



HAL
open science

On the hardness of problems around s-clubs on split graphs

Cristina Bazgan, Pinar Heggernes, André Nichterlein, Thomas Pontoizeau

► **To cite this version:**

Cristina Bazgan, Pinar Heggernes, André Nichterlein, Thomas Pontoizeau. On the hardness of problems around s-clubs on split graphs. *Discrete Applied Mathematics*, 2025, 364, pp.247-254. 10.1016/j.dam.2025.01.023 . hal-04941471

HAL Id: hal-04941471

<https://univ-paris-dauphine.hal.science/hal-04941471v1>

Submitted on 11 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



On the hardness of problems around s -clubs on split graphs

Cristina Bazgan^{a,*}, Pinar Heggenes^b, André Nichterlein^c, Thomas Pontoizeau^a

^a Université Paris-Dauphine, PSL Research University, CNRS, LAMSADE, Paris, 75016, France

^b Department of Informatics, University of Bergen, Norway

^c Algorithmics and Computational Complexity, Faculty IV, TU, Berlin, Germany



ARTICLE INFO

Article history:

Received 10 June 2024

Received in revised form 4 December 2024

Accepted 12 January 2025

Available online 24 January 2025

Keywords:

s -club

Split graphs

Bounded diameter

Spanning subgraph

Edge insertion

ABSTRACT

We investigate the complexity of problems related to s -clubs. Given a graph, an s -club is a subset of vertices such that the subgraph induced by it has diameter at most s . We show that partitioning a split graph into two 2-clubs is NP-hard. Moreover, we prove that finding the minimum number of edges to add to a split graph in order to obtain a diameter of at most 2 is $W[2]$ -hard with respect to the number of edges to add. Finally we show that finding the minimum number of edges to keep within a split graph of diameter 2 or 3 in order to maintain its diameter is NP-complete.

© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The recent development of social networks such as Facebook or LinkedIn and online dating services has motivated the investigation of community detection in such networks. A standard abstract model for those networks are graphs in which a community should intuitively describe some cohesion, oftenly corresponding to some density in a subgraph. The first intuitive way to define a community is to look for a group of people in which all members of this group know each other. In graphs, it corresponds to a clique. The problem of finding a clique of maximum size is a well known NP-hard problem [16]. However, considering communities as cliques is too restrictive: a subgraph with all possible internal edges except one would not be considered as a community under this assumption, even if it probably should be in real world social networks.

We can consider a less restrictive condition which still reflects cohesion. In this way, a community can be defined as a group of people such that every two members have a ‘chain’ of relationship between them: the first person knows someone who knows someone ... who knows the second member, with a restricted length for this chain. Given a graph, Mokken [19] introduced the notion of s -club which is a vertex set such that the subgraph induced by the vertex set has diameter at most s .

Problems around s -clubs have been well studied in the literature. A main one is, given a graph, to find an s -club of maximum size, which has been well-studied [2,6,13–15,22].

Another problem consists in finding a partition into k parts of vertices that are all s -clubs, that we discuss in this paper. This problem has been studied before [1,7,8,20]. The problem has been shown linear-time solvable in trees by Parley et al. [20]. Deogun et al. [8] proved that the problem is NP-complete for any $k \geq 3$ and $s \geq 2$ even in the case where the graph is both split and undirected path. Moreover, for $s = 2$ the minimum number of parts is bounded by

* Corresponding author.

E-mail addresses: bazgan@lamsade.dauphine.fr (C. Bazgan), pinar.heggenes@ii.uib.no (P. Heggenes), andre.nichterlein@tu-berlin.de (A. Nichterlein), thomas.pontoizeau@lamsade.dauphine.fr (T. Pontoizeau).

the domination number of a graph and is equal to the domination number on strongly chordal graphs [8]. Abbas and Stewart [1] proved that minimizing the number of parts in a partition of a graph into s -clubs is NP-hard for any $s \geq 2$ for bipartite graphs, NP-hard on chordal graphs, and also NP-hard for split graphs with $s = 2$ and any $k \geq 3$.

It is interesting to also study dynamic versions of community detection. In fact, real social networks are constantly changing and links between members can either appear or disappear. In this way, we also study problems related to s -clubs around adding and removing edges. Plesnik [21] studied the following problem: given a graph G , a cost function c on the edges and an integer B , finding a spanning subgraph G' of G with cost $\sum_{e \in G'} c(e) < B$ and with minimum diameter. The associated decision problem has been proven NP-complete. Biló et al. [4] studied the two following problems. Given a graph $G = (V, E)$, and two positive integers D and B , find a minimum-cardinality set E' of edges to be added to G in such a way that the diameter of $G' = (V, E \cup E')$ is at most D . Given a graph $G = (V, E)$, find a set E' of B edges to be added to G in such a way that the diameter of $G' = (V, E \cup E')$ is minimized. Both are known to be NP-hard. Deleting at most t edges to a graph in order to obtain a graph of diameter at least s was proven NP-hard for $k = |E| - |V| + 1$ $s = |V| - 1$ by Schoone et al. [23].

The paper is organized as follows. In Section 2, we formally introduce the problems around s -clubs we studied. In Section 3, we rectify an error in the literature [7] and show that the problem of finding a partition into two 2-clubs is NP-complete in split graphs. In Section 4, we discuss the problem of adding a minimum number of edges in a graph in order to become of diameter at most 2. In Section 5, we study the problem of finding the minimum number of edges to keep in a graph while maintaining its diameter.

2. Preliminaries

In this section we define the basic notions we use and the problems we study in the paper.

Given a graph $G = (V, E)$, we denote by xy the edge in E between two vertices x and y of V . A *partition* of V is a set of subsets $\{V_1, V_2, \dots, V_p\}$ of V for some integer p such that $\cup_{i=1}^p V_i = V$ and for any two $i, j \in \{1, 2, \dots, p\}$, $i \neq j$, $V_i \cap V_j = \emptyset$. A k -*partition* of V is a partition of V into k subsets. A *spanning subgraph* of G is a subgraph $H = (V(H), E(H))$ of G such that $V(H) = V$, $E(H) \subseteq E$, and H is connected. The k th *power* $G^k = (V_k, E_k)$ of a graph $G = (V, E)$ is a graph such that $V_k = V$, and for any two vertices $x, y \in V_k$, $xy \in E_k$ if and only if the length of a shortest path between x and y in G is smaller than or equal to k . For $k = 2$, we name such graph the *squared graph* of G .

For a vertex x adding a *twin copy* of x consists of adding a vertex x' that has the same set of neighbors as x . For any vertices x, y of a given graph, we denote $d(x, y)$ the length of a shortest path between x and y . Given a subset $V_1 \subseteq V$ we denote $d_{V_1}(x, y)$ the length of a shortest path between x and y in the subgraph induced by V_1 . The *diameter* of a (sub)graph is the longest distance between any pair of vertices in the (sub)graph. An s -*club* is a vertex subset $S \subseteq V$ such that the subgraph of G induced by S has diameter at most s . A *dominating set* is a set of vertices D such that any vertex of the graph is either in D or is adjacent to a vertex in D . A *split graph* is a graph such that its vertex set can be partitioned into two subsets, one inducing an independent set S and the other one inducing a clique K . Since we have several results in this class of graphs, we introduce some notation for an easier reading. For any split graph $G = (V = S \cup K, E)$, S always corresponds to the independent set and K to the clique. Moreover, for any dominating set D of G , we always consider that D is included in K . Indeed, if a vertex v of D belongs to S , we can take any neighbor v' of v from K in D instead of v from S without compromising the fact that D is a dominating set.

We are interested in the following decision problems:

In Section 3, we investigate the following problem:

k -PARTITION INTO s -CLUBS

- Input:** A graph $G = (V, E)$ and two integers k, s .
Question: Is there a partition $\{P_1, P_2, \dots, P_k\}$ of V , such that for each $i \in \{1, \dots, k\}$ the part P_i is an s -club?

In Section 4, we investigate the following problem:

s -CLUB EDGE INSERTION

- Input:** A graph $G = (V, E)$ and two integers s, t .
Question: Is there a set of edges E' of size at most t , such that V is an s -club in the graph $G' = (V, E \cup E')$?

In Section 5, we investigate the following problem:

SPANNING s -CLUB

- Input:** A graph $G = (V, E)$ and an integer t .
Question: Is there a set of edges $E' \subset E$ of size at most t , such that the graph $G' = (V, E')$ has diameter at most s ?

In order to prove some NP-completeness results, we use the following NP-complete problem [12]:

DOMINATING SET

Input: A graph $G = (V, E)$ and an integer t .
Question: Does G contain a size- t dominating set?

3. Partitioning into two 2-clubs

In this section we discuss which graphs can be partitioned into two (non-empty) 2-clubs. We start with bipartite graphs.

Since a 2-club in a bipartite graph is a biclique, partitioning a bipartite graph G into k 2-clubs is equivalent to partition G into k bicliques. A partition of a bipartite graph into two bicliques can be computed in polynomial time [12, problem GT15]. Moreover, partitioning a bipartite graph into k bicliques for any fixed $k \geq 3$ is NP-complete [10]. Thus, we can conclude that partitioning a bipartite graph into two 2-clubs can be done in polynomial time whereas partitioning a bipartite graph into k 2-clubs, for any fixed $k \geq 3$ is NP-complete.

Note that an interesting approach to detect s -clubs in graphs is to consider the power graph of the initial graph [7,13]. Chang et al. [7] claimed that the minimum number of parts in a partition into 2-clubs in a graph G equals the minimum number of cliques in a partition into cliques in G^2 . Actually, there exist graphs (even split graphs) in which there is no partition into two 2-clubs but the squared graph contains a partition into two cliques. Consider the graph in Fig. 1 as an example.

In order to prove that partitioning a graph into two 2-clubs is NP-complete on split graphs, we reduce from the NP-complete MONOTONE 3-SAT [12].

MONOTONE 3-SAT

Input: A set X of variables, a collection C of clauses over X which contains either only negated variables or only positive variables such that for each clause $c \in C$, $|c| = 3$.
Question: Is there a satisfying truth assignment for C ?

Theorem 1. 2-PARTITION INTO 2-CLUBS is NP-complete even on split graphs.

Proof. We reduce MONOTONE 3-SAT to 2-PARTITION INTO 2-CLUBS on split graphs. Let $I = (X, C)$ be an instance of MONOTONE 3-SAT with X the set of variables and C the set of clauses of size three, each clause being either positive or negative. We denote C_1 the subset of C of positive clauses and C_0 the subset of C of negative ones. We define a split graph $G = (V = S \cup K, E)$ as an instance of 2-PARTITION INTO 2-CLUBS as follows (see Fig. 2).

We construct two sets S_p and S_n that contain at the beginning the vertices $v_p \in S_p$ and $v_n \in S_n$, respectively. For each clause $c_i \in C$, we introduce two vertices $c_{i,1}$ and $c_{i,2}$. If c_i is a positive clause ($c_i \in C_1$), then these two vertices are introduced to S_p , otherwise to S_n . For each variable $x_i \in X$ we introduce a vertex x_i in a subset K_x . Notice that $|K_x| = |X|$. Let c_i be a clause that contains either x_j or \bar{x}_j , $x_j \in X$. Then add the two edges $x_j c_{i,1}$ and $x_j c_{i,2}$. For each pair of vertices $u, v \in S_p$ (resp. in S_n) that do not correspond to the same clause, that is, $\{u, v\} \subseteq S_p \times S_p \setminus \{\{c_{i,1}, c_{i,2}\} \mid c_i \in C_1\}$ (resp. $\{u, v\} \subseteq S_n \times S_n \setminus \{\{c_{i,1}, c_{i,2}\} \mid c_i \in C_0\}$), do the following: introduce a new vertex z in a subset K_p (resp. K_n) and introduce two edges uz, vz in E . Let $S = S_p \cup S_n$ and $K = K_p \cup K_x \cup K_n$. Join any two vertices of K by an edge in E so the subgraph induced by K becomes a clique, and thus we obtain a split graph $G = (S \cup K, E)$. Notice that $|S| = O(|C|)$ and $|K| = O(|C|^2 + |X|)$. Thus, the size of the instance G is polynomial in $|X|$ and $|C|$.

Now we show that there is an assignment for the variables of X such that every clause of C is satisfied if and only if there is a partition of V into two 2-clubs.

Suppose that there is an assignment satisfying all clauses from C . Then, we define the partition $\{V_1, V_0\}$ of V into two 2-clubs as follows. Let K_1 (resp. K_0) be the set of all vertices $x_i \in K_x$ such that the corresponding variables $x_i \in X$ have been assigned to true (resp. false). Define $V_1 = S_p \cup K_p \cup K_1$ (resp. $V_0 = S_n \cup K_n \cup K_0$). Next we show that for every pair of vertices $\{z_1, z_2\}$ in V_1 , $d_{V_1}(z_1, z_2) \leq 2$. First, for any $z_1, z_2 \in V_1 \cap K$, $d_{V_1}(z_1, z_2) = 1$. Moreover, for any $z_1 \in V_1 \cap K$, $z_2 \in V_1 \cap S$, $d_{V_1}(z_1, z_2) \leq 2$ since any vertex of $V_1 \cap S$ has a neighbor in $V_1 \cap K$. Finally, let $z_1, z_2 \in V_1 \cap S$. If they belong to the same clause (say $z_1 = c_{i,1}$ and $z_2 = c_{i,2}$ for some i), then $c_{i,1}$ and $c_{i,2}$ must share at least one common neighbor in K_1 as the corresponding clause $c_i \in C_1$ is satisfied by the assignment. Otherwise, they have a common neighbor in K_p by construction. Thus, z_1 and z_2 always have a common neighbor either in K_p or K_1 and $d_{V_1}(z_1, z_2) = 2$. Then, for any pair of vertices $\{z_1, z_2\}$ in V_1 , $d_{V_1}(z_1, z_2) \leq 2$. The same reasoning can be done for V_0 by symmetry. Thus, $\{V_1, V_0\}$ is a partition of V into two 2-clubs.

Suppose now that there is a partition $\{V_1, V_0\}$ of V such that V_1 and V_0 are 2-clubs. First notice that, by construction, $d(v_p, u) = 3$ for each $u \in S_n$. Symmetrically, $d(v_n, u) = 3$ for each $u \in S_p$. Thus, we can assume without loss of

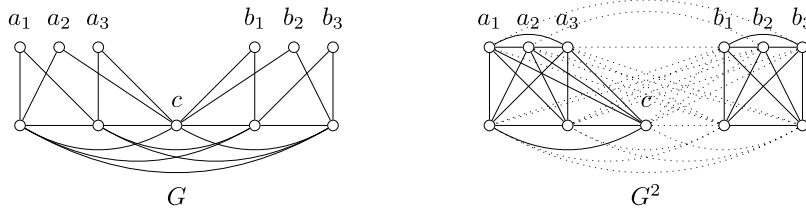


Fig. 1. A split graph in which there is no partition into two 2-clubs but there is a partition into two cliques in the squared graph. Some edges are dotted in G^2 in order to highlight the partition into two cliques. The vertices a_1 and b_1 are at distance 3 in G . Thus, if a partition into two 2-clubs exists in G , a_1 and b_1 must be in different parts. We can apply this reasoning for the vertices a_1 and b_2 , a_1 and b_3 and conclude that b_1, b_2, b_3 are in the same part. By symmetry, a_1, a_2, a_3 are also in the same part. Since the only common neighbor of a_2, a_3 and b_1, b_2 is c , c needs to be in both parts of G in order to have a partition into 2-clubs, which is impossible. Thus, there is no partition into two 2-clubs.

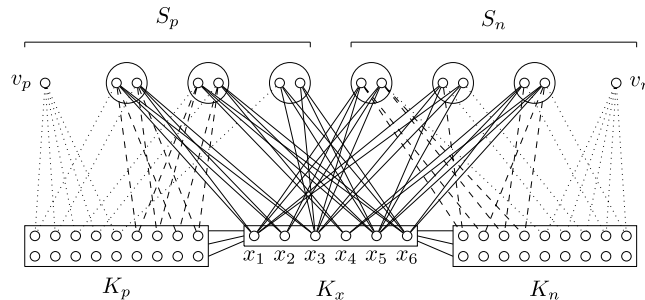


Fig. 2. The split graph G defined from the instance $I = (X, C)$ with $X = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ and $C = \{x_1 \vee x_2 \vee x_3, x_4 \vee x_5 \vee x_6, x_3 \vee x_5 \vee x_6, \bar{x}_1 \vee \bar{x}_2 \vee \bar{x}_3, \bar{x}_1 \vee \bar{x}_3 \vee \bar{x}_5, \bar{x}_4 \vee \bar{x}_5 \vee \bar{x}_6, \}$.

generality that $S_p \subset V_1$ and $S_n \subset V_0$. Consider the assignment α that sets every variable $x_j \in K_x \cap V_1$ to true and every variable $x_j \in K_x \cap V_0$ to false. Let $c_i \in C_1$ be an arbitrary positive clause. By construction of K_p , the two vertices $c_{i,1}$ and $c_{i,2}$ do *not* have a common neighbor in K_p . Since $c_{i,1}, c_{i,2} \in V_1$, which is a 2-club, it follows that there is a vertex $x_j \in K_x \cap V_1$ that is adjacent to $c_{i,1}$ and $c_{i,2}$. Thus, α sets x_j to true and satisfies c_i . Hence, α satisfies all clauses in C_1 . Similarly, for each $c_i \in C_0$ there is a vertex $x_j \in K_x \cap V_0$ that is adjacent to $c_{i,1}$ and $c_{i,2}$. Hence, α satisfies all clauses in C_0 as well. \square

Chang et al. [7] claimed that determining if the complement of the squared graph is bipartite allows to determine if there is a partition of a graph into two 2-clubs. Actually, this only allows to determine if there is a covering into two 2-clubs (i.e. a set $\{C_1, C_2\}$ of subsets of V such that $C_1 \cup C_2 = V$ and C_1, C_2 are 2-clubs). In fact, there is a covering into two 2-clubs if and only if the complement of the squared graph is bipartite. Indeed, if the complement of the squared graph is bipartite with the 2-partition (A, B) , a covering into two 2-clubs can be determined by considering the partition (A, B) and for any two vertices which are not at distance 2 in the subgraph induced by one part, add a common neighbor of those two vertices into this part (without removing it from the original part). Notice that the existence of such bipartition ensures that such vertex always exists. In Fig. 1, c would belong to both 2-clubs.

Partitioning 2-clubs into two 2-clubs. Theorem 1 implies that not every split graph can be partitioned into two 2-clubs. In contrast, this is always possible in diameter-two split graphs: Considering any vertex of the independent set in one part and the rest of the graph in the other part constitutes a 2-partition into two 2-clubs. For arbitrary graphs of diameter two, this changes again:

Theorem 2. *There exist graphs of diameter two that cannot be partitioned into two 2-clubs.*

We construct a graph G satisfying the properties of the above theorem, see Fig. 3 for an illustration of G . Start with a cycle $C = (\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2, \dots, \alpha_\ell, \beta_\ell, \gamma_\ell)$ on 3ℓ vertices with ℓ sufficiently large (any $\ell \geq 4$ will work). Add three vertices $U = \{u_\alpha, u_\beta, u_\gamma\}$ and add edges from all α - (β -, γ -)vertices to u_α (u_β, u_γ), respectively. Now, for each pair of vertices $v, w \in C \cup U$ that has distance more than two add a “connector-vertex” $c_{v,w}$ adjacent to v and w . Make all these connector-vertices a clique K . This completes the construction of $G = (V = C \cup U \cup K, E)$.

Lemma 3. *The constructed graph G is a 2-club that cannot be partitioned into two 2-clubs.*

Proof. It is easy to verify that G has indeed diameter two: the set K ensures that each pair of vertices in $C \cup U$ has distance at most two. Moreover, K is a clique and each vertex outside K has at least one neighbor in K .

Assume towards a contradiction that there is such a partition. That is, we can color each vertex red or blue so that the red and blue vertices induce the two 2-clubs G_r and G_b , respectively. Moreover, there is at least one red and one blue vertex.

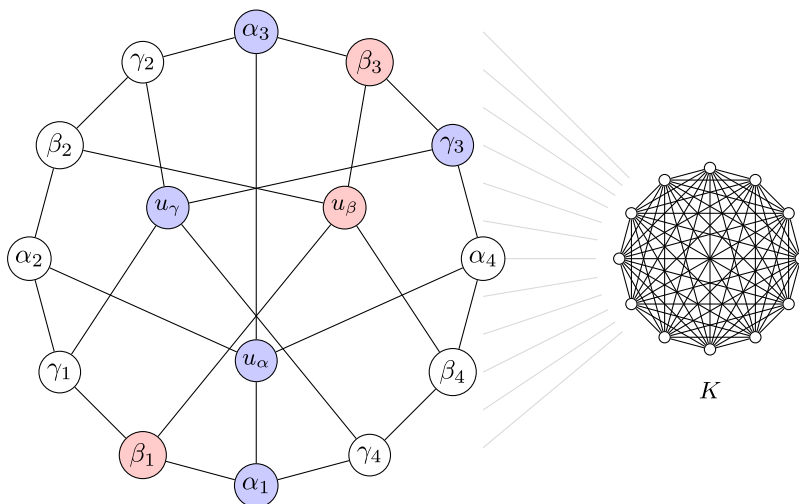


Fig. 3. The graph provided to show Theorem 2. For any pair of vertices on the left side, that have distance more than two, there is a common neighbor in the clique K on the right side. These edges between left and right are only indicated for better visibility. The coloring of the vertices on the left corresponds to the cases discussed in the proof of Lemma 3.

Assume that all vertices in C have the same color, say red. Then also all vertices in U have to be red (otherwise for some $v \in \{\alpha, \beta, \gamma\}$ we have $\text{dist}_{G_r}(v_1, v_2) > 2$). By construction, all vertices in K are red (otherwise G_r is not a 2-club), leaving no blue vertex, a contradiction.

We are left with the case where there are red and blue vertices in C . As all vertices are colored, this implies that there has to be a pair of adjacent vertices in C that has different colors. Assume without loss of generality (the construction is symmetrical) that α_1 is blue and its neighbor β_1 is red, see also Fig. 3. Then u_α cannot be red as otherwise the only common neighbor of the red u_α and β_1 is the blue α_1 , contradicting the fact that G_r has diameter two. Thus, u_α is blue. Similarly, u_β is red. It follows that at most one α_i is red and at most one β_i is blue.

Consider the case where u_γ is blue (the case where it is red is analogous). Thus, at most one γ_i is red. Since $\ell \geq 4$, it follows that there is an $i \in [\ell]$ so that α_i and γ_i are blue and β_i is red (in Fig. 3 this is $i = 3$). This is a contradiction to G_b having diameter two as β_i is the only common neighbor of α_i and γ_i . \square

4. Edge Insertion

The s -CLUB EDGE INSERTION problem has been proved NP-complete for $s = 3$ by Schoone et al. [23] and NP-complete for $s = 2$ by Li et al. [17]. The case $s = 1$ is trivial since it corresponds to adding edges between every pair of nonadjacent vertices. Gao et al. [11] proved the $W[2]$ -hardness of the problem w.r.t. the number t of edges to add, for any $s \geq 2$ by establishing a reduction from DOMINATING SET. Figiel et al. [9] extended the $W[2]$ -hardness to graphs of diameter three.

We strengthen this further and prove that 2-CLUB EDGE INSERTION is $W[2]$ -hard even on split graphs. Notice that even if a solution is not required to be a split graph, we show in Theorem 4 that it is always possible to obtain a split graph of diameter 2 with less edges than any solution.

Theorem 4. 2-CLUB EDGE INSERTION is $W[2]$ -hard w.r.t. the number t of edges to add, even on split graphs.

Proof. We reduce DOMINATING SET on split graphs of diameter 2, which has been proved $W[2]$ -hard by Lokshtanov et al. [18], to 2-CLUB EDGE INSERTION on split graphs. Let $G = (V = S \cup K, E)$ be a split graph of diameter 2, instance of DOMINATING SET, where S corresponds to the independent set and K to the clique. We construct an instance $G' = (V', E')$ of 2-CLUB EDGE INSERTION as follows. Consider a copy of G and add two new vertices s_0, k_0 . The graph G' is a split graph with $V' = S' \cup K'$ where $S' = S \cup \{s_0\}$ and $K' = K \cup \{k_0\}$. The set E' is obtained from E by adding edges between k_0 and every vertex $v \in K$ and the edge k_0s_0 (see Fig. 4). We show now that there is a dominating set of size at most t in G if and only if we can add at most t edges to G' such that G' has diameter 2.

Suppose that D is a dominating set of size t in G . We assume that $D \subseteq K$. The graph $(V', E' \cup D')$ where $D' = \{s_0x : x \in D\}$ has diameter 2 since G has diameter 2, and any vertex from S has a neighbor in D and then s_0 is at distance two from any vertex from S .

Suppose now that D' is a set of non edges of G' of size t such that $G'' = (V', E' \cup D')$ has diameter 2. We first show that we can assume that all edges in D' are between s_0 and K . Let $xy \in D'$. If $x = s_0$ and $y \in S$, let y' be a neighbor of y in K and the graph $G'' = (V', E' \cup D' \cup \{s_0y'\} \setminus \{xy\})$ is still of diameter 2. If both x and y are different from s_0 and $x \in S, y \in K$,

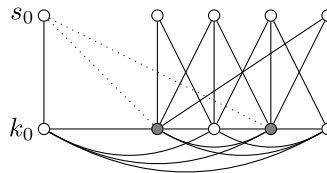


Fig. 4. The graph G' and a set of edges (represented by dotted lines) of minimum size to add to make G' having diameter 2 (a minimum dominating set is given in gray).

then the graph $G'' = (V', E' \cup D' \setminus \{xy\})$ is still of diameter 2 since G has diameter 2. If now both x and y are different from s_0 and $x \in S, y \in S$, then the graph $G'' = (V', E' \cup D' \setminus \{xy\})$ is still of diameter 2 since G has diameter 2. Thus, we can assume that all edges in D' are between s_0 and K (by updating D' if necessary) and $|D'| \leq t$. Then, the set of vertices adjacent to s_0 by an edge in D' is necessarily a dominating set of G since s_0 and any vertex of S must be at distance 2. Thus, we obtain a dominating set in G of size at most t . \square

5. Spanning subgraphs

We investigate the complexity of SPANNING s -CLUB. We show that in split graphs, SPANNING 2-CLUB and SPANNING 3-CLUB are NP-complete while SPANNING s -CLUB is polynomial time solvable for any $s \geq 4$. We start with the latter.

SPANNING s -CLUB is easy to solve for any $s \geq 4$ in split graphs. Indeed, for any split graph $G = (S \cup K, E)$, define the partial graph $G' = (V', E')$ as a tree such that for any $v \in S$, choose a unique edge $e \in E$ adjacent to v in G and let introduce e in E' . Then choose any vertex $x \in K$ and for any $x' \in K$ such that $x \neq x'$, let xx' be in E' . The remaining graph G' has diameter 4 and the number of edges in E' is minimum since G' is a tree (see Fig. 5).

In order to prove that SPANNING 3-CLUB is NP-complete on split graphs, we introduce the following lemma:

Lemma 5. *Let $G = (V = S \cup K, E)$ be a split graph. Let $G' = (V, E')$ be a spanning subgraph of G . If G' has diameter three and every vertex in S has degree one in G' , then the neighbors $N_{G'}(S)$ of S in G' form a clique and a dominating set for G .*

Proof. Since the spanning subgraph $G' = (V, E')$ is of diameter 3 and any vertex in S has degree one in G' then $N_{G'}(S)$ is a clique in G' . Assume by contradiction that it is not the case. Then there are two vertices $u, v \in N_{G'}(S)$ such that $uv \notin E(G')$. Thus, the vertices from S adjacent to u, v in G' are at distance more than 3, contradicting the assumption that G' is a 3-club. Moreover, by definition, each vertex in S has a neighbor in $N_{G'}(S)$. Also each vertex in $K \setminus N_{G'}(S)$ has a vertex in $N_{G'}(S)$ as the vertices in S have degree one. Thus, $N_{G'}(S)$ is a dominating set for G . \square

This lemma provides a clear way to show NP-completeness of SPANNING 3-CLUB. Reduce from DOMINATING SET on split graphs and modify the graph such that a minimum dominating set remains unchanged but the minimum spanning 3-club has all vertices in the independent set as degree-one vertices. This is exactly what we are doing to obtain the next result.

Theorem 6. SPANNING 3-CLUB is NP-complete even on split graphs.

Proof. We reduce from the NP-complete problem DOMINATING SET on split graphs [3] to SPANNING 3-CLUB on split graphs. Let $(G = (V = S \cup K, E), b)$ be an instance of DOMINATING SET, where S corresponds to the independent set, K to the clique, and b to the size of the dominating set. We construct an instance $\widehat{G} = (\widehat{V}, \widehat{E})$ of SPANNING 3-CLUB starting from G and inserting $\ell = |K|^2$ twin copies v^1, \dots, v^ℓ for each vertex $v \in S$. We show that there is a dominating set of size at most b in G if and only if there is a spanning subgraph containing at most $b(b-1)/2 + |\widehat{V}| - b$ edges which is a 3-club.

Let D be a dominating set of size b in $G, D \subseteq K$. Define the following spanning subgraph $G' = (\widehat{V}, E')$ with $E' \subset E(\widehat{G})$. For each pair of vertices $\{x, y\}$ in D, xy belongs to E' . Moreover, for each $v \in \widehat{V} \setminus D$, choose any edge vx with $x \in D$ in E' . Note that $|E'| = b(b-1)/2 + |\widehat{V}(G)| - b$ and it is easy to see that G' is a spanning subgraph of diameter 3.

Now let $G' = (\widehat{V}, E')$ be a spanning subgraph of \widehat{G} of diameter 3 with $|E'| \leq b(b-1)/2 + |\widehat{V}(G)| - b$ for some integer b . If there exists at least one group of ℓ twin vertices v^1, \dots, v^ℓ such that each of these vertices has at least two neighbors in G' , then we change G' as follows: make K a clique in G' and have for each vertex v^i one neighbor in K . This results in another spanning subgraph G'' of diameter 3. Moreover, G'' has potentially $\binom{|K|}{2}$ more edges than G' (making K a clique) but saves $\ell = |K|^2 > \binom{|K|}{2}$ edges by having all v^i of degree one. Thus, overall G'' has less edges than G' . Applying this argument to each twin group results in a spanning subgraph G''' of diameter 3 where each twin group has at least one vertex v^i of degree one. Let S' be the set of all these degree-one vertices. By Lemma 5 the set $N_{G'''}(S')$ is a dominating set for $G'[K \cup S']$. Since S' contains a vertex from each twin group, that is, for each vertex $v \in S$, it follows that $N_{G'''}(S')$ is a dominating set for G . Moreover, since G''' contains at most $b(b-1)/2 + |\widehat{V}(G)| - b$ edges and $N_{G'''}(S')$ is, by Lemma 5, a clique, it follows that $|N_{G'''}(S')| \leq b$. Thus, $N_{G'''}(S')$ is a dominating set of size at most b for G . \square



Fig. 5. A split graph and its spanning subtree of diameter 4.

Theorem 7. SPANNING 2-CLUB is NP-complete even on split graphs.

Proof. We reduce DOMINATING SET, which remains NP-complete on split graphs of diameter 2 [18], to SPANNING 2-CLUB on split graphs. Let $G = (V = S \cup K, E)$ be the split graph of diameter 2 in the DOMINATING SET instance, where S corresponds to the independent set and K to the clique. Set $\ell = |E| + 1$. We construct an instance \widehat{G} of SPANNING 2-CLUB as follows. Let S' be an independent set of size ℓ that will be completely connected to K in \widehat{G} . Thus, $\widehat{G} = (\widehat{V}, \widehat{E}) = (S \cup S' \cup K, E \cup (S' \times K))$. We show that there is a dominating set of size at most b in G if and only if there is a spanning diameter-2 subgraph in \widehat{G} containing at most $b\ell + |E|$ edges.

Let D be a dominating set of size at most b in G , $D \subseteq K$. Consider the following spanning subgraph $G' = (\widehat{V}, E')$ with $E' \subseteq \widehat{E}$, where E' contains all edges from E and edges between any vertex in D and any vertex in S' , that is $|E'| = |E| + \ell|D| \leq |E| + \ell b$. Clearly G' is a 2-club.

Now let $G'' = (\widehat{V}, E'')$ be a spanning subgraph of \widehat{G} of diameter 2 such that $|E''| \leq b\ell + |E|$. Since $b\ell + |E| < (b + 1)\ell$, there exists $v' \in S'$ such that $d_{G''}(v') \leq b$ (if not, then $d_{G''}(v') \geq b + 1$ for any $v' \in S'$ and $|E''| \geq (b + 1)\ell$ that is a contradiction). Since G'' is a 2-club then $N_{G''}(v')$ is a dominating set of size at most b in G . \square

The above two theorems imply that SPANNING s -Club is NP-complete in general graphs for any $s \geq 1$: This can be proven by using the ℓ -path extension of a graph: For any $\ell \geq 1$, the ℓ -path extension of a graph G is the graph G' constructed from the graph G by adding for each vertex v of G a path of length ℓ starting at v (see Fig. 6). Notice that a graph G is the 0-path extension of itself.

Theorem 8. SPANNING s -Club is NP-complete on general graphs for any $s \geq 2$.

Proof. For $s = 2$ and $s = 3$ the result follows from Theorems 6 and 7. Thus, let $s \geq 4$ be an integer.

If s is even, then we reduce from SPANNING 2-CLUB, otherwise from SPANNING 3-CLUB. In both cases, the reduction is the same: Set $\ell = \lfloor (s - 2)/2 \rfloor$. Given the graph $G = (V, E)$ we construct the ℓ -path extension \widehat{G} of G . Subsequently, we consider only the case where s is odd. The case where s is even is analogous.

We show that G has a spanning 3-club containing at most t edges if and only if \widehat{G} contains a spanning s -club containing at most $t + \ell|V|$ edges.

Let H be a spanning 3-club with at most t edges. Hence, we set \widehat{H} as H plus all the added paths with the ℓ -path extension. Clearly, \widehat{H} has at most $t + \ell|V|$ edges. Moreover, \widehat{H} is an s -club: Any pair of vertices within H has distance at most 3. Thus, any pair of vertices in \widehat{H} has distance at most $3 + 2\ell = s$.

Conversely, let \widehat{H} be a spanning s -club in \widehat{G} with at most $t + \ell|V|$ edges. We claim that $H := \widehat{H}[V]$ is a spanning 3-club with at most t edges: As \widehat{H} is connected, it contains all $\ell|V|$ edges in the added paths. Thus, H has at most t edges. Moreover, we claim H has diameter 3. Assume otherwise: There are two vertices u, v at distance 4 or more. However, then the endpoints of the two ℓ -paths starting in u and v , respectively, have distance at least $4 + 2\ell = s + 1$ in \widehat{H} , a contradiction. \square

6. Conclusion

In this paper, we saw that partitioning a graph into two 2-clubs is harder than expected [7]. We further proved that 2-CLUB EDGE INSERTION is $W[2]$ -hard with respect to the number of edges to add. We also showed that SPANNING 2-CLUB and SPANNING 3-CLUB are NP-complete. All these hardness results also hold on split graphs. Moreover, SPANNING s -CLUB is shown NP-complete on general graphs for any integer $s \geq 2$. It will be interesting to determine large subclasses of split graphs [5] for which our problems are still NP-complete or they become polynomial-time solvable. Finally, what is the complexity of determining whether a 2-club (or in general s -club) can be partitioned into two 2-clubs (two s -clubs)?

Data availability

No data was used for the research described in the article.

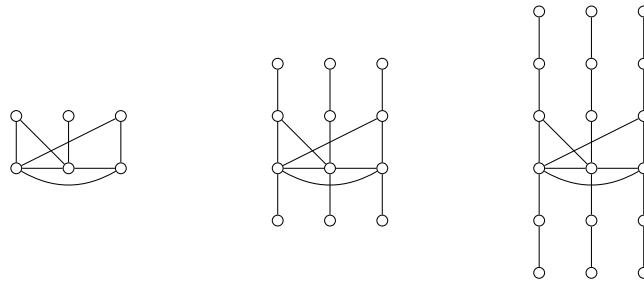


Fig. 6. A split graph and its 1-path extension and 2-path extension.

References

- [1] N. Abbas, L.K. Stewart, Clustering bipartite, chordal graphs: Complexity, sequential, parallel algorithms, *Discrete Appl. Math.* 91 (1–3) (1999) 1–23.
- [2] Y. Asahiro, E. Miyano, K. Samizo, Approximating maximum diameter-bounded subgraphs, *LATIN 2010*, in: *Proceedings of the 9th Latin American Symposium on Theoretical Informatics*, Vol. LNCS 6034, 2010, pp. 615–626.
- [3] A.A. Bertossi, Dominating sets for split and bipartite graphs, *Inform. Process. Lett.* 19 (1) (1984) 37–40.
- [4] D. Bilò, L. Gualà, G. Proietti, Improved approximability and non-approximability results for graph diameter decreasing problems, *Theoret. Comput. Sci.* 417 (2012) 12–22.
- [5] A. Brandstädt, V.B. Le, J.P. Spinrad, *Graph classes: A survey*, in: *SIAM Monographs on Discrete Mathematics and Applications*, 1987.
- [6] M. Chang, L. Hung, C. Lin, P. Su, Finding large k -clubs in undirected graphs, *Computing* 95 (9) (2013) 739–758.
- [7] J.M. Chang, J.S. Yang, S.L. Peng, On the complexity of graph clustering with bounded diameter, in: *Proceedings of the 18th International Computer Science and Engineering Conference, ICSEC 2014*, IEEE, 2014, pp. 18–22.
- [8] J.S. Deogun, D. Kratsch, G. Steiner, An approximation algorithm for clustering graphs with dominating diametral path, *Inform. Process. Lett.* 61 (3) (1997) 121–127.
- [9] A. Figiel, A-S. Himmel, A. Nichterlein, R. Niedermeier, On 2-clubs in graph-based data clustering: Theory and algorithm engineering, *J. Graph Algorithms Appl.* 25 (1) (2021) 521–547.
- [10] H. Fleischner, E. Mujuni, D. Paulusma, S. Szeider, Covering graphs with few complete bipartite subgraphs, *Theoret. Comput. Sci.* 410 (21–23) (2009) 2045–2053.
- [11] Y. Gao, D.R. Hare, J. Nastos, The parametric complexity of graph diameter augmentation, *Discrete Appl. Math.* 161 (10–11) (2013) 1626–1631.
- [12] M.R. Garey, D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman & Co, 1979.
- [13] P. Golovach, P. Heggernes, D. Kratsch, A. Rafiey, Finding clubs in graph classes, *Discrete Appl. Math.* 174 (2014) 57–65.
- [14] S. Hartung, C. Komusiewicz, A. Nichterlein, Parameterized algorithmics and computational experiments for finding 2-clubs, *J. Graph Algorithms Appl.* 19 (1) (2015) 155–190.
- [15] S. Hartung, C. Komusiewicz, A. Nichterlein, O. Suchý, On structural parameterizations for the 2-club problem, *Discrete Appl. Math.* 185 (2015) 79–92.
- [16] R.M. Karp, *Reducibility Among Combinatorial Problems*, Springer US, 1972, pp. 85–103.
- [17] C. Li, S.T. McCormick, D. Simchi-Levi, On the minimum-cardinality-bounded-diameter and the bounded-cardinality-minimum-diameter edge addition problems, *Oper. Res. Lett.* 11 (1992) 303–308.
- [18] D. Lokshantov, N. Misra, G. Philip, M.S. Ramanujan, S. Saurabh, Hardness of r -dominating set on graphs of diameter $(r-1)$, *IPEC 2013*, in: *Proceedings of the 8th International Symposium on Parameterized and Exact Computation*, Vol. LNCS 8246, 2013, pp. 255–267.
- [19] R. Mokken, Cliques, clubs and clans, *Qual. Quantity: Int. J. Methodol.* 13 (2) (1979) 161–173.
- [20] A. Parley, S. Hedetniemi, A. Proskurowski, Partitioning trees: Matching, domination, and maximum diameter, *Int. J. Comput. Inf. Sci.* 10 (1) (1981) 55–61.
- [21] J. Plesnik, The complexity of designing a network with minimum diameter, *Networks* 11 (1) (1981) 77–85.
- [22] A. Schäfer, C. Komusiewicz, H. Moser, R. Niedermeier, Parameterized computational complexity of finding small-diameter subgraphs, *Optim. Lett.* 6 (5) (2012) 883–891.
- [23] A.A. Schoone, H.L. Bodlaender, J.V. Leeuwen, Diameter increase caused by edge deletion, *J. Graph Theory* 11 (3) (1987) 409–427.