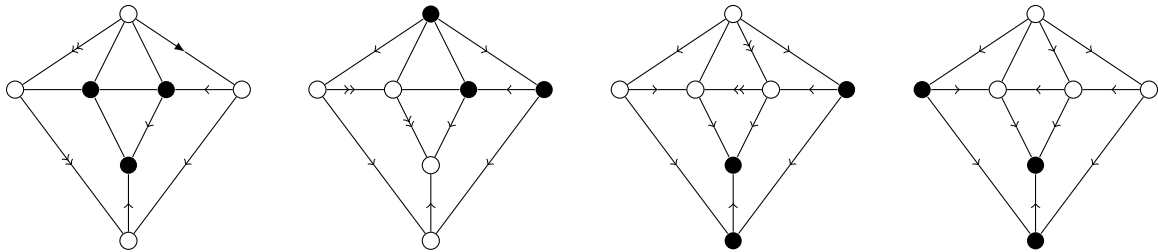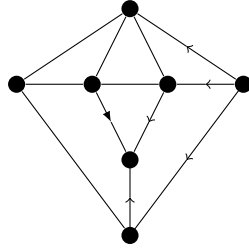
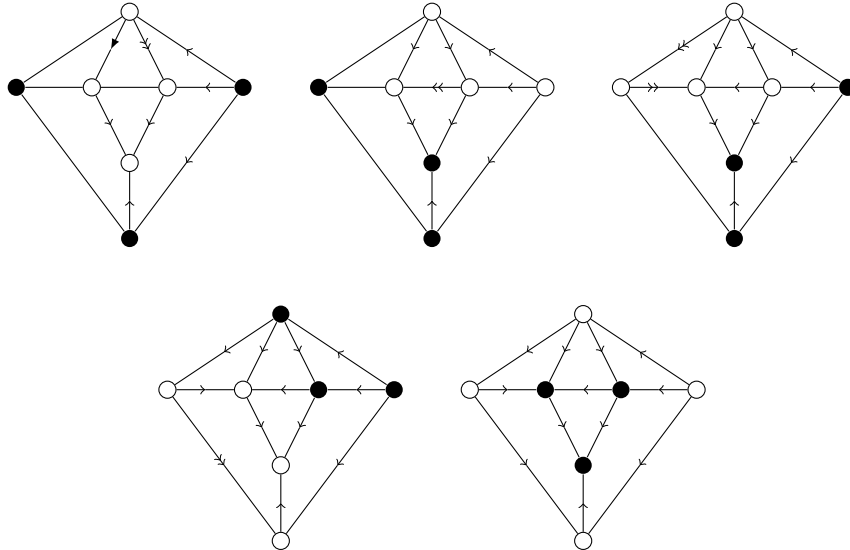*Case* 1.1.2: In this case the orientation of the edge chosen in Case 1.1.1 is reversed.

Case 1.1.2.1: By orienting one more edge (solid arrow) we obtain a contradiction through repeated use of Lemma 1 (a cycle of four white vertices with three consecutively oriented edges in the final diagram).
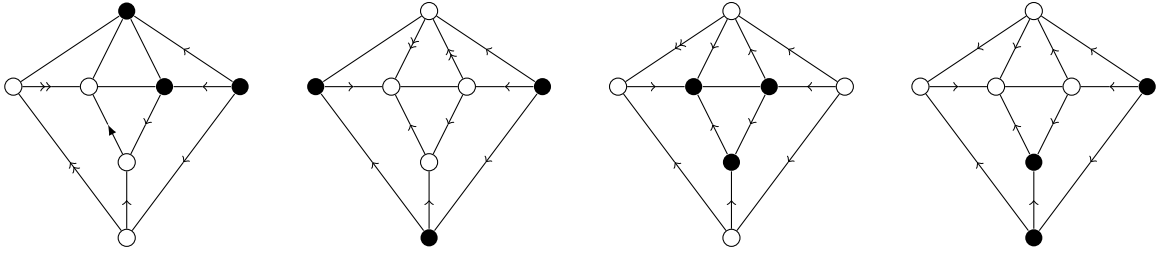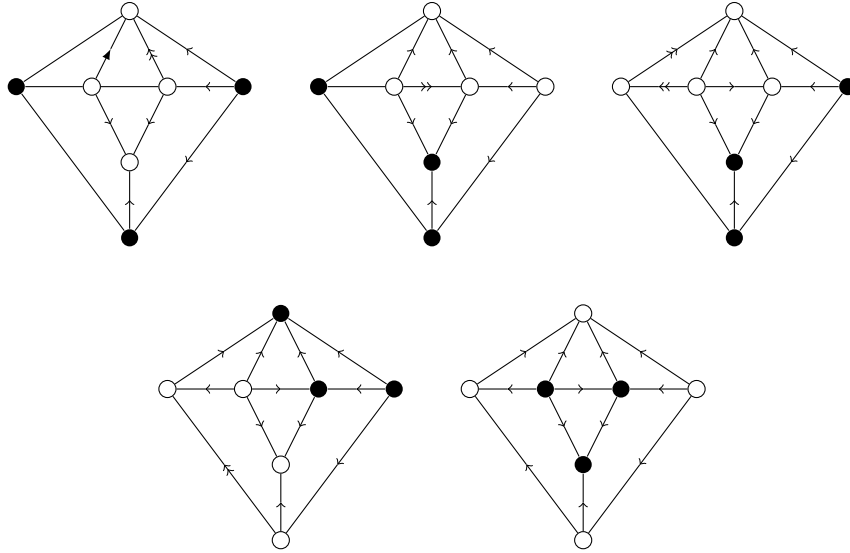
Case 1.1.2.2: Now we reverse the orientation of the edge chosen in Case 1.1.2.1.

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

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

Case 1.2: In this case, we reverse the orientation of the edge chosen in Case 1.1 and obtain a contradiction by analogy with Case 1.1.

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

4   Asymptotic enumeration of word-representable graphs

Given a class X of graphs, we write X_n for the number of graphs in X with vertex set $\{1, 2, \ldots, n\}$, i.e. the number of *labelled* graphs in X. Following [4], we call X_n the *speed* of X.

A class X is *hereditary* if it is closed under taking induced subgraphs. Alternatively, X is hereditary if $G \in X$ implies $G - x \in X$ for any vertex $x \in V(G)$. Clearly, if G is a word-representable graph and w is a word representing G, then for any vertex $x \in V(G)$ the word obtained from w by deleting all appearances of x represents $G - x$. This observation leads to the following obvious conclusion.

Theorem 5. *The class of word-representable graphs is hereditary.*

The speed of hereditary classes and their asymptotic structure have been extensively studied in the literature. In particular, it is known that for every hereditary class X different from the class of all finite graphs,

$$\lim_{n \to \infty} \frac{\log_2 X_n}{\binom{n}{2}} = 1 - \frac{1}{k(X)},$$

where $k(X)$ is a natural number, called the *index* of X. To define this notion let us denote by $\mathcal{E}_{i,j}$ the class of graphs whose vertices can be partitioned into at most i independent sets and j cliques. In particular, $\mathcal{E}_{p,0}$ is the class of p-colorable graphs. Then $k(X)$ is the largest k such that X contains $\mathcal{E}_{i,j}$ with $i + j = k$ for some i and j. This result was obtained independently by Alekseev [2] and Bollobás and Thomason [5, 6] and is known nowadays as the Alekseev-Bollobás-Thomason Theorem (see e.g. [3]).

Since $\binom{n}{2}$ is the minimum number of bits needed to represent an arbitrary n-vertex labeled graph and $\log_2 X_n$ is the minimum number of bits needed to represent an n-vertex labeled graph in the class X, the ratio $\log_2 X_n / \binom{n}{2}$ can be viewed as the coefficient of compressibility for representing graphs in X and its asymptotic value was called by Alekseev [1] the *entropy* of X.

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

Theorem 6.
$$\lim_{n\to\infty} \frac{\log_2 b_n}{\binom{n}{2}} = \frac{2}{3}.$$

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

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

Theorem 7.
$$\lim_{n\to\infty} \frac{\log_2 a_n}{\binom{n}{2}} = \frac{2}{3}.$$

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

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

On the other hand, obviously $b_n \ge a_n$ and hence $\lim_{n\to\infty} \frac{\log_2 b_n}{\binom{n}{2}} \ge \lim_{n\to\infty} \frac{\log_2 a_n}{\binom{n}{2}}$. Combining, we obtain $\lim_{n\to\infty} \frac{\log_2 b_n}{\binom{n}{2}} = \lim_{n\to\infty} \frac{\log_2 a_n}{\binom{n}{2}}$. Together with Theorem 6, this proves the result. $\qquad\square$

Corollary 1.
$$a_n = 2^{\frac{n^2}{3} + o(n^2)}.$$

5 Acknowledgment

References

[1] V.E. ALEKSEEV, Hereditary classes and coding of graphs, *Probl. Cybernet.* 39 (1982) 151–164 (in Russian).

[2] V.E. ALEKSEEV, Range of values of entropy of hereditary classes of graphs, *Diskret. Mat.* 4 (1992), no. 2, 148–157 (in Russian; translation in *Discrete Mathematics and Applications*, 3 (1993), no. 2, 191–199).

[3] N. ALON, J. BALOGH, B. BOLLOBÁS and R. MORRIS, The structure of almost all graphs in a hereditary property, *J. Combin. Theory* B 79 (2011) 131–156.

[4] J. BALOGH, B. BOLLOBÁS and D. WEINREICH, The speed of hereditary properties of graphs, *J. Combin. Theory* B 79 (2000) 131–156.

[5] B. BOLLOBÁS and A. THOMASON, Projections of bodies and hereditary properties of hypergraphs, *Bull. London Math. Soc.*, 27 (1995) 417–424.

[6] B. BOLLOBÁS and A. THOMASON, Hereditary and monotone properties of graphs, "The mathematics of Paul Erdős, II" (R.L. Graham and J. Nesetril, Editors), *Alg. and Combin.*, Vol. 14, Springer-Verlag, New York/Berlin (1997), 70–78.

[7] A. Brandstädt, V.B. Le, J. Spinrad, Graph classes: a survey. SIAM Monographs on Discrete Mathematics and Applications. SIAM, Philadelphia, PA, 1999. xii+304 pp. ISBN: 0-89871-432-X.

[8] M. HALLDÓRSSON, S. KITAEV AND A. PYATKIN, Graphs capturing alternations in words, *Lect. Notes Comput. Sci.* 6224 (2010) 436–437. *Proc. 14th Conf. on Developments in Language Theory*, DLT 2010, London, Ontario, Canada, August 17–20, 2010.

[9] M. HALLDÓRSSON, S. KITAEV AND A. PYATKIN, Alternation graphs, *Lect. Notes Comput. Sci.* 6986 (2011) 191–202. Proceedings of the *37th International Workshop on Graph-Theoretic Concepts in Computer Science*, WG 2011, Teplá Monastery, Czech Republic, June 21–24, 2011.

[10] S. KITAEV AND S. SEIF, Word problem of the Perkins semigroup via directed acyclic graphs, *Order* 25 (2008) 3, 177–194.

[11] S. KITAEV AND A. PYATKIN, On representable graphs, *J. Autom., Lang. and Combin.* 13 (2008) 1, 45–54.

[12] S. KITAEV, P. SALIMOV, C. SEVERS AND H. ÚLFARSSON, On the representability of line graphs, *Lecture Notes in Computer Science* 6795 (2011) 478–479. *Proc. 15th Conf. on Developments in Language Theory*, DLT 2011, University of Milano-Bicocca, Milan, Italy, July 19–22, 2011.

[13] S. KITAEV, P. SALIMOV, C. SEVERS AND H. ÚLFARSSON, Word-representability and line graphs, *Open J. Discr. Math.* 1 (2011) 2, 96–101.

[14] S. SEIF, Monoids with sub-log-exponential free spectra, *J. Pure Appl. Algebra* 212 (2008) 5, 1162–1174.