Word-representability of split graphs

Sergey Kitaev[*], Yangjing Long[†], Jun Ma[‡] and Hehui Wu[§]

September 15, 2020

Abstract

Two letters x and y alternate in a word w if after deleting in w all letters but the copies of x and y we either obtain a word $xyxy\cdots$ (of even or odd length) or a word $yxyx\cdots$ (of even or odd length). 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 $xy \in E$. It is known that a graph is word-representable if and only if it admits a certain orientation called semi-transitive orientation.

Word-representable graphs generalize several important classes of graphs such as 3-colorable graphs, circle graphs, and comparability graphs. There is a long line of research in the literature dedicated to word-representable graphs. However, almost nothing is known on word-representability of split graphs, that is, graphs in which the vertices can be partitioned into a clique and an independent set. In this paper, we shed a light to this direction. In particular, we characterize in terms of forbidden subgraphs word-representable split graphs in which vertices in the independent set are of degree at most 2, or the size of the clique is 4. Moreover, we give necessary and sufficient conditions for an orientation of a split graph to be semi-transitive.

[*]Department of Mathematics and Statistics, University of Strathclyde, 26 Richmond Street Glasgow G1 1XH, United Kingdom. **Email:** sergey.kitaev@strath.ac.uk.

[†]School of Mathematical Sciences, Shanghai Jiao Tong University, Dongchuan Road 800, 200240 Shanghai, China; and Department of Mathematics and Computer Science, University of Greifswald, Walther-Rathenau-Straße 47, D-17487 Greifswald, Germany. **Email:** yjlong@sjtu.edu.cn.

[‡]Department of Mathematics, Shanghai Jiao Tong University, Shanghai 200240, China. **Email:** majun904@sjtu.edu.cn.

[§]Shanghai Center for Mathematical Sciences, Fudan University, 220 Handan Road, Shanghai 200433, China. **Email:** hhwu@fudan.edu.cn.

1 Introduction

There is a long line of research papers dedicated to the theory of word-representable graphs (e.g. see [10]), and the core of the book [11] is devoted to the theory of word-representable graphs. The motivation to study these graphs is their relevance to algebra, graph theory, computer science, combinatorics on words, and scheduling [11]. In particular, word-representable graphs generalize several fundamental classes of graphs (e.g. *circle graphs*, 3-*colorable graphs* and *comparability graphs*).

A graph $G = (V, E)$ is *word-representable* if and only if there exists a word w over the alphabet V such that letters x and y, $x \neq y$, alternate in w if and only if $xy \in E$. The class of word-representable graphs is *hereditary*. That is, removing a vertex v in a word-representable graph G results in a word-representable graph G'. Recognizing word-representable graphs is an NP-complete problem [11].

Even though much is understood about word-representable graphs [10, 11], almost nothing is known on word-representability of *split graphs*, that is, graphs in which the vertices can be partitioned into a clique and an independent set. The only known examples in the literature of minimal non-word-representable split graphs are shown in Figure 1. These graphs are three out of the four graphs on the last line in Figure 3.9 on page 48 in [11] showing all 25 non-word-representable graphs on 7 vertices. We note that non-representability of T_1 is discussed, e.g. in [2], and non-word-representability of T_2 follows from Theorem 2 of Section 3 coming from [12]. The minimality by the number of vertices for the graphs follows from the fact that the *wheel graph* W_5 (see Section 3 for the definition) is the only non-word-representable on 6 vertices.

In this paper we characterize in terms of forbidden subgraphs word-representable split graphs in which vertices in the independent set are of degree at most 2 (see Theorem 10), or the size of the clique is 4 (see Theorem 12). To achieve these results, we introduce the following classes of graphs:

- $K_\ell^{\triangle}$, $\ell \geq 3$, in Definition 3 that are always word-representable by Theorem 8. This class of graphs is generalized in Corollary 16 to word-representable graphs K_ℓ^k.
- A_ℓ, $\ell \geq 4$, in Definition 4 that are minimal non-word-representable by Theorem 9. This class of graphs generalizes the known non-word-representable graph T_1 in Figure 1 corresponding to $\ell = 4$.

Also, in Theorem 15 we give necessary and sufficient conditions for an orientation of a split graph to be semi-transitive. In Theorem 17 we establish a particular property of semi-transitive orientations. Finally, directions for further research are in Section 8.

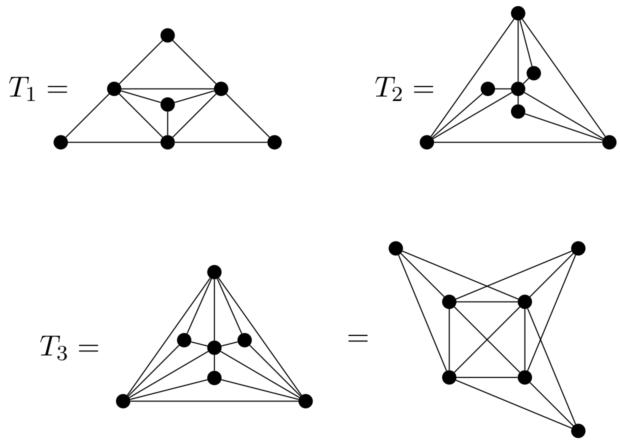
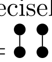


Figure 1: The minimal (by the number of vertices) non-word-representable split graphs T_1, T_2 and T_3

2 Split graphs

Let S_n be a split graph on n vertices. The vertices of S_n can be partitioned into a maximal clique K_m and an independent set E_{n-m}, i.e. the vertices in E_{n-m} are of degree at most $m-1$. We only consider such "maximal" partitions throughout the paper and let $S_n := (E_{n-m}, K_m)$.

Based on [4] it can be shown [11, Theorem 2.2.10] that the class of split graphs is the intersection of the classes of *chordal graphs* (those avoiding all cycle graphs C_m, $m \geq 4$, as induced subgraphs) and their complements, and this is precisely the class of graphs not containing the graphs C_4, C_5 and $2K_2 =$  as induced subgraphs. More relevant to our studies is the following result (see Section 3 for the definition of a comparability graph).

Theorem 1 ([6]). *Split comparability graphs are characterized by avoiding the three graphs in Figure 2 as induced subgraphs.*

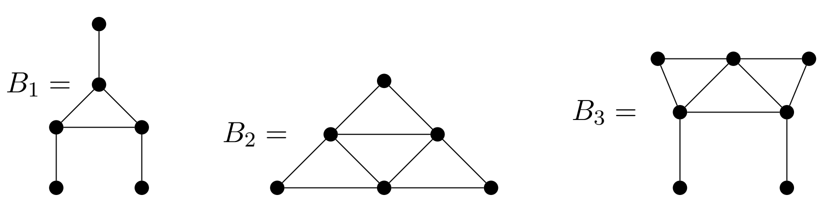


Figure 2: Forbidden induced subgraphs for split comparability graphs

It is known that any comparability graph is word-representable, and such a graph on n vertices can be represented by a word, which is a concatenation of (several) permutations of length n [13, 11]. Thus, when studying word-representability of a split graph, we can assume that one of the graphs in Figure 2 is present as an induced subgraph, because otherwise the split graph in question is a comparability graph and thus is word-representable.

3 Word-Representable graphs

Suppose that w is a word over some alphabet and x and y are two distinct letters in w. We say that x and y *alternate* in w if after deleting in w all letters but the copies of x and y we either obtain a word $xyxy\cdots$ (of even or odd length) or a word $yxyx\cdots$ (of even or odd length). For example, in 23125413241362, the letters 2 and 3 alternate. So do the letters 5 and 6, while 1 and 3 do not alternate.

Definition 1. *A graph $G = (V, E)$ is* word-representable *if there exists a word w over V such that letters x and y, $x \neq y$, alternate in w if and only if $xy \in E$. (By definition, w must contain each letter in V.) We say that w* represents *G, and that w is a* word-representant.

For example, a complete graph K_n can be represented by any permutation π of $\{1, 2, \ldots, n\}$. Also, the empty graph E_n (also known as edgeless graph, or null graph) on vertices $\{1, 2, \ldots, n\}$ can be represented by $12\cdots(n-1)nn(n-1)\cdots 21$. Definition 1 works for both labeled and unlabeled graphs because any labeling of a graph G is equivalent to any other labeling of G with respect to word-representability.

An orientation of a graph is *transitive* if presence of edges $u \to v$ and $v \to z$ implies presence of the edge $u \to z$. An unoriented graph is a *comparability graph* if it admits a transitive orientation. It is well known [11, Section 3.5.1], and is not difficult to show that the smallest non-comparability graph is the cycle graph C_5.

Theorem 2 ([12])**.** *If a graph G is word-representable then the neighbourhood of each vertex in G is a comparability graph.*

Theorem 2 allows to construct examples of non-word-representable graphs. For example, the *wheel graph* W_5, obtained from the cycle graph C_5 by adding an apex (all-adjacent vertex) is the minimal (by the number of vertices) non-word-representable graph.

4 Semi-transitive orientations

A *shortcut* is an *acyclic non-transitively oriented* graph obtained from a directed cycle graph forming a directed cycle on at least four vertices by changing the orientation of one of the edges, and possibly by adding more directed edges connecting some of the vertices (while keeping the graph be acyclic and non-transitive). Thus, any shortcut

- is acyclic (that it, there are no directed cycles);
- has at least 4 vertices;
- has exactly one source (the vertex with no edges coming in), exactly one sink (the vertex with no edges coming out), and a directed path from the source to the sink that goes through every vertex in the graph;
- has an edge connecting the source to the sink that we refer to as the *shortcutting edge*;
- is not transitive (that it, there exist vertices u, v and z such that $u \to v$ and $v \to z$ are edges, but there is no edge $u \to z$).

Definition 2. *An orientation of a graph is* semi-transitive *if it is acyclic and shortcut-free.*

Lemma 3. *Let K_m be a clique in a graph G. Then any acyclic orientation of G induces a transitive orientation on K_m. In particular, any semi-transitive orientation of G induces a transitive orientation on K_m with a single source and a single sink.*

Proof. Oriented K_m is called a tournament, and it is well known, and is not difficult to prove that any tournament contains a Hamiltonian path, that is, a path going through each vertex exactly once. Taking into account that the orientation of K_m is acyclic, it must be transitive with the unique source and sink given by the Hamiltonian path. □

A key result in the theory of word-representable graphs is the following theorem.

Theorem 4 ([9])**.** *A graph is word-representable if and only if it admits a semi-transitive orientation.*

A corollary of Theorem 4 is the following useful theorem.

Theorem 5 ([9])**.** *Any* 3*-colorable graph is word-representable.*

5 Preliminaries

We begin with a result that allows us to assume in our studies that the size of a maximal clique in a split graph is at least 4.

Theorem 6. $S_n = (E_{n-m}, K_m)$ *is word-representable for* $m \leq 3$.

Proof. S_n is 3-colorable, and so, by Theorem 5, it is word-representable. □

The following lemma allows us to assume in our studies that (i) each vertex in a split graph is of degree at least 2, and (ii) no two vertices have the same set of neighbours.

Lemma 7. *Let $S_n = (E_{n-m}, K_m)$ and S_{n+1} be split graphs, where S_{n+1} is obtained from S_n by either adding a vertex of degree 0 (to E_{n-m}), or adding a vertex of degree 1 (to E_{n-m}), or by "copying" a vertex (either in E_{n-m} or in K_m), that is, by adding a vertex whose neighbourhood is identical to the neighbourhood of a vertex in S_n. Then S_n is word-representable if and only if S_{n+1} is word-representable.*

Proof. Suppose a vertex x of degree 0 is added to a word-representable S_n with a word-representant w. Then the word xxw represents S_{n+1}.

Connecting two word-representable graphs by an edge gives a word-representable graph (see [11, Section 5.4.3]), which is easy to see using semi-transitive orientations and Theorem 4. A 1-vertex graph is word-representable, so the lemma is true for adding a vertex of degree 1.

Copying a vertex v in S_n (either connected, or not, to v) is a particular case of replacing any vertex in a word-representable graph by a *module*, which is a comparability graph. It is known (see [11, Section 5.4.4]) that such a replacement gives a word-representable graph, which completes the proof of the lemma. □

Definition 3. *For $\ell \geq 3$, the graph $K_\ell^\triangle$ is obtained from the complete graph K_ℓ labeled by $1, 2, \ldots, \ell$, by adding a vertex i' of degree 2 connected to vertices i and $i+1$ for each $i \in \{1, 2, \ldots, \ell-1\}$. Also, a vertex ℓ' connected to the vertices 1 and ℓ is added. See Figure 3 for the graph $K_6^\triangle$.*

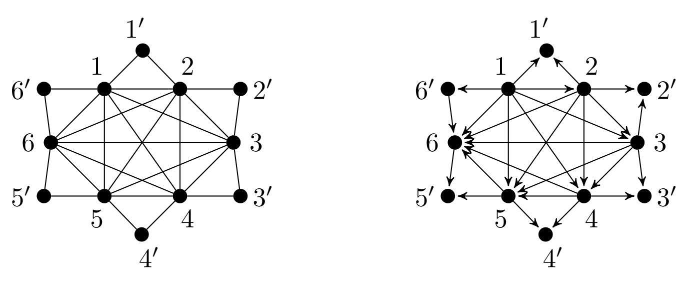


Figure 3: The graph $K_6^\triangle$ and one of its semi-transitive orientations

Theorem 8. *$K_\ell^\triangle$ is word-representable.*

Proof. In the case of odd ℓ, it is not difficult to come up with a word representing $K_\ell^\triangle$ based on the representation $12\cdots\ell12\cdots\ell$ of K_ℓ and adding i's as follows (where we present the resulting word on two lines):

$$1'121'3'343'5'565'\cdots(\ell-2)'(\ell-2)(\ell-1)(\ell-2)'\ell'\ell 1\ell'2'232'$$
$$4'454'\cdots(\ell-1)'(\ell-1)\ell(\ell-1)'.$$

However, we next provide a semi-transitive orientation of $K_\ell^{\triangle}$ that works for any ℓ, so that the statement will follow from Theorem 4.

First, orient the K_ℓ transitively so that there is a directed path $1 \to 2 \to \cdots \to \ell$ as shown for the case $\ell = 6$ in Figure 3. Next, for $i \in \{1, 2, \ldots, \ell-1\}$ orient the edges incident to i' as $i \to i'$ and $(i+1) \to i'$. Finally, orient the edges incident to ℓ' as $1 \to \ell'$ and $\ell' \to \ell$ as again shown for the case $\ell = 6$ in Figure 3.

We claim that the orientation obtained is semi-transitive. Indeed, it is easy to see that there are no directed cycles. Furthermore, because K_ℓ is transitively oriented, any possible shortcut must involve a vertex i'. Clearly, $\ell' \to \ell$ and $1 \to \ell'$ are not shortcutting edges because ℓ' is neither a sink nor a source. Note that $a < b$ whenever $a \to b$ for $a, b \in \{1, 2, \ldots, \ell\}$. Using this observation, $(i+1) \to i'$, for $i \in \{1, 2, \ldots, \ell-1\}$, is not a shortcutting edge because there is no path from a vertex $(i+1)$ to a vertex i. Finally, $i \to i'$, for $i \in \{1, 2, \ldots, \ell-1\}$, cannot be a shortcutting edge because there is no path of length greater than 2 from a vertex i to a vertex i'. □

Definition 4. *For $\ell \geq 4$, let A_ℓ be the graph obtained from $K_{\ell-1}^{\triangle}$ by adding a vertex ℓ connected to the vertices $1, 2, \ldots, \ell-1$ and no other vertices. Note that $A_4 = T_1$ in Figure 1. A schematic way to represent a graph A_ℓ is shown in Figure 4.*

Theorem 9. *A_ℓ is a minimal non-word-representable graph.*

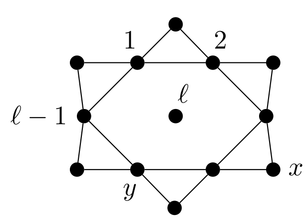


Figure 4: A schematic way to represent A_ℓ

Proof. **Minimality.** Because of the symmetries, we only need to consider three cases with a reference to Figure 4.

- Removing the vertex ℓ we obtain the graph $K_{\ell-1}^{\triangle}$ which is word-representable by Theorem 8.

- Removing the vertex x we get a graph isomorphic to the graph obtained from $K_\ell^\triangle$ by removing the vertices $1'$ and $2'$. Such a graph is word-representable by Theorem 8 taking into account the hereditary nature of word-representability.
- Removing the vertex y, and then the two obtained vertices of degree 1 not affecting word-representability by Lemma 7, we get a graph isomorphic to the graph obtained from $K_{\ell-1}^\triangle$ by removing the vertices $1'$ and $2'$, which is word-representable by Theorem 8 taking into account the hereditary nature of word-representability.

Non-word-representability. We will show that A_ℓ does not admit a semi-transitive orientation, and the result will follow by Theorem 4.

Suppose A_ℓ admits a semi-transitive orientation. By Lemma 3, this orientation induces a transitive orientation on the clique of size $\ell-1$ obtained by removing the vertex ℓ. We claim that without loss of generality, we can assume that the Hamiltonian path on this clique is $1 \to 2 \to \cdots \to (\ell-1)$, or its cyclic shift (e.g. $2 \to 3 \to \cdots \to (\ell-1) \to 1$, or $3 \to 4 \to \cdots \to (\ell-1) \to 1 \to 2$, etc). Indeed, if that were not the case, then changing all orientations to the opposite, if necessary, there must exist an i such that

- $P_i = i \to x_1 \to x_2 \to \cdots \to x_j \to (i+1)$ is part of the Hamiltonian path for $j \geq 1$; if $i = (\ell-1)$ then $(i+1) := 1$;
- either $x \to i$, or $(i+1) \to y$, or both, are present in the Hamiltonian path for some vertices x and y.

If the orientation of the edge $i'(i+1)$ is $(i+1) \to i'$ then this edge, along with P_i and the edge ii' will either induce a directed cycle, or a shortcut, a contradiction. Thus, the orientation of $i'(i+1)$ must be $i' \to (i+1)$. Furthermore, to avoid a shortcut involving the edge $i' \to (i+1)$ and P_i, we must orient the edge ii' as $i \to i'$. But now, the graph induced by P_i, $i \to i'$, $i' \to (i+1)$, and $x \to i$ or $(i+1) \to y$ (whatever exists) will induce a shortcut. Indeed, in the former case, the edge $x \to (i+1)$ is present, but the edge $x \to i'$ is not, while in the latter case, the edge $i \to y$ is present, while $i' \to y$ is not. Thus, renaming the vertices, if necessary (which is equivalent to a cyclic shift), we can assume that the partial orientation of the semi-transitively oriented A_ℓ is as in the graph to the left in Figure 5. In that figure, we do not draw the edges $i \to j$ for $|j-i| \geq 2$, except for the edge $1 \to (\ell-1)$, to arrange a better look for the figure (although existence of these edges is assumed).

Now, if $(\ell-1) \to \ell'$ were an edge, then the edge $1\ell'$ would either be a shortcutting edge (e.g. $2 \to \ell'$ is missing), or would form a cycle taking into account the directed path $1 \to 2 \to \cdots \to (\ell-1)$. Thus,

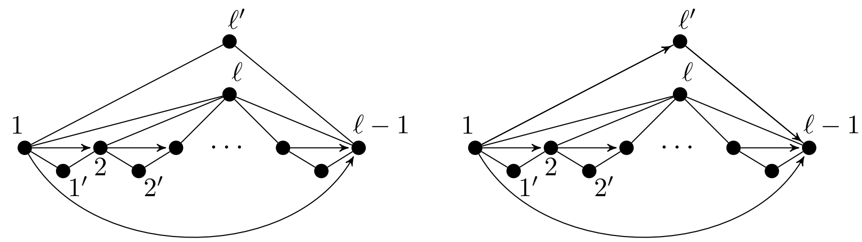


Figure 5: Non-word-representability of A_ℓ

we must have $\ell' \to (\ell-1)$, and not to have a shortcut, we must also have $1 \to \ell'$, as shown in the graph to the right in Figure 5.

Next, consider the triangle $121'$. Orienting it as $2 \to 1'$ and $1' \to 1$ gives a cycle, while orienting it as $1 \to 1'$ and $1' \to 2$ gives a shortcut induced by the vertices 1, $1'$, 2 and 3 with the shortcutting edge $1 \to 3$. On the other hand, similarly to the proof of Theorem 8, one can see that none of the orientations $1 \to 1'$ and $2 \to 1'$, or $1' \to 1$ and $1' \to 2$, results in a shortcut or a cycle. Similarly, no matter which of these orientations is selected, when considering the graph induced by the vertices 1, 2, $1'$ and ℓ, we see that the orientation of the edges 1ℓ and 2ℓ must either be $1 \to \ell$ and $2 \to \ell$, or $\ell \to 1$ and $\ell \to 2$.

Similar arguments as above can be applied to the graphs induced by i, i', $(i+1)$ and ℓ for $i=2$, then $i=3$, etc, up to $i=\ell-2$, except for now the orientations of the edges $i\ell$ and $(i+1)\ell$ will be uniquely defined based on the orientation of the edge 1ℓ. Thus, we see that ℓ must either be a sink or a source. Considering the graph induced by the vertices 1, $(\ell-1)$, ℓ and ℓ' we see that in the former case, $1 \to \ell$ is a shortcutting edge, while in the later case $\ell \to (\ell-1)$ is a shortcutting edge, a contradiction. Thus, A_ℓ does not admit a semi-transitive orientation and thus is not word-representable. □

6 Our characterization results

6.1 Restricting degrees in E_{n-m} to be at most 2

Definition 5. *For a split graph (E_{n-m}, K_m), any triangle induced by two vertices in K_m and one vertex in E_{n-m} is a* non-clique triangle.

Theorem 10. *Let $m \geq 1$ and $S_n = (E_{n-m}, K_m)$ be a split graph. Also, let the degree of any vertex in E_{n-m} be at most 2. Then S_n is word-representable if and only if S_n does not contain the graphs T_2 in Figure 1 and A_ℓ in Definition 4 as induced subgraphs.*

Proof. Note that the necessary condition is given by Theorem 9 and

Figure 1 containing non-word-representable graphs.

For the other direction, suppose that S_n is A_ℓ-free and T_2-free. By Lemma 7, we can assume that each vertex in E_{n-m} is of degree 2, and no two vertices in E_{n-m} have the same neighbourhood. Because S_n is T_2-free, no three non-clique triangles can be incident to the same vertex. Moreover, because S_n is A_ℓ-free, K_m cannot have a cycle of size less than m such that each edge in the cycle is an edge in a non-clique triangle. These observations imply that either K_m has a cycle of length m formed by edges in non-clique triangles, or it contains disjoint paths, such that each edge in a path is an edge in a non-clique triangle, as shown schematically in Figure 6. But then we can redraw the graph, if necessary, to see that S_n is exactly the graph $K_m^{\triangle}$ with possibly some non-clique triangles missing, and this graph is word-representable by Theorem 8 taking into account the hereditary nature of word-representability. □

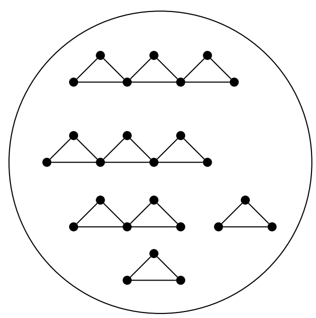


Figure 6: Schematic structure of the graph S_n in Theorem 10

6.2 Cliques of size 4

We restrict our attention to the case of cliques of size 4 ($m = 4$). If the degrees of vertices in E_{n-4} are at most 2, we can apply Theorem 10 to see that word-representability is characterized by avoidance of the graphs T_1 and T_2 in Figure 1 as induced subgraphs. However, E_{n-4} may also have vertices of degree 3. Theorem 12 below gives a complete characterization for word-representability of (E_{n-4}, K_4).

Our methodology to prove Theorem 12 is in using Lemma 7 to come up with the largest possible split graph S_n in the context. We then identify a minimal non-word-representable induced subgraph in such S_n and consider a smaller graph S_{n-1} obtained from S_n by removing one vertex. We need to consider all possibilities of removing a vertex in S_n, but we use symmetries, whenever possible, to reduce the number of cases to consider. If S_{n-1} is word-representable, there is nothing to

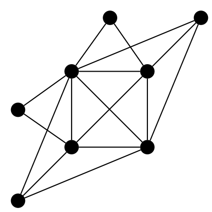
$T_4 =$ 

Figure 7: A minimal non-word-representable split graph T_4

do. Otherwise, we repeat the process for S_{n-1} instead of S_n. This way, we located all minimal non-word-representable induced subgraphs. We note that in the proof, orientations claimed by us to be semi-transitive, can be checked to be such either by hand, or using the software [5].

Lemma 11. *The split graph T_4 in Figure 7 is a minimal non-word-representable graph.*

Proof. The neighbourhood of the universal vertex in T_4 is isomorphic to the non-comparability graph B_3 in Figure 2, and thus T_4 is not word-representable by Theorem 2.

The minimality of T_4 follows from the fact that removing a vertex in T_4 we do not obtain one of the graphs in Figure 3.9 on page 48 in [11] showing all 25 non-word-representable graphs on 7 vertices. □

Theorem 12. *Let $S_n = (E_{n-4}, K_4)$ be a split graph. Then S_n is word-representable if and only if S_n does not contain the graphs T_1, T_2 and T_3 in Figure 1, and T_4 in Figure 7 as induced subgraphs.*

Proof. We can assume that E_{n-4} contains at least one vertex of degree 3, or else we are done by Theorem 10 with $T_1 = A_3$ and T_2 being forbidden induced subgraphs. Further, recall that by Lemma 7, we can assume that each vertex in E_{n-4} is of degree 2 or 3, and no vertices in E_{n-4} have the same neighbourhood.

Note that the necessary condition has already been proved, so for the opposite direction we assume that S_n is T_i-free for $i = 1, 2, 3, 4$.

Assuming that E_{n-4} only contains vertices of degree 3, we can see that T_1 and T_3 are the only minimal non-word-representable induced subgraphs to be avoided by S_n to be word-representable. Indeed, since no vertices in E_{n-4} have the same neighbourhood, E_{n-4} can have at most 4 vertices in this case. If all 4 vertices are present, S_n contains the minimal non-word-representable T_3 as an induced subgraph (to see this, T_3 is redrawn in a different way in Figure 1). Removing one of the 4 vertices in E_{n-4} (any one due to the symmetries) we obtain exactly T_3 which is a minimal non-word-representable graph.

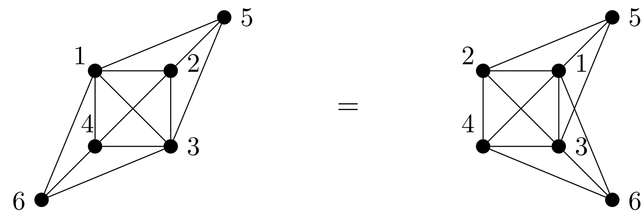


Figure 8: The single maximal possibility, up to isomorphism, for (E_{n-4}, K_4) with vertices of degree 3 in E_{n-4}

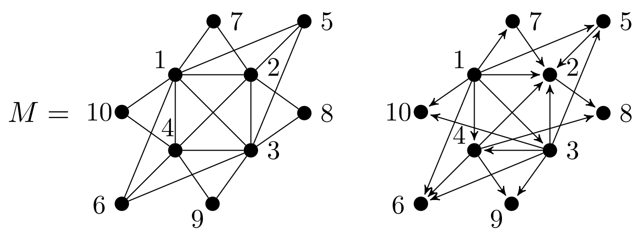


Figure 9: Maximal non-isomorphic possibilities for (E_{n-4}, K_4) with vertices of degree 2 and 3 in E_{n-4}

It remains to notice that if E_{n-4} contains 4 vertices and we remove a vertex in K_4 we will obtain the minimal non-word-representable graph T_1. Thus, E_{n-4} can have at most two vertices in this case, resulting, up to isomorphism, in a single case to consider that is presented in Figure 8 (along with a justification that two of seemly different graphs are actually isomorphic). But S_n is T_1-free and T_3-free.

We next consider adding vertices of degree 2 to E_{n-4} in the graph in Figure 8. Since S_n is T_2-free, no three non-clique triangles (with disjoint vertices) can be incident to the same vertex. Thus, at most four vertices (with distinct neighbourhoods) of degree 2 can be present in E_{n-4}, and there are just two non-isomorphic ways to add these vertices to the graph in Figure 8 that are given in Figure 9.

The graph to the right in Figure 9 is word-representable and we provide one of its semi-transitive orientations to justify this (we omit a justification that the orientation is semi-transitive). On the other hand, the graph M in Figure 9 contains T_4 (just remove the vertices 7 and 10 to see this) while S_n is assumed to be T_4-free. However, we are not done yet since M is not a minimal non-word-representable graph. To find all minimal non-word-representable induced subgraphs in M, we will consider removing one vertex from it. Note that there are only

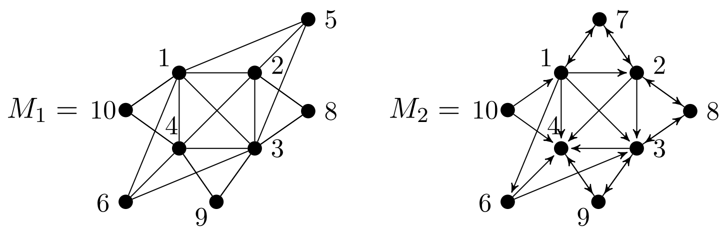


Figure 10: Two cases to consider in the proof of Theorem 12

four cases to consider up to isomorphism.

1. If vertex 1 is removed in M, then vertices 7 and 10 will be of degree 1 and can also be removed by Lemma 7. Moreover, the vertices 6 and 9 will have the same neighbourhoods, and by the same lemma, one of these vertices can be removed. The same applies to vertices 5 and 8, resulting in a graph on 5 vertices induced by, say, vertices 2, 3, 4, 5, 6, and any graph on 5 vertices is word-representable.
2. If vertex 2 is removed in M, then vertices 7 and 8 will be of degree 1 and thus can also be removed by Lemma 7. This leaves us with a graph on 6 vertices which is word-representable because it is different from W_5, the only non-word-representable graph on 6 vertices.
3. If vertex 5 is removed, then we obtain the graph M_2 in Figure 10, which is word-representable because of the semi-transitive orientation we provide in the figure (we omit a justification that the orientation is semi-transitive).
4. Finally, if vertex 7 is removed in M, we will obtain the non-word-representable graph M_1 in Figure 10 (it contains T_4).

To complete our proof, we need to remove a vertex in M_1. No symmetries can be applied here, so we have to consider 9 cases.

1. If vertex 1 is removed then vertex 10 will be of degree 1 and it can be removed by Lemma 7. The resulting graph is word-representable because it is clearly a subgraph of T_4, and T_4 is a minimal non-word-representable.
2. If vertex 2 is removed then vertex 8 will be of degree 1 and it can be removed by Lemma 7. The resulting graph is precisely the non-word-representable graph T_1, which is avoided by S_n.
3. If vertex 3 is removed then vertices 8 and 9 will be of degree 1 and they can be removed by Lemma 7. The resulting graph is on 6 vertices, it is not W_5 and thus is word-representable.

4. If vertex 4 is removed then vertices 9 and 10 will be of degree 1 and they can be removed by Lemma 7. The resulting graph is on 6 vertices, it is not W_5 and thus is word-representable.
5. If vertex 5 is removed then we will obtain a word-representable graph M_3 in Figure 11, where we provide a semi-transitive orientation of the graph without justification.
6. If vertex 6 is removed then we will obtain a word-representable graph M_4 in Figure 11, where we provide a semi-transitive orientation of the graph without justification.
7. If vertex 8 is removed then we will obtain the graph M_5 in Figure 11. This graph contains T_1 as an induced subgraph (remove vertex 2 to see it). To complete this case, we need to remove a vertex in M_5 other than vertex 2, to make sure that a word-representable graph would be obtained.
   (a) Removing vertex 1, which is clearly equivalent to removing vertex 3, gives vertex 8 of degree 1 which can be removed by Lemma 7. Moreover, one of vertices 6 and 9 can be removed by Lemma 7 because they have the same neighbourhood. This results in a graph on 5 vertices, but any such graph is word-representable.
   (b) Removing vertex 4 gives two vertices, 9 and 10, that can be removed by Lemma 7. The resulting graph is on 5 vertices and it must be word-representable.
   (c) Removing vertex 5 is equivalent to removing vertices 7 and 8 in the graph M_2 in Figure 10, so this graph is word-representable.
   (d) Removing vertex 6 is equivalent to removing vertices 9 and 10 in the graph M_2 in Figure 10
   (e) Finally, removing vertex 9, which is clearly equivalent to removing vertex 10, gives the graph obtained from the semi-transitively oriented graph M_6 in Figure 11, and it is word-representable.
8. If vertex 9 is removed then the semi-transitively oriented graph M_6 in Figure 11 is obtained (we omit justification that the orientation is indeed semi-transitive).
9. Finally, if vertex 10 is removed then we will obtain the minimal non-word-representable graph T_4, which is avoided by S_n.

We have shown that there are no other minimal non-word-representable graphs apart from T_i, $i = 1, 2, 3, 4$, that need to be avoided in order for S_n to be word-representable. □

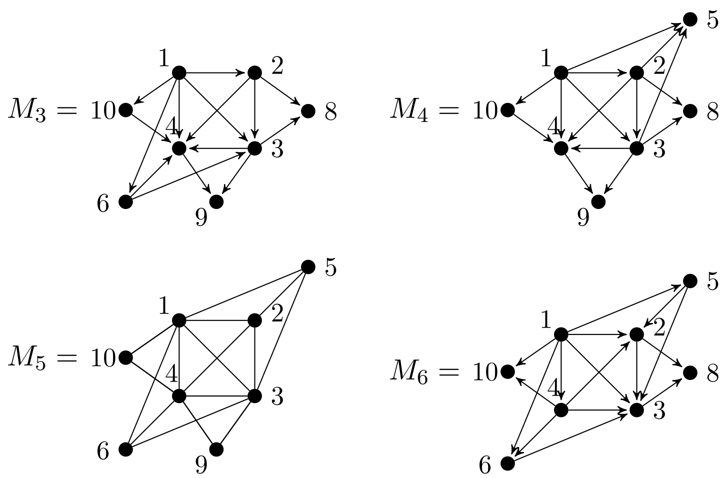


Figure 11: Four subcases to consider in the proof of Theorem 12

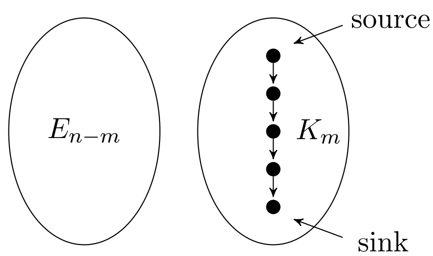


Figure 12: A schematic structure of a semi-transitively oriented split graph

7 Semi-transitive orientations on split graphs

Let $S_n = (E_{n-m}, K_m)$ be a word-representable split graph. Then by Theorem 4, S_n admits a semi-transitive orientation. Further, by Lemma 3 we known that any such orientation induces a transitive orientation on K_m that can be presented schematically as in Figure 12, where we show the longest directed path in K_m, denoted by $\vec{P}$, but do not draw the other edges in K_m even though they exist.

Lemmas 13 and 14 below describe the structure of semi-transitive orientations in an arbitrary word-representable split graph.

Lemma 13. *Any semi-transitive orientation of $S_n = (E_{n-m}, K_m)$ subdivides the set of all vertices in E_{n-m} into three, possibly empty, groups corresponding to each of the following types presented schemat-*

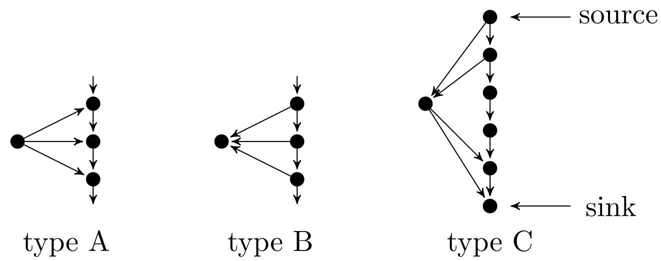


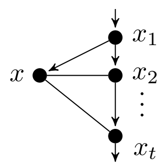
Figure 13: Three types of vertices in E_{n-m} under a semi-transitive orientation of (E_{n-m}, K_m). The vertical oriented paths are a schematic way to show (parts of) $\vec{P}$

ically in Figure 13, where $\vec{P} = p_1 \to \cdots \to p_m$ *is the longest directed path in* K_m*:*

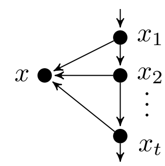
- *A vertex in* E_{n-m} *is of* type A *if it is a source and is connected to all vertices in* $\{p_i, p_{i+1}, \ldots, p_j\}$ *for some* $1 \leq i \leq j \leq m$*;*
- *A vertex in* E_{n-m} *is of* type B *if it is a sink and is connected to all vertices in* $\{p_i, p_{i+1}, \ldots, p_j\}$ *for some* $1 \leq i \leq j \leq m$*;*
- *A vertex* $v \in E_{n-m}$ *is of* type C *if there is an edge* $x \to v$ *for each* $x \in I_v = \{p_1, p_2, \ldots, p_i\}$ *and there is an edge* $v \to y$ *for each* $y \in O_v = \{p_j, p_{j+1}, \ldots, p_m\}$ *for some* $1 \leq i < j \leq m$*.* I_v *(resp.,* O_v*) is called the* source-group *(resp.,* sink-group*) or vertices.*

Proof. Let x be a vertex in E_{n-m} and $x_1, x_2, \ldots, x_t$ be the vertices in K_m that are connected to x. First observe that to avoid directed cycles, the partial orientation

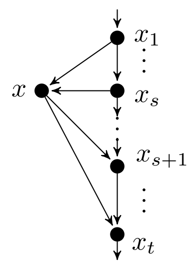
Moreover, the vertices $x_1, x_2, \ldots, x_t$ must be consecutive on $\vec{P}$. Indeed, if x_i and x_{i+1} are not consecutive for some i, $1 \leq i \leq t-1$ (there is a vertex on $\vec{P}$ between x_i and x_{i+1} not connected to x) then the vertices on $\vec{P}$ between x_1 and x_{i+1}, along with x, form a shortcut with the shortcutting edge $x \to x_{i+1}$. On the other hand, the partial orientation



can either be extended to



with x_1, $x_2, \ldots, x_t$ being consecutive to avoid $x_1 \to x$ being a shortcutting edge (by the reasons similar to the previous case), or to avoid directed cycles, all edges of the form $x_i \to x$ must be above of all edges of the form $x \to x_i$ as shown schematically in the following picture where $s := \max\{i \mid x_i \to x \text{ is an edge}\}$:

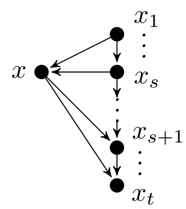


One can use arguments as above to see that to avoid shortcuts, the vertices x_1, $x_2, \ldots, x_s$ corresponding to the edges oriented towards x must be consecutive on $\vec{P}$. The vertices x_{s+1}, $x_{s+2}, \ldots, x_t$ must also be consecutive on $\vec{P}$. On the other hand, there are no restrictions on the vertices x_s and x_{s+1}, so there can be some other vertices there on the path $\vec{P}$.

To complete the theorem, we show that x_1 (resp., x_t) must be the source (resp., sink) in $\vec{P}$. Indeed, suppose there exists a vertex y on $\vec{P}$ such that $y \to x_1$ is an edge. Then the subgraph induced by the vertices y, x_1, x, x_{s+1} is a shortcut with the shortcutting edge $y \to x_{s+1}$ (because the edge $y \to x$ is missing), a contradiction. Similarly, if there exists a vertex z on $\vec{P}$ such that $x_t \to z$ is an edge, then the graph induced by the vertices x_1, x, x_t, z is a shortcut with the shortcutting edge $x_1 \to z$ (because the edge $x \to z$ is missing), a contradiction. □

There are additional restrictions on relative positions of the neighbours of vertices of the types A, B and C. These restrictions are given by the following lemma.

Lemma 14. *Let $S_n = (E_{n-m}, K_m)$ be oriented semi-transitively. For a vertex $x \in E_{n-m}$ of the type C, presented schematically as*



there is no vertex $y \in E_{n-m}$ of the type A or B, which is connected to both x_s and x_{s+1}. Also, there is no vertex $y \in E_{n-m}$ of the type C such that either the source-group, or the sink-group of vertices given by y (see the statement of Lemma 13 and its proof for the definitions) contains both x_s and x_{s+1}.

Proof. If y is of the type A, then the subgraph induced by the vertices y, x_s, x and x_{s+1} is a shortcut with the shortcutting edge being $y \to x_{s+1}$ (the edge $y \to x$ is missing).

Similarly, if y is of the type B, then the subgraph induced by the vertices y, x_s, x and x_{s+1} is a shortcut with the shortcutting edge being $x_s \to y$ (the edge $x \to y$ is missing).

If y is of the type C and both x_s and x_{s+1} belong to the same group of y's neighbours, then $x_1 \to x_t$ will be a shortcutting edge. Indeed, if both x_s and x_{s+1} belong to

- the source-group then $x_1 \to x_s \to x \to x_{s+1} \to y \to x_t$ induces a non-transitive subgraph (the edge $y \to x$ is missing).
- the sink-group then $x_1 \to y \to x_s \to x \to x_t$ induces a non-transitive subgraph (the edge $y \to x$ is missing). □

The following theorem is a classification theorem for semi-transitive orientations on split graphs.

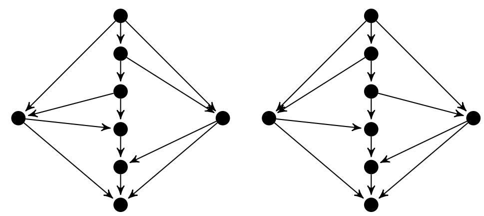
Theorem 15. *An orientation of a split graph $S_n = (E_{n-m}, K_m)$ is semi-transitive if and only if*

- *K_m is oriented transitively,*
- *each vertex in E_{n-m} is of one of the three types presented in Figure 13, and*
- *the restrictions in Lemma 14 are satisfied.*

Proof. The forward direction follows from Lemmas 3, 13 and 14.

For the opposite direction, suppose that all restrictions are satisfied, but a shortcut is created with the longest path $\vec{X}$ from the source to the sink. Note that $\vec{X}$ must involve a node in E_{n-m} because K_m is oriented transitively. Also, $\vec{X}$ cannot involve more that one vertex of the type A or B because otherwise we obtain a contradiction with the beginning of $\vec{X}$ not being the beginning of the shortcutting edge (vertices of the type A or B are sinks or sources).

Next, we note that $\vec{X}$ cannot pass through two vertices of the type C if they satisfy the conditions of Lemma 14, which is easy to see from the following two figures representing schematically all possibilities:



Finally, we need to consider the situations when $\vec{X}$ passes through

- a vertex y of type A and a vertex x of type C, and
- a vertex y of type B and a vertex x of type C

while respecting the conditions of Lemma 14. In either of these cases, both the shortcutting edge and the beginning of $\vec{X}$ must clearly start, or end, at y (depending on y's type). But then, in order for $\vec{X}$ to visit x, the vertex y must be connected to both x_s and x_{s+1} in the terminology of Lemma 14, a contradiction. □

The following corollary of Theorem 15 generalizes Theorem 8 (which is the case $k = 2$ in the corollary).

Corollary 16. *Let the split graph K_ℓ^k be obtained from the complete graph K_ℓ with vertices drawn on a circle, by adding ℓ vertices so that*

- *each such vertex is connected to k consecutive (on the circle) vertices in K_ℓ;*
- *neighbourhoods of all these vertices are distinct; and*
- *$\ell \geq 2k - 1$.*

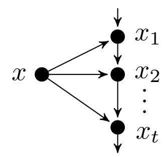
Then K_ℓ^k is word-representable.

Proof. Orient the clique in K_ℓ^k transitively with the Hamiltonian path going around the circle, and then assign to the vertices in the independent set types A (or B) and C. Because $\ell \geq 2k - 1$, no vertex of the type A or B will be violating the condition of Lemma 14, and thus by Theorem 15, the obtained orientation is semi-transitive. □

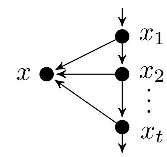
We complete this section with the following theorem.

Theorem 17. *Let $S_n = (E_{n-m}, K_m)$ be semi-transitively oriented. Then any vertex in E_{n-m} of the type A can be replaced by a vertex of the type B, and vice versa, keeping orientation be semi-transitive.*

Proof. Suppose that a vertex x of the type A



becomes a vertex of the type B



while no other orientation is changed in the semi-transitively oriented S_n. Clearly, if the change has resulted in a non-semi-transitive orientation then the vertex x must be involved in a shortcut (it cannot be involved in a directed cycle) with $x_i \to x$ being a shortcutting edge for some i. This contradicts to the vertices $x_1, \ldots, x_t$ being consecutive on $\vec{P}$ and inducing a transitive orientation together with x.

Essentially identical arguments, with a shortcutting edge being $x \to x_i$ this time, show that switching from type B to type A for a vertex $x \in E_{n-m}$ does not result in a non-semi-transitive orientation. □

8 Concluding remarks

In this paper, we characterized in terms of forbidden subgraphs word-representable split graphs $S_n = (E_{n-m}, K_m)$ in which vertices in E_{n-m} are of degree at most 2 (see Theorem 10), or the size of K_m is 4 (see Theorem 12). Moreover, in Theorem 15 we give necessary and sufficient conditions for an orientation of a split graph to be semi-transitive. Our results were the basis for (computational) characterization of word-representable graphs with cliques of size 5 in the follow up paper [1].

There are several natural directions of further research. For example, one can consider vertices of degree at most 3 in E_{n-m} (thus extending the results in Theorem 10), or letting the clique be of size 6 (thus extending the results in Theorem 12 and in [1]). Either of these directions is challenging due to a large number of cases to consider. It is conceivable that our classification result, Theorem 15, on semi-transitive orientations of split graphs will eventually be the key for a complete classification of word-representable split graphs, but for the moment it is difficult to state any conjectures on how such a characterization would look like.

9 Acknowledgments

The second author acknowledges support of National Natural Science Foundation of China (No. 11671258) and Postdoctoral Science Foundation of China (No. 2016M601576).

References

[1] T. Z. Q. Chen, S. Kitaev, A. Saito. Representing split graphs by words, *Discussiones Mathematicae Graph Theory*, to appear.

[2] T. Z. Q. Chen, S. Kitaev, B. Y. Sun. Word-representability of face subdivisions of triangular grid graphs. *Graphs and Combinatorics* **32(5)** (2016) 1749–1761.

[3] A. Collins, S. Kitaev, V. Lozin. New results on word-representable graphs. *Discr. Appl. Math.* **216** (2017) 136–141.

[4] S. Foldes, P. L. Hammer. Split graphs. In Proceedings of the *Eighth Southeastern Conference on Combinatorics, Graph Theory and Computing* (Louisiana State Univ., Baton Rouge, 1977), pages 311–315. Congressus Numerantium, No. XIX.

[5] M. Glen. Software available at `personal.cis.strath.ac.uk/sergey.kitaev/word-representable-graphs.html`

[6] M. Golumbic. *Algorithmic Graph Theory and Perfect Graphs*, Academic Press (1980) ISBN 0-12-289260-7.

[7] M. Halldórsson, S. Kitaev, A. Pyatkin. Graphs capturing alternations in words. *Lecture Notes Comp. Sci.* **6224** (2010) 436–437.

[8] M. Halldórsson, S. Kitaev, A. Pyatkin. Alternation graphs. *Lecture Notes Comp. Sci.* **6986** (2011) 191–202.

[9] M. Halldórsson, S. Kitaev, A. Pyatkin. Semi-transitive orientations and word-representable graphs. *Discr. Appl. Math.* **201** (2016) 164–171.

[10] S. Kitaev. A Comprehensive Introduction to the Theory of Word-Representable Graphs, *Lecture Notes Comp. Sci.* **10396** (2017) 36–67.

[11] S. Kitaev, V. Lozin. Words and Graphs, *Springer*, 2015.

[12] S. Kitaev, A. Pyatkin. On representable graphs. *J. Autom., Lang. and Combin.* **13** (2008) 1, 45–54.

[13] S. Kitaev, S. Seif. Word problem of the Perkins semigroup via directed acyclic graphs. *Order* **25** (2008) 3, 177–194.