

Vertex Deletion Problems on Chordal Graphs^{*†}

Yixin Cao¹, Yuping Ke², Yota Otachi³, and Jie You⁴

- 1 Department of Computing, Hong Kong Polytechnic University, Hong Kong, China
yixin.cao@polyu.edu.hk
- 2 Department of Computing, Hong Kong Polytechnic University, Hong Kong, China
yu.ke@polyu.edu.hk
- 3 Faculty of Advanced Science and Technology, Kumamoto University, Kumamoto, Japan
otachi@cs.kumamoto-u.ac.jp
- 4 School of Information Science and Engineering, Central South University and Department of Computing, Hong Kong Polytechnic University, Hong Kong, China
jie.you@polyu.edu.hk

Abstract

Containing many classic optimization problems, the family of vertex deletion problems has an important position in algorithm and complexity study. The celebrated result of Lewis and Yannakakis gives a complete dichotomy of their complexity. It however has nothing to say about the case when the input graph is also special. This paper initiates a systematic study of vertex deletion problems from one subclass of chordal graphs to another. We give polynomial-time algorithms or proofs of NP-completeness for most of the problems. In particular, we show that the vertex deletion problem from chordal graphs to interval graphs is NP-complete.

1998 ACM Subject Classification F.2.2 Analysis of Algorithms and Problem Complexity, G.2.2 Graph Theory

Keywords and phrases vertex deletion problem, maximum subgraph, chordal graph, (unit) interval graph, split graph, hereditary property, NP-complete, polynomial-time algorithm

Digital Object Identifier 10.4230/LIPIcs.FSTTCS.2017.22

1 Introduction

Generally speaking, a vertex deletion problem asks to transform an input graph to a graph in a certain class by deleting a minimum number of vertices. Many classic optimization problems belong to the family of vertex deletion problems, and their algorithms and complexity have been intensively studied. For example, the clique problem and the independent set problem are nothing but the vertex deletion problems to complete graphs and to edgeless graphs respectively. Most interesting graph properties are *hereditary*: If a graph satisfies this property, then so does every induced subgraph of it. For all the vertex deletion problems to hereditary graph classes, Lewis and Yannakakis [26] have settled their complexity once and for all with a dichotomy result: They are either NP-hard or trivial. Thereafter algorithmic efforts

* A full version of the paper is available at <https://arxiv.org/abs/1707.08690>.

† Supported in part by the Hong Kong Research Grants Council (RGC) under grants 252026/15E and 152261/16E, and the National Natural Science Foundation of China (NSFC) under grant 61572414.



© Yixin Cao, Yuping Ke, Yota Otachi, and Jie You;
licensed under Creative Commons License CC-BY

37th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2017).

Editors: Satya Lokam and R. Ramanujam; Article No. 22; pp. 22:1–22:14



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

were mostly focused on the nontrivial ones, and the major approaches include approximation algorithms [27], parameterized algorithms [6], and exact algorithms [15].

Chordal graphs make one of the most important graph classes. Together with many of its subclasses, it has played important roles in the development of structural graph theory. (We defer their definitions to the next section.) Many algorithms have been developed for vertex deletion problems to chordal graphs and its subclasses,—most notably (unit) interval graphs, cluster graphs, and split graphs; see, e.g., [17, 4, 10, 9, 8, 33, 12, 25, 1] for a partial list. After the long progress of algorithmic achievements, some natural questions arise: What is the complexity of transforming a chordal graph to a (unit) interval graph, a cluster graph, a split graph, or a member of some other subclass of chordal graphs? It is quite surprising that this type of problems has not been systematically studied, save few concrete results, e.g., the polynomial-time algorithms for the clique problem, the independent set problem, and the feedback vertex set problem (the object class being forests) [21, 32].

The same question can be asked for other pair of source and object graph classes. The most important source classes include planar graphs [20, 18, 16], bipartite graphs [31], and degree-bounded graphs [19]. As one may expect, with special properties imposed on input graphs, the problems become easier, and some of them may not remain NP-hard. Unfortunately, a clear-cut answer to them seems very unlikely, since their complexity would depend upon both the source class and the object class. Indeed, some are trivial (e.g., vertex cover on split graphs), some remain NP-hard (e.g., vertex cover on planar graphs), while some others are in P but can only be solved by very nontrivial polynomial-time algorithms (e.g., vertex cover on bipartite graphs).

Throughout the paper we write the names of graph classes in small capitals; e.g., CHORDAL and BIPARTITE stand for the class of chordal graphs and the class of bipartite graphs respectively. We use \mathcal{C} , commonly with subscripts, to denote an unspecified hereditary graph class, and use $\mathcal{C}_1 \rightarrow \mathcal{C}_2$ to denote the vertex deletion problem from class \mathcal{C}_1 to class \mathcal{C}_2 :

Given a graph G in \mathcal{C}_1 , find a minimum set $V_- \subseteq V(G)$ such that $G - V_-$ is in \mathcal{C}_2 .

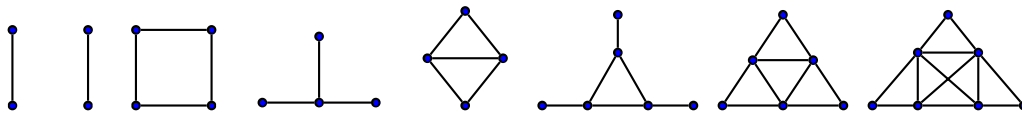
It is worth noting that \mathcal{C}_2 may or may not be a subclass of \mathcal{C}_1 , and when it is not, the problem is equivalent to $\mathcal{C}_1 \rightarrow \mathcal{C}_1 \cap \mathcal{C}_2$: Since \mathcal{C}_1 is hereditary, $G - V_-$ is necessarily in \mathcal{C}_1 . For almost all classes \mathcal{C} , the complexity of problems PLANAR $\rightarrow \mathcal{C}$ and BIPARTITE $\rightarrow \mathcal{C}$ has been answered in a systematical manner [26, 31], while for most graph classes \mathcal{C} , the complexity of problem DEGREE-BOUNDED $\rightarrow \mathcal{C}$ has been satisfactorily determined [19].

Apart from CHORDAL, we would also consider vertex deletion problems on its subclasses. Therefore, our purpose in this paper is a focused study on the algorithms and complexity of $\mathcal{C}_1 \rightarrow \mathcal{C}_2$ with both \mathcal{C}_1 and \mathcal{C}_2 being subclasses of CHORDAL. Since it is generally acknowledged that the study of chordal graphs motivated the theory of perfect graphs [24, 2], the importance of chordal graphs merits such a study from the aspect of structural graph theory. However, our main motivation is from the recent algorithmic progress in vertex deletion problems. It has come to our attention that to transform a graph to class \mathcal{C}_1 , it is frequently convenient to first make it a member of another class \mathcal{C}_2 that contains \mathcal{C}_1 as a proper subclass, followed by an algorithm for the $\mathcal{C}_2 \rightarrow \mathcal{C}_1$ problem [29, 9, 7, 33].

There being many subclasses of CHORDAL, the number of problems fitting in our scope is quite prohibitive. The following simple observations would save us a lot of efforts.

► **Proposition 1.** *Let \mathcal{C}_1 and \mathcal{C}_2 be two graph classes.*

- (1.) *If the $\mathcal{C}_1 \rightarrow \mathcal{C}_2$ problem is in P, then so is $\mathcal{C} \rightarrow \mathcal{C}_2$ for any subclass \mathcal{C} of \mathcal{C}_1 .*
- (2.) *If the $\mathcal{C}_1 \rightarrow \mathcal{C}_2$ problem is NP-complete, then so is $\mathcal{C} \rightarrow \mathcal{C}_2$ for any superclass \mathcal{C} of \mathcal{C}_1 .*



■ **Figure 1** Small subgraphs: $2K_2$, C_4 , claw ($K_{1,3}$), diamond, net, tent, rising sun.

For example, the majority of our hardness results for problems $\text{CHORDAL} \rightarrow \mathcal{C}$ are obtained by proving the hardness of $\text{SPLIT} \rightarrow \mathcal{C}$. Indeed, this is very natural as in literature, most (NP-)hardness of problems on chordal graphs is proved on split graphs, e.g., dominating set [3], Hamiltonian path [28], and maximum cut [5]. The most famous exception is probably the pathwidth problem, which can be solved in polynomial time on split graphs but becomes NP-complete on chordal graphs [23]. No problem like this surfaces during our study, though we do have the following hardness result proved directly on chordal graphs, for which we have no conclusion on split graphs.

► **Theorem 2.** *Let F be a biconnected chordal graph. If F is not complete, then the $\text{CHORDAL} \rightarrow F\text{-FREE}$ problem is NP-complete.*

Another simple observation of common use to us is about complement graph classes. The complement \overline{G} of graph G is defined on the same vertex set $V(G)$, where a pair of distinct vertices u and v is adjacent in \overline{G} if $uv \notin E(G)$. It is easy to see that the complement of \overline{G} is G . In Figure 1, for example, the net and the tent are the complements of each other. The complement of a graph class \mathcal{C} , denoted by $\overline{\mathcal{C}}$, comprises all graphs whose complements are in \mathcal{C} ; e.g., the complement of COMPLETE SPLIT is $\{2K_2, P_3\}$ -FREE. A graph class \mathcal{C} is self-complementary if it is its own complement, i.e., a graph $G \in \mathcal{C}$ if and only if $\overline{G} \in \mathcal{C}$. For example, both SPLIT and THRESHOLD are self-complementary. As usual, n denotes the number of vertices in the input graph. Note that we need an n^2 item because it takes $O(n^2)$ time to compute the complement of a graph.

► **Proposition 3.** *Let \mathcal{C}_1 and \mathcal{C}_2 be two graph classes. If the $\mathcal{C}_1 \rightarrow \mathcal{C}_2$ problem can be solved in $f(n)$ time, then the $\overline{\mathcal{C}}_1 \rightarrow \overline{\mathcal{C}}_2$ problem can be solved in $O(f(n) + n^2)$ time.*

Our results (besides Theorem 2) are summarize in Figure 2. Unfortunately, we have to leave the complexity of some problems open, particularly $\text{CHORDAL} \rightarrow \text{CLUSTER}$, $\text{CHORDAL} \rightarrow \text{UNIT INTERVAL}$, and $\text{INTERVAL} \rightarrow \text{UNIT INTERVAL}$.

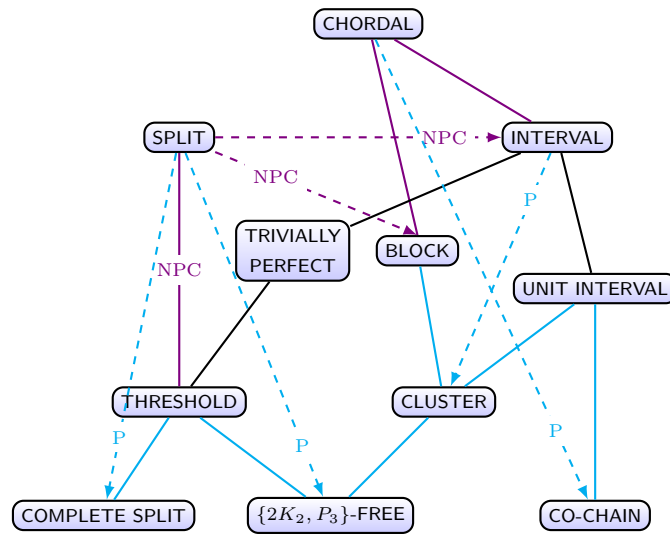
Let us also mention the approximation algorithms. All the problems have constant-ratio approximations, which follow from either [7, 8] or the general observation of Lund and Yannakakis [27]. On the other hand, none of the NP-complete problems admits a polynomial-time approximation scheme.

2 Preliminaries

All graphs discussed in this paper are undirected and simple. A graph G is given by its vertex set $V(G)$ and edge set $E(G)$, whose cardinalities will be denoted by n and m respectively.

For a subset $X \subseteq V(G)$, denote by $G[X]$ the subgraph induced by X , and by $G - X$ the subgraph $G[V(G) \setminus X]$; we use $E(X)$ as a shorthand for $E(G[X])$, i.e., all edges among vertices in X . For a subset $E_- \subseteq E(G)$ of edges, we use $G - E_-$ to denote the subgraph with vertex set $V(G)$ and edge set $E(G) \setminus E_-$. We write $G - v$ and $G - e$ instead of $G - \{v\}$ and $G - \{e\}$ for $v \in V(G)$ and $e \in E(G)$ respectively.

For $\ell \geq 2$, we use P_ℓ , K_ℓ , and I_ℓ to denote an induced path, a clique, and an independent set, respectively, on ℓ vertices. For $\ell \geq 4$, we use C_ℓ to denote an induced cycle on ℓ vertices;



■ **Figure 2** A summary of major graph classes studied by this paper and our results. Two classes are connected by a solid edge when the lower one is a subclass of the higher one. A directed dashed edge from \mathcal{C}_1 to \mathcal{C}_2 is used when \mathcal{C}_2 is not an immediate subclass of \mathcal{C}_1 . We omit here results implied by Proposition 1; e.g., $\mathcal{C} \rightarrow \text{CO-CHAIN}$ is in P for all \mathcal{C} , and $\text{CHORDAL} \rightarrow \mathcal{C}$ is NP-complete when \mathcal{C} is THRESHOLD, BLOCK, or INTERVAL. The cyan, violet, and black edges indicate that the complexity of the representing problems is in P, NP-complete, and unknown, respectively.

such a cycle is also called a *hole*. Some small graphs that will be used in this paper are depicted in Figure 1. Note that C_4 and $2K_2$ are complements to each other, while the complements of P_4 and C_5 are themselves.

We say that a graph G *contains* a subgraph F if F is isomorphic to some induced subgraph of G . A graph is *F-free* if it does not contain F ; for a set \mathcal{F} of graphs, a graph G is *F-free* if it is F -free for every $F \in \mathcal{F}$. Each set \mathcal{F} defines a hereditary graph class, and every hereditary graph class can be defined as such; in other words, for any hereditary graph class \mathcal{C} , there is a (possibly infinite) set \mathcal{F} of subgraphs such that a graph $G \in \mathcal{C}$ if and only if it is \mathcal{F} -free. Each graph F in \mathcal{F} is usually assumed to be minimal, in the sense that F is not in \mathcal{C} but every proper induced subgraph of F is; they are called the *minimal obstructions* of \mathcal{C} . One should note that a minimal obstruction of a graph class may not be a minimal obstruction of its subclass; e.g., the minimal obstruction C_5 of SPLIT is not a minimal obstruction of THRESHOLD, because C_5 contains the non-threshold graph P_4 as a proper induced subgraph.

The vertex deletion problem with object class \mathcal{C} can also be defined as finding a maximum subgraph in the class \mathcal{C} . For example, both vertex cover and independent set refer to the vertex deletion problem to the class EDGELESS, which is exactly the K_2 -free graphs. Although these formulations may behave different with respect to approximation, they are the same for our purpose. We may use both formulations interchangeably, dependent on which is more convenient in the context. Yet another way to view the vertex deletion problem toward property \mathcal{F} -free is to find a minimum set of vertices from a graph to hit all its induced subgraphs in \mathcal{F} .

We now define the graph classes we are going to study. Although the containment relationships of all the graph classes to be studied can be readily checked with their obstruction characterizations, sometimes it would be far more informative and inspiring if we look at them from the lens of the definitions and/or geometric representations of these graph classes.

A graph is *chordal* if every cycle of length larger than three has a chord, i.e., an edge between two non-consecutive vertices of the cycle. A graph is an *interval graph* if its vertices can be assigned to intervals on the real line such that there is an edge between two vertices if and only if their corresponding intervals intersect, and a *unit interval graph* if all the intervals have the same length. A graph G is a *trivially perfect graph* if for every induced subgraph of G , the size of the largest independent set is equivalent to the number of all maximal cliques [22]. Chordal graphs are precisely graphs that are intersection graphs of subtrees of a tree, while interval graphs are intersection graphs of sub-paths of a path. Therefore, INTERVAL \subset CHORDAL. A trivially perfect graph can be represented by a set of *non-overlapping* intervals; in other words, if two intervals intersect, then one is contained in the other. Therefore, TRIVIALY PERFECT \subset INTERVAL.

A graph is a *cluster graph* if every component is a clique. A graph is a *block graph* if the deletion of all cut vertices leaves a cluster graph. It is known that a graph is $\{2K_2, P_3\}$ -free if it is a cluster graph of which at most one clique is nontrivial, i.e., having more than one vertex. It is immediate from their definitions that $\{2K_2, P_3\}$ -FREE \subset CLUSTER \subset BLOCK. Moreover, block graphs are precisely those chordal graph of which any two maximal cliques share at most one vertex.

A graph is a *split graph* if its vertices can be partitioned into a clique C and an independent set I , and a *complete split graph* if every vertex in C is adjacent to all vertices in I ; we use $C \uplus I$ to denote the split partition. Note that either of the two sets may be empty. A graph G is a *threshold graph* if there is a real number t , the so-called *threshold*, and an assignment $f : V(G) \rightarrow \mathbb{R}$ such that $uv \in E(G)$ if and only if $f(u) + f(v) \geq t$ [11]. It is easy to verify that COMPLETE SPLIT \subset THRESHOLD \subset SPLIT: The first can be witnessed by $t = 1$ and assignment $f(v) = 1$ if $v \in C$ and 0 otherwise; and the second by the clique partition $\{v : f(v) \geq t/2\} \uplus \{v : f(v) < t/2\}$. Further, if we order the vertices in the independent set I of a threshold graph such that $f(v_1) \leq \dots \leq f(v_{|I|}) < t/2$, then $N(v_1) \subseteq \dots \subseteq N(v_{|I|})$. Likewise, there is an ordering of vertices $u_1, \dots, u_{|C|}$ in C such that $N[u_1] \subseteq \dots \subseteq N[u_{|C|}]$.

The reader may have noticed the striking resemblance between split graphs and bipartite graphs. Indeed, if we add edges to make one side of a bipartite graph into a clique, we end with a split graph; or equivalently, given a split graph G with split partition $C \uplus I$, the subgraph $G - E(C)$ is bipartite. Clearly, $G - E(C)$ is a complete bipartite graph if and only if G is a complete split graph. If G is a threshold graph, then $G - E(C)$ is a *chain graph* [31, 30]. Finally, CO-CHAIN denotes the complement of CHAIN.

Recall that Yannakakis [31] has given a dichotomy on the vertex deletion problem from bipartite graphs. Inspired by this and the aforementioned connection between bipartite graphs and split graphs, a natural attempt at problems SPLIT $\rightarrow \mathcal{C}$ would be reducing them to the corresponding problem on bipartite graphs (for algorithms) or the other way (for hardness results). This approach however turns out to be less straightforward as one may expect. See the full version for discussions.

3 Algorithmic results

This section gives the polynomial-time algorithms. Our focus would be laid on the use of structural properties, and if possible, we would present the simplest algorithms without elaborating on the implementation details. These problems may have more efficient algorithms, and with more complex data structures and algorithmic finesses, some of them may even be solved in linear time.

Our first two results are on split graphs, for which we need to put split partitions under

0. $\mathcal{S} \leftarrow \emptyset$;
1. build a bipartite graph G' by removing all edges among C from G ;
2. find a minimum vertex cover of G' , and add it to \mathcal{S} ;
3. **for each** $v \in I$ **do**
 - find a minimum vertex cover X of $G' - (C \setminus N(v)) - v$;
 - add $X \cup (C \setminus N(v))$ to \mathcal{S} ;
4. **return** a set in \mathcal{S} with the minimum cardinality.

■ **Figure 3** Algorithm for $\text{SPLIT} \rightarrow \{2K_2, P_3\}$ -FREE.

scrutiny. Let $C \uplus I$ be a split partition of a split graph G . If some vertex in I is completely adjacent to C , then we can move such a vertex v to C to make another split partition $C' = C \cup \{v\}$ and $I' = I \setminus \{v\}$. Note that the vertex v may not be unique, and the resulting graphs by moving them would be isomorphic. Moreover, after such a move, no vertex of I' can be completely adjacent to C' . The following proposition fully characterizes split graphs with more than one different split partition.

► **Proposition 4.** *Let G be a split graph with at least two split partitions, and let $C \uplus I$ and $C' \uplus I'$ be two different split partitions of G .*

- (i) *The difference between $|C|$ and $|C'|$ is at most 1.*
- (ii) *If $|C| = |C'| + 1$, then C is a maximum clique, and I' is a maximum independent set of G ; moreover, $C' \subset C$.*
- (iii) *If $|C| = |C'|$, then $G - E(C)$ and $G - E(C')$ are isomorphic.*

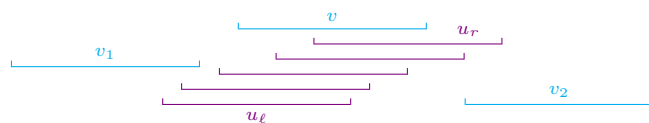
As a result, a split graph has either one or two essentially distinct split partitions. On the other hand, of all split partitions of a complete bipartite graph, only one, whose independent set is the largest, satisfies the definition of complete bipartite graphs, and we will exclusively refer to it when we are discussing a complete split graph.

Let G be a split graph with split partition $C \uplus I$ and let \underline{G} be a $\{2K_2, P_3\}$ -free subgraph of G . If \underline{G} has edges, all of them must be in the same nontrivial clique. At most one vertex of this clique can be from I ; therefore, all other vertices of I either are deleted or become isolated in \underline{G} . In other words, for each other vertex v in I , either v or all its neighbors have to be deleted.

► **Theorem 5.** *The $\text{SPLIT} \rightarrow \{2K_2, P_3\}$ -FREE problem is in P .*

Proof. Let G be the input graph to the $\text{SPLIT} \rightarrow \{2K_2, P_3\}$ -FREE problem and let $C \uplus I$ be a split partition of G . We use the algorithm in Figure 3 to find a minimum solution to G . To argue its correctness, we show that (i) every set in \mathcal{S} , added in step 2 or 3, is a solution to G , and (ii) at least one of them is minimum. For (i), it is easy to verify that any vertex cover of $G' = G - E(C)$ is a solution: There is no edge between C and I after its deletion. The situation in step 3 is similar; note that $N[v] \uplus (I \setminus \{v\})$ is a split partition of $G - (C \setminus N(v))$.

Let V_- be a minimum solution to G . In the first case, every vertex $v \in I \setminus V_-$ is isolated in $G - V_-$. In other words, V_- contains a vertex cover of $G' = G - E(C)$, and then the solution found by step 2 is already the minimum. Henceforth we assume that there exists a vertex $v \in I \setminus V_-$ such that $N(v) \not\subseteq V_-$. Since any vertex $u \in N(v)$ and $w \in C \setminus N(v)$ induce a P_3 with v , in this case all vertices in $C \setminus N(v)$ must be in V_- . Note that the vertex v is unique: If two vertices in $I \setminus V_-$ have neighbors in $C \setminus V_-$, then they are in a non-clique component. Therefore, after removing $C \setminus N(v)$ and v from the graph, it reduces to the first case. This justifies step 3.



■ **Figure 4** A connected split graph with split partition $C \uplus I$ that is also a unit interval graph. Violet intervals are for vertices in C and cyan for I . Note that the vertex v from I is completely adjacent to C .

The algorithm makes $O(n)$ calls to an algorithm for the bipartite vertex cover problem, each taking $O(m\sqrt{n})$ time, and hence the whole algorithm runs in $O(mn\sqrt{n})$ time. ◀

Noting that $\text{SPLIT} \cap \text{CLUSTER}$ is precisely $\{2K_2, P_3\}$ -FREE, we can apply the algorithm of Theorem 5 to the $\text{SPLIT} \rightarrow \text{CLUSTER}$ problem. Moreover, since SPLIT is self-complementary, while the complement of $\{2K_2, P_3\}$ -FREE is COMPLETE SPLIT , it follows from Proposition 3 that the $\text{SPLIT} \rightarrow \text{COMPLETE SPLIT}$ problem is also in P .

► **Corollary 6.** *Problems $\text{SPLIT} \rightarrow \text{CLUSTER}$ and $\text{SPLIT} \rightarrow \text{COMPLETE SPLIT}$ are in P .*

A similar observation can be used to solve the $\text{SPLIT} \rightarrow \text{UNIT INTERVAL}$ problem. We start from a simple property of connected graphs in $\text{SPLIT} \cap \text{UNIT INTERVAL}$.

► **Proposition 7.** *Let G be a connected split graph and let $C \uplus I$ be a split partition of G . If G is a unit interval graph, then $|I| \leq 3$, and the equality holds only when there is a vertex $v \in I$ adjacent to all vertices in C .*

Proof. We prove $|I| \leq 2$ if C is a maximum clique of G , and then the proposition follows from Proposition 4(i). Let u_ℓ and u_r be the vertices in C with respectively the leftmost and rightmost intervals. Suppose for contradiction $|I| > 2$. Let v_1 and v_2 be the vertices in I with respectively the leftmost and rightmost intervals. Then $\text{lp}(u_\ell) < \text{rp}(v_1) < \text{lp}(u_r) < \text{rp}(u_\ell) < \text{lp}(v_2) < \text{rp}(u_r)$, where the second and the fourth inequalities follow from that C is a maximum clique, and the others from the selections of the four vertices. Since G is connected, the interval for any other vertex v in $I \setminus \{v_1, v_2\}$, which is nonempty, has to lie in $(\text{rp}(v_1), \text{lp}(v_2))$. But then it has to contain $[\text{lp}(u_r), \text{rp}(u_\ell)]$, and $\{v\} \cup C$ is a clique, contradicting that C is a maximum clique of G . ◀

Similar as Theorem 5, our algorithm for $\text{SPLIT} \rightarrow \text{UNIT INTERVAL}$ separates into two cases, based on whether there is a vertex of $I \setminus V_-$ adjacent to all vertices in $C \setminus V_-$.

► **Theorem 8.** *The $\text{SPLIT} \rightarrow \text{UNIT INTERVAL}$ problem is in P .*

Proof. Let G be the input graph to the $\text{SPLIT} \rightarrow \text{UNIT INTERVAL}$ problem and let $C \uplus I$ be a split partition of G . We use the algorithm in Figure 5 to find a solution. To argue its correctness, we show that all sets put into \mathcal{S} in steps 1–4 are solutions to G , and at least one of them is minimum. It is clear for step 1. After the deletion of a solution found in step 2, only v in I remains adjacent to the remaining vertices of C . In step 3, only v_1 and v_2 from I can remain adjacent to vertices in C . In step 3.2, no vertex in C is adjacent to both v_1 and v_2 ; in step 3.3, every vertex in C is adjacent to at least one of v_1 and v_2 . In either case, it is easy to verify that the graph is a unit interval graph by building a unit interval model directly. Step 4 follows from the same argument as above: After the deletion of $C \setminus N(v)$, it reduces to one of the three previous steps.

```

0.  $\mathcal{S} \leftarrow \emptyset$ ;
1. solve the SPLIT  $\rightarrow \{2K_2, P_3\}$ -FREE problem on  $G$ ; add the solution to  $\mathcal{S}$ ;
    $\parallel$  case 1:
2. for each  $v \in I$  do
   find a minimum vertex cover of  $G - v - E(C)$ , and add it to  $\mathcal{S}$ ;
3. for each  $v_1, v_2 \in I$  do
3.1.  $G' \leftarrow G - \{v_1, v_2\} - E(C)$ ;
3.2. find a minimum vertex cover of  $G' - N(v_1) \cap N(v_2)$ ,
   and add its union with  $N(v_1) \cap N(v_2)$  to  $\mathcal{S}$ ;
3.3. find a minimum vertex cover of  $G' - C \setminus (N(v_1) \cup N(v_2))$ ,
   and add its union with  $C \setminus (N(v_1) \cup N(v_2))$  to  $\mathcal{S}$ ;
    $\parallel$  case 2:
4. for each  $v \in I$  do
    $G'' \leftarrow G - (C \setminus N(v))$  with split partition  $N[v]$  and  $I \setminus \{v\}$ ;
   solve  $G''$  as case 1, but append  $C \setminus N(v)$  to each solution found;
5. return a set in  $\mathcal{S}$  with the minimum cardinality.
    
```

■ **Figure 5** Algorithm for SPLIT \rightarrow UNIT INTERVAL.

Let V_- be a minimum solution to G . If $G - V_-$ is $\{2K_2, P_3\}$ -free, then the solution found by step 1 is the minimum. Henceforth we assume that $G - V_-$ contains a non-clique component U ; note that such a component contains all vertices in $C \setminus V_-$ and hence is unique.

In the first case, every vertex $v \in U \cap I$ has at least one non-neighbor in $C \setminus V_-$, i.e., $N(v) \setminus V_- \subset C \setminus V_-$. According to Proposition 7, $|U \cap I| \leq 2$. If $U \cap I = \{v\}$, then $G - (V_- \cup \{v\})$ is $\{2K_2, P_3\}$ -free and the only nontrivial clique $U \setminus \{v\}$ is a subset of C ; hence step 2 always find a minimum solution. In the rest of this case, $U \cap I$ has two different vertices; let them be v_1 and v_2 . Since any $u_1 \in N(v_1) \cap N(v_2)$ and $u_2 \in C \setminus (N(v_1) \cup N(v_2))$ induce a claw with $\{v_1, v_2\}$, at least one of the two sets needs to be empty or completely contained in V_- . Steps 3.2 and 3.3 take care of these two situations separately.

We are now in the second case, where $C \setminus V_- \subseteq N(v)$ for some vertex $v \in I \setminus V_-$; in other words, V_- contains all vertices in $C \setminus N(v)$. There might be two of such vertices, when we can take v to be either of them. Clearly, $N[v]$ and $I \setminus \{v\}$ is then a split partition of $G'' = G - (C \setminus N(v))$, which has a solution $V_- \setminus (C \setminus N(v))$. Moreover, under this new split partition, we reduce it to the first case.

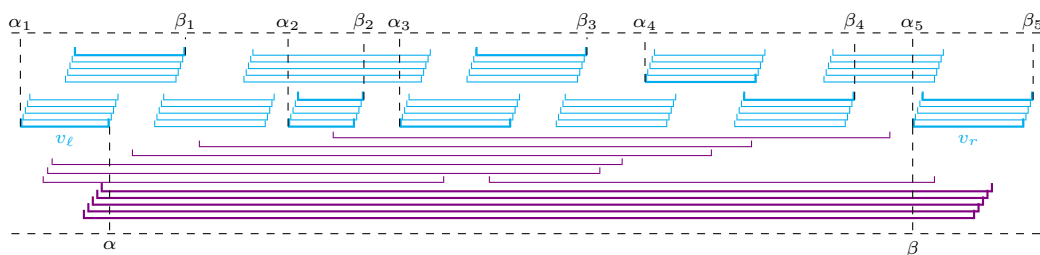
The algorithm makes $O(n^3)$ calls to the algorithm for the bipartite vertex cover problem, each taking $O(m\sqrt{n})$ time, and hence the whole algorithm runs in $O(mn^{3.5})$ time. ◀

We now turn to problems whose inputs are interval graphs, for which we rely on interval models. Recall that an interval model for an interval graph is a set of intervals representing its vertices. In this paper, all intervals are closed. An interval model can be specified by the $2n$ endpoints for the n intervals, the interval for vertex v being by $[\mathbf{lp}(v), \mathbf{rp}(v)]$.

For the UNIT INTERVAL \rightarrow COMPLETE SPLIT problem, the clique is from some maximal clique of the input graph G and can be enumerated. On the other hand, according to Proposition 7, there are at most three vertices in the independent set, which can be easily found. However, for interval graphs, it can be more complicated.

► **Theorem 9.** *Problems* INTERVAL \rightarrow COMPLETE SPLIT *and* INTERVAL \rightarrow CLUSTER *are in* P .

Proof. We solve both problems by finding the maximum subgraphs, for which we work on interval models. Let us fix an interval model for the input graph G ; we may assume without loss of generality that no distinct intervals can share an endpoint.



■ **Figure 6** Given is an interval model for an interval graph G .

The top line is for the INTERVAL \rightarrow CLUSTER problem. A maximum cluster subgraph of G has five cliques, each specified by a pair of α_i and β_i . (In this example, the maximum clique in each range $[\alpha_i, \beta_i]$ comprises all intervals in this range.)

The bottom line is for the INTERVAL \rightarrow COMPLETE SPLIT problem. A maximum complete split subgraph of G contains 12 vertices. The clique contains the vertices represented by the lowest five intervals, and the independent set contains v_ℓ and v_r , together with a maximum independent set of all intervals completely lying in $[\alpha, \beta]$.

For the INTERVAL \rightarrow COMPLETE SPLIT problem, we consider a maximum complete split subgraph $G[U]$. It is trivial if $G[U]$ is a clique; hence we assume otherwise. Let $C \uplus I$ be the split partition of $G[U]$, and let

$$\alpha = \text{rp}(v_\ell) = \min_{v \in I} \text{rp}(v) \text{ and } \beta = \text{lp}(v_r) = \max_{v \in I} \text{lp}(v).$$

Note that $|I| \geq 2$, as otherwise $G[U]$ is a clique; hence $v_\ell \neq v_r$ and $\alpha < \beta$. See Figure 6. It is easy to see that a vertex is in C if and only if its interval fully contains $[\alpha, \beta]$; on the other hand, the maximality of U requires us to take all such vertices. The independent set I would then consist of v_ℓ, v_r , and a maximum independent set of the subgraph induced by intervals satisfying $\alpha < \text{lp}(v) < \text{rp}(v) < \beta$. There are $O(n^2)$ pairs of indices to enumerate, and for each pair, both the clique and a maximum independent set can be found in $O(n)$ time. The whole algorithm runs in $O(n^3)$ time.

We now consider the INTERVAL \rightarrow CLUSTER problem. Suppose that $G[U]$ is a maximum cluster subgraph of G and that it has k cliques. For the i th clique B_i , we can find two endpoints

$$\alpha_i = \min_{v \in B_i} \text{lp}(v) \text{ and } \beta_i = \max_{v \in B_i} \text{rp}(v).$$

Then all intervals for vertices in B_i are completely contained in the interval $[\alpha_i, \beta_i]$. The k intervals defined as such are pairwise disjoint: There cannot be edges between two cliques in $G[U]$. Therefore, B_i must be a maximum clique in the subgraphs induced by $\{v : \alpha_i \leq \text{lp}(v) < \text{rp}(v) \leq \beta_i\}$, which can be found easily. See Figure 6. The problem can thus be reduced to find the k pairs of endpoints α_i and β_i .

We build another weighted interval model as follows. For each $\text{lp}(v_\ell)$ and each $\text{rp}(v_r)$ with $\text{lp}(v_\ell) < \text{rp}(v_r)$, possibly $v_\ell = v_r$, we add an interval $[\text{lp}(v_\ell), \text{rp}(v_r)]$, whose weight is set to be the size of maximum cliques in the subgraphs induced by $\{v : \text{lp}(v_\ell) \leq \text{lp}(v) < \text{rp}(v) \leq \text{rp}(v_r)\}$. We then find a set of pairwise disjoint intervals with the maximum weight sum (or equivalently, a maximum-weight independent set of the weighted interval graph represented by the new interval model). All the steps can be done in polynomial time. ◀

It is easy to verify the following greedy algorithm solves the TREE \rightarrow CLUSTER problem. We root the input graph at an arbitrary vertex, and work on any leaf at the lowest level: If it

22:10 Vertex Deletion Problems on Chordal Graphs

has siblings (i.e., its parent has degree larger than 2), then delete its parent and put it into the solution; otherwise the parent of its parent. As we see below, a similar idea would enable us to solve the `BLOCK` \rightarrow `CLUSTER` problem. Recall that a *block* (also known as biconnected component) of a graph G is a maximal biconnected subgraph of G . The *block-cut tree* of a block graph has a vertex for each block and for each cut vertex, and an edge for each pair of a block and a cut vertex that belongs to that block. Note that every block of a block graph is a clique.

► **Theorem 10.** *The `BLOCK` \rightarrow `CLUSTER` problem can be solved in polynomial time.*

Proof. We construct the block-cut tree T of the input graph G . A cut vertex v of G is denoted by the same label in T , while for a block vertex u of T , we use $B(u)$ to denote the vertices in the block of G . We arbitrarily root T at some block vertex. Note that all leaves of T are block vertices, and their neighbors are not; this invariant will be maintained during our algorithm. Until the tree becomes empty, the algorithm always picks a leaf vertex u at the lowest level. Let v be its parent. If v has other children, we remove v and its children from T and put v in the solution V_- . In the rest u is the only child of v ; let u' be the parent of v , and let v' be the parent of u' . If at least one vertex in the clique $B(u')$ is not a cut vertex, then we remove v, u from T and put v in V_- . Otherwise, we remove the subtree rooted at v' from T ; we put $B(u') \setminus \{v\}$ into the solution, and for each other child u_i of v' that is not a leaf, we solve the subgraph induced by $B(u_i)$ and its children. The correctness is quite straightforward, so we omit here. ◀

The proofs of the following results are left in the full version.

► **Theorem 11.** *The `CHORDAL` \rightarrow `CO-CHAIN` problem is in P .*

► **Theorem 12.** *For any $p > 1$, the `CHORDAL` \rightarrow κ_p -FREE problem is in P .*

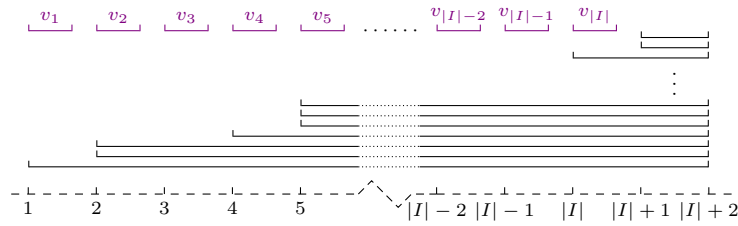
► **Theorem 13** ([14]). *The `CHORDAL` \rightarrow `SPLIT` problem is in P .*

4 Hardness

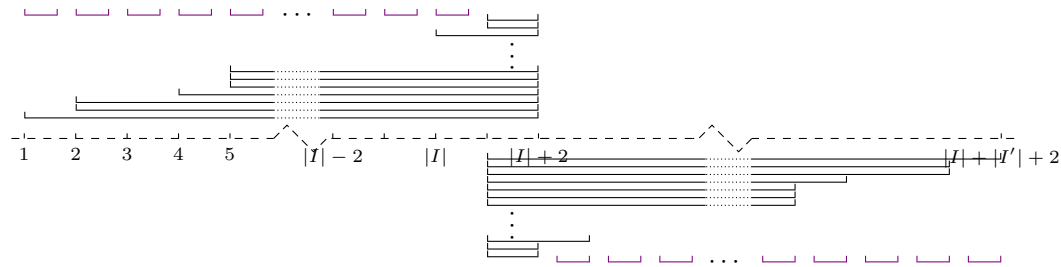
We now turn to hardness results. Here the problems should be understood to be their decision versions: The input includes, apart from a graph G from \mathcal{C}_1 , a positive integer k , and the problem is to decide whether G can be made a graph in \mathcal{C}_2 by deleting at most k vertices. All of them are in NP because all the concerned graph classes can be recognized in polynomial time. Our first hardness result, on `SPLIT` \rightarrow `THRESHOLD`, follows easily from the results of Yannakakis [31] on bipartite graphs. Recall that a bipartite graph is not a chain graph if and only if it contains some $2K_2$, and a split graph is not a threshold graph if and only if it contains some P_4 .

► **Lemma 14.** *The `SPLIT` \rightarrow `THRESHOLD` problem is NP-complete.*

Proof. Let G be a bipartite graph with partition C and I . We add all possible edges among C to make it a clique. Let G' be the resulting graph, which is clearly a split graph, witnessed by the split partition $C \uplus I$. We argue for every vertex set U that $G[U]$ is a chain graph, i.e., being $2K_2$ -free, if and only if $G'[U]$ is a threshold graph, i.e., being P_4 -free. Let X be any set of four vertices. If $G[X]$ is $2K_2$, then $|X \cap C| = |X \cap I| = 2$, but then $G'[X]$ would be a P_4 . The other direction can be argued similarly. Since the `BIPARTITE` \rightarrow `CHAIN` problem is NP-hard [31], the lemma follows. ◀



■ **Figure 7** The interval model for a threshold graph given by (1).



■ **Figure 8** The interval model for $G_1 \otimes G_2$.

Recall that every threshold graph is an interval graph, and this can be generalized as follows. Let G_1 and G_2 be two threshold graphs with split partitions $C \uplus I$ and $C' \uplus I'$ respectively. We let $G_1 \bowtie_{(C,C')} G_2$, or simply $G_1 \bowtie G'$ as in the rest of the paper the partitions are always clear from context, denote the graph obtained from them by adding all possible edges between C and C' —i.e., its vertex set and edge set are $V(G_1) \cup V(G_2)$ and $E(G_1) \cup E(G_2) \cup (C \times C')$ respectively. This is clearly a split graph with split partition $C \cup C'$ and $I \cup I'$. One can verify that $G_1 \bowtie G_2$ is also an interval graph by their obstructions as follows. A split graph that is not an interval graph has to contain a tent, a net, or a rising sun (see Figure 1). Each of them has three independent vertices, which have to be from $I \cup I'$, but a quick inspection of these three graphs will convince us that this cannot be possible.

► **Proposition 15.** *For any threshold graphs G_1, G_2 , the graph $G_1 \bowtie G_2$ is an interval graph.*

A better way to look at Proposition 15 is probably through interval models. Let G be a threshold graph with split partition $C \uplus I$, and let vertices in I be ordered in a way that $N(v_1) \subseteq N(v_2) \subseteq \dots \subseteq N(v_{|I|})$. We can build an interval model for G by setting intervals

$$\begin{aligned}
 & [i, i + 0.5] && \text{for every } v_i \in I, \\
 & [\min\{i : v_i \in N(v)\}, |I| + 2] && \text{for every } v \in N(I), \text{ and} \\
 & [|I| + 1, |I| + 2] && \text{otherwise (i.e., } v \in C \setminus N(I)).
 \end{aligned} \tag{1}$$

See Figure 7 for illustration.

An interval model for $G_1 \bowtie G_2$ can be built from the interval models for G_1 and G_2 by (i) keeping the intervals for G_1 , and (ii) setting the interval to be $[|I| + |I'| + 3 - \text{rp}(v), |I| + |I'| + 3 - \text{lp}(v)]$ for each $v \in V(G_2)$. See Figure 8.

► **Theorem 16.** *The SPLIT \rightarrow INTERVAL problem is NP-complete.*

Proof. It is clear that the problem is in NP. Let G be a split graph with split partition $C \uplus I$. We take a complete split graph G' with split partition $C' \uplus I'$, where $|C'| = |I'| = |C|$, and

let $H = G \bowtie G'$. We argue that (G, k) is a yes-instance of the SPLIT \rightarrow THRESHOLD problem if and only if (H, k) is a yes-instance of the SPLIT \rightarrow INTERVAL problem. Since both problems are trivial yes-instances when $k \geq |C|$, we may assume henceforth $k < |C|$.

Suppose that $G - V_-$, where $|V_-| \leq k$, is a threshold graph. According to Proposition 15, $(G - V_-) \bowtie G'$ is an interval graph. It is the same graph as $H - V_-$. Therefore, V_- is a solution of (H, k) . This verifies the only if direction.

Now suppose that $H - V_-$, where $|V_-| \leq k$, is an interval graph. Suppose for contradiction that $\underline{G} = G - (V_- \cap V(G))$ is not a threshold graph. Then \underline{G} must contain some P_4 ; let it be $v_1u_1u_2v_2$. Since \underline{G} is a split graph, we must have $u_1, u_2 \in C$ and $v_1, v_2 \in I$. On the other hand, by the assumption $k < |C|$, neither $C' \setminus V_-$ nor $I' \setminus V_-$ can be empty. Let $u \in C' \setminus V_-$ and $v \in I' \setminus V_-$. By the construction, the only edges between $\{u, v\}$ and $\{v_1, u_1, u_2, v_2\}$ are uu_1 and uv_2 , but then these six vertices together induce a net in $H - V_-$, a contradiction. \blacktriangleleft

► **Corollary 17.** *The CHORDAL \rightarrow INTERVAL problem is NP-complete.*

The last result is on the deletion of any biconnected subgraph from chordal graphs. Recall that a vertex v is *simplicial* in G if $N[v]$ is a clique. A graph is chordal if and only if we can make it empty by deleting simplicial vertices in the remaining graph [13].

► **Theorem 18.** *Let F be a biconnected chordal graph. If F is not complete, then the CHORDAL \rightarrow F -FREE problem is NP-complete. Moreover, if F is a complete split graph with $|C| = 2$ and $|I| \geq 2$, then the SPLIT \rightarrow F -FREE problem is NP-complete.*

Proof. We use the following reduction from the vertex cover problem. Let G be an input graph to the vertex cover problem, we conduct the following operations.

1. For each edge $uv \in E(G)$, add a distinct copy of F such that each of them uses uv as one of its edges. We say that u, v are the attachments for this copy of F .
2. Add all possible edges among $V(G)$ to make it complete.

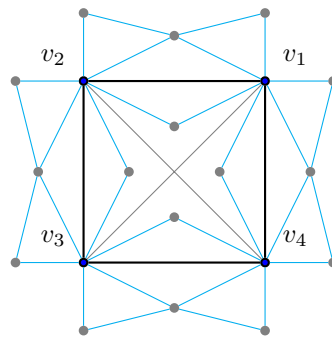
Let G' be the obtained graph. To see that G' is chordal, we give an explicit way of eliminating simplicial vertices to make G' empty. A chordal graph either is a clique or contains two nonadjacent simplicial vertices; all vertices are simplicial when it is a clique. For each copy of F , we can find a simplicial vertex in $V(F) \setminus \{u, v\}$. We keep doing this, and then only vertices in $V(G)$ remain. They have been made a clique, and thus all of them simplicial.

We argue that G has a vertex cover of size k if and only if we can delete k vertices from G' to make it F -free. The following fact would be essential. We consider any copy X of F with attachments u and v . If we delete u or v , then the other becomes a cut vertex, and $X \setminus \{u, v\}$ are in different blocks from other vertices of $V(G')$. But any other copy of F , if it exists, must be completely contained in a block, and thus it cannot contain any vertex in X .

Suppose that V_- is a vertex cover of size k in G . We claim that $\underline{G} = G' - V_-$ has no copy of F . For each copy of F with attachments u and v . Therefore, a copy of F in \underline{G} , if one exists, has all its vertices from $V(G)$. But this is not possible because F is not a clique.

Suppose now that V_- is a solution to G' of size k . We may assume that V_- contains no new vertex: If it contains a vertex from a copy of F with attachments u and v , we can replace it by u . (Note that the new set remains a solution to G' because the aforementioned fact.) Since $G' - V_-$ does not contain F , each copy of it has at least one of the attachments in V_- . Therefore, each edge of G , at least one end is in V_- , which means that V_- is a vertex cover of G . \blacktriangleleft

► **Corollary 19.** *Problems SPLIT \rightarrow BLOCK and CHORDAL \rightarrow BLOCK are NP-complete.*



■ **Figure 9** Reduction for Theorem 18, with F being a tent. The original graph G , drawn with blue vertices and thick edges, is a C_4 . The new vertices are gray and new edges thin. The set $\{v_1, v_3\}$ is a solution to both problems.

References

- 1 Akanksha Agrawal, Daniel Lokshtanov, Pranabendu Misra, Saket Saurabh, and Meirav Zehavi. Feedback vertex set inspired kernel for chordal vertex deletion. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1383–1398. SIAM, 2017.
- 2 Claude Berge. *Some classes of perfect graphs*. Internat. Computation Centre, 1966.
- 3 Alan A. Bertossi. Dominating sets for split and bipartite graphs. *Information processing letters*, 19(1):37–40, 1984.
- 4 Ivan Bliznets, Fedor V. Fomin, Michał Pilipczuk, and Yngve Villanger. Largest chordal and interval subgraphs faster than 2^n . In *European Symposium on Algorithms*, pages 193–204. Springer, 2013.
- 5 Hans L. Bodlaender and Klaus Jansen. On the complexity of the maximum cut problem. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 769–780. Springer, 1994.
- 6 Leizhen Cai. Fixed-parameter tractability of graph modification problems for hereditary properties. *Information Processing Letters*, 58(4):171–176, 1996.
- 7 Yixin Cao. Linear recognition of almost interval graphs. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1096–1115. Society for Industrial and Applied Mathematics, 2016.
- 8 Yixin Cao. Unit interval editing is fixed-parameter tractable. *Information and Computation*, 253:109–126, 2017.
- 9 Yixin Cao and Dániel Marx. Interval deletion is fixed-parameter tractable. *ACM Transactions on Algorithms (TALG)*, 11(3):21, 2015.
- 10 Yixin Cao and Dániel Marx. Chordal editing is fixed-parameter tractable. *Algorithmica*, 75(1):118–137, 2016.
- 11 Václav Chvátal and Peter L. Hammer. Aggregation of inequalities in integer programming. *Annals of discrete mathematics*, 1:145–162, 1977.
- 12 Marek Cygan and Marcin Pilipczuk. Split vertex deletion meets vertex cover: new fixed-parameter and exact exponential-time algorithms. *Information Processing Letters*, 113(5):179–182, 2013.
- 13 Gabriel Andrew Dirac. On rigid circuit graphs. In *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, volume 25, pages 71–76. Springer, 1961.
- 14 Tinaz Ekin and Dominique de Werra. On split-coloring problems. *Journal of Combinatorial Optimization*, 10(3):211–225, 2005.

- 15 Fedor V. Fomin, Serge Gaspers, Daniel Lokshtanov, and Saket Saurabh. Exact algorithms via monotone local search. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 764–775. ACM, 2016.
- 16 Fedor V. Fomin, Daniel Lokshtanov, Neeldhara Misra, and Saket Saurabh. Planar F-deletion: Approximation and optimal FPT algorithms. *FOCS'12*, pages 470–479, 2012.
- 17 Fedor V. Fomin, Ioan Todinca, and Yngve Villanger. Large induced subgraphs via triangulations and CMSO. *SIAM Journal on Computing*, 44(1):54–87, 2015.
- 18 Michael R. Garey and David S. Johnson. The rectilinear steiner tree problem is NP-complete. *SIAM Journal on Applied Mathematics*, 32(4):826–834, 1977.
- 19 Michael R. Garey and David S. Johnson. Computers and intractability: a guide to the theory of NP-completeness. *WH Free. Co., San Fr*, pages 90–91, 1979.
- 20 Michael R. Garey, David S. Johnson, and Larry Stockmeyer. Some simplified NP-complete graph problems. *Theoretical computer science*, 1(3):237–267, 1976.
- 21 Fănică Gavril. Algorithms for minimum coloring, maximum clique, minimum covering by cliques, and maximum independent set of a chordal graph. *SIAM Journal on Computing*, 1(2):180–187, 1972.
- 22 Martin Charles Golumbic. Trivially perfect graphs. *Discrete Mathematics*, 24(1):105–107, 1978.
- 23 Jens Gusted. On the pathwidth of chordal graphs. *Discrete Applied Mathematics*, 45(3):233–248, 1993.
- 24 András Hajnal and János Surányi. Über die auflösung von graphen in vollständige teilgraphen. *Ann. Univ. Sci. Budapest, Eötvös Sect. Math*, 1:113–121, 1958.
- 25 Bart M.P. Jansen and Marcin Pilipczuk. Approximation and kernelization for chordal vertex deletion. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1399–1418. SIAM, 2017.
- 26 John M. Lewis and Mihalis Yannakakis. The node-deletion problem for hereditary properties is NP-complete. *Journal of Computer and System Sciences*, 20(2):219–230, 1980.
- 27 Carsten Lund and Mihalis Yannakakis. The approximation of maximum subgraph problems. *Automata, Languages and Programming*, pages 40–51, 1993.
- 28 Haiko Müller. Hamiltonian circuits in chordal bipartite graphs. *Discrete Mathematics*, 156(1-3):291–298, 1996.
- 29 René Van Bevern, Christian Komusiewicz, Hannes Moser, and Rolf Niedermeier. Measuring indifference: Unit interval vertex deletion. In *WG*, pages 232–243. Springer, 2010.
- 30 Mihalis Yannakakis. Computing the minimum fill-in is NP-complete. *SIAM Journal on Algebraic Discrete Methods*, 2(1):77–79, 1981.
- 31 Mihalis Yannakakis. Node-deletion problems on bipartite graphs. *SIAM Journal on Computing*, 10(2):310–327, 1981.
- 32 Mihalis Yannakakis and Fănică Gavril. The maximum k -colorable subgraph problem for chordal graphs. *Information Processing Letters*, 24(2):133–137, 1987.
- 33 Jie You, Jianxin Wang, and Yixin Cao. Approximate association via dissociation. *Discrete Applied Mathematics*, 219:202–209, 2017.