Journal of Graph Algorithms and Applications
http://jgaa.info/ vol. 23, no. 2, pp. 271–292 (2019)
DOI: 10.7155/jgaa.00491

Covering a Graph with Clubs

Riccardo Dondi[1] *Giancarlo Mauri*[2] *Florian Sikora*[3] *Italo Zoppis*[2]

[1]Università degli Studi di Bergamo, Bergamo, Italy
[2] Università degli Studi di Milano-Bicocca, Milano - Italy
[3]Université Paris-Dauphine, PSL University,
CNRS, LAMSADE, 75016 Paris, France

Abstract

Finding cohesive subgraphs in a network has been investigatSeveral alternative formulations of cohesive subgraph have been proposed, a notable one of them is s-club, which is a subgraph whose diameter is at most s. Here we consider a natral variant of the well-known Minimum Clique Cover problem, where we aim to cover a given graph with the minimum number of s-clubs, instead of cliques. We study the computational and approximation complexity of this problem, when s is equal to 2 or 3. We show that deciding if there exists a cover of a graph with three 2-clubs is NP-complete, and that deciding if there exists a cover of a graph with two 3-clubs is NP-complete. Then, we consider the approximation complexity of covering a graph with the minimum number of 2-clubs and 3-clubs. We show that, given a graph $G = (V, E)$ to be covered, covering G with the minimum number of 2-clubs is not approximable within factor $O(|V|^{1/2-\varepsilon})$, for any $\varepsilon > 0$, and covering G with the minimum number of 3-clubs is not approximable within factor $O(|V|^{1-\varepsilon})$, for any $\varepsilon > 0$. On the positive side, we give an approximation algorithm of factor $2|V|^{1/2} \log^{3/2} |V|$ for covering a graph with the minimum number of 2-clubs.

Submitted:	Reviewed:	Revised:	Accepted:	Final:
September 2018	February 2019	February 2019	March 2019	March 2019
		Published:		
		April 2019		
	Article type:		Communicated by:	
	Regular paper		P. Ferragina	

A preliminary version of this paper appeared in [9]. Research partially supported by the project "ESIGMA" (ANR-17-CE23-0010).

E-mail addresses: riccardo.dondi@unibg.it (Riccardo Dondi) mauri@disco.unimib.it (Giancarlo Mauri) florian.sikora@dauphine.fr (Florian Sikora) zoppis@disco.unimib.it (Italo Zoppis)

1 Introduction

The quest for modules inside a network is a well-known and deeply studied problem in network science, with applications in different fields, for example the analysis of biological or social network. A highly investigated problem is that of finding cohesive subgroups inside a network (for example see [26]), which in graph theory translates in highly connected subgraphs. A common approach is to look for cliques, that is graphs whose vertices are all pairwise connected. Several combinatorial problems based on clique have been considered, notable examples being the Maximum Clique problem ([12, GT19]), the Minimum Clique Cover problem ([12, GT17]), and the Minimum Clique Partition problem ([12, GT15]). This last is a classical problem in theoretical computer science, that, given a graph, asks for a partition of the vertices into the minimum number of cliques. The Minimum Clique Partition problem has been deeply studied since the seminal paper of Karp [17], studying its complexity in several graph classes, like cubic graphs [5], unit-disk graphs [6, 24, 10] and bounded clique-width graphs [11].

When analyzing networks, asking for a complete subgraph is sometimes too restrictive, as interesting highly connected graphs do not always have connections between all pairs of vertices, for example due to noise in the data considered.

To overcome this limitation of the clique approach, alternative definitions of highly connected graphs have been proposed, leading to the concept of *relaxed clique* [18]. A relaxed clique is a graph $G = (V, E)$ whose vertices satisfy a property which is a relaxation of the clique property. Indeed, a clique is a subgraph whose vertices are all at distance one from each other, that is the diameter of the graph is one. Moreover, the vertices of a clique have the same degree (the size of the vertices in the clique minus one). Different definitions of relaxed clique are obtained by modifying one of the properties of clique. Some variants relax the distance between the vertices of the subgraph sought, thus leading to distance-based relaxed cliques, other variants relax the degree of the subgraph sought, leading to degree-based relaxed cliques, and so on (see [18] for a survey on different definitions of relaxed clique and their algorithmic properties).

In this paper, we focus on a distance-based relaxation. In a clique all the vertices are required to be at distance at most one from each other. Here this constraint is relaxed, so that the vertices have to be at distance at most s, for an integer $s \geqslant 1$. A subgraph whose vertices are all distance at most s is called an s-*club* (notice that, when $s = 1$, an s-club is exactly a clique). The identification of s-clubs inside a network has been defined for the analysis of networks [21, 1] and has been recently applied for the analysis of social networks [20, 22, 27], and biological networks [23, 3]. Interesting recent studies have shown the relevance of finding s-clubs in a network [20, 22], in particular focusing on finding 2-clubs in real networks like DBLP or a European corporate network.

Contributions to the study of s-clubs mainly focus on the Maximum s-Club problem, that is the problem of finding an s-club of maximum size. Maximum s-Club is known to be NP-hard, for each $s \geqslant 1$ [4]. Even deciding whether there exists an s-club larger than a given size in a graph of diameter $s+1$ is NP-complete, for each $s \geqslant 1$ [3]. The Maximum s-Club problem has been studied also in the approximability and parameterized complexity framework. A polynomial-time approximation algorithm with factor $O(|V|^{1/2})$ for every $s \geqslant 2$ on an input graph $G = (V, E)$ has been designed [2]. This is optimal, since the problem is not approximable within factor $O(|V|^{1/2-\varepsilon})$, on an input graph $G = (V, E)$, for each $\varepsilon > 0$ and $s \geqslant 2$ [2].

Maximum s-Club has been studied also in parameterized complexity framework. Maximum s-Club, unlike the problem of finding a clique of maximum size, is known to be fixed-parameter tractable, when parameterized by the size of an s-club [25, 19, 7]. The Maximum s-Club problem has been investigated also for structural parameters and specific graph classes [15, 14].

In this paper, we consider a different combinatorial problem, where we aim at covering the vertices of a network with a set of subgraphs. Similar to Minimum Clique Partition, we consider the problem of covering a graph with the minimum number of s-clubs such that each vertex belongs to an s-club. We denote this problem by Min s-Club Cover, and we focus in particular on the cases $s = 2$ and $s = 3$. We show some analogies and differences between Min s-Club Cover and Minimum Clique Partition. We start in Section 3 by considering the computational complexity of the problem of covering a graph with two or three s-clubs. This is motivated by the fact that Clique Partition is known to be in P when we ask whether there exists a partition of the graph consisting of two cliques, while it is NP-hard to decide whether there exists a partition of the graph consisting of three cliques [13], since Clique Partition is equivalent to GraphColoring on the complementary graph. As for Clique Partition, we show that it is NP-complete to decide whether there exist three 2-clubs that cover a graph. On the other hand, we show that, unlike Clique Partition, it is NP-complete to decide whether there exist two 3-clubs that cover a graph. These two results imply also that Min 2-Club Cover and Min 3-Club Cover do not belong to the class XP for the parameter "number of clubs" in a cover. Notice that when we ask for the existence of a single s-club that covers a graph, we have to simply check in polynomial-time if the given graph is an s-club.

Then, we consider the approximation complexity of Min 2-Club Cover and Min 3-Club Cover. We recall that, given an input graph $G = (V, E)$, Minimum Clique Partition is not approximable within factor $O(|V|^{1-\varepsilon})$, for any $\varepsilon > 0$, unless $P = NP$ [28]. Here we show that Min 2-Club Cover has a slightly different behavior, while Min 3-Club Cover is similar to Clique Partition. Indeed, in Section 4 we prove that Min 2-Club Cover is not approximable within factor $O(|V|^{1/2-\varepsilon})$, for any $\varepsilon > 0$, unless $P = NP$, while Min 3-Club Cover is not approximable within factor $O(|V|^{1-\varepsilon})$, for any $\varepsilon > 0$, unless $P = NP$. In Section 5, we present a greedy approximation algorithm that has factor $2|V|^{1/2} \log^{3/2} |V|$ for Min 2-Club Cover, which almost match the inapproximability result for the problem.

We start the paper by giving in Section 2 some definitions and by formally defining the problem we are interested in.

2 Preliminaries

Given a graph $G = (V, E)$ and a subset $V' \subseteq V$, we denote by $G[V']$ the subgraph of G induced by V'. Given two vertices $u, v \in V$, the distance between u and v in G, denoted by $d_G(u, v)$, is the length of a shortest path from u to v in G.

The diameter of a graph $G = (V, E)$ is the maximum distance between two vertices of V. Given a graph $G = (V, E)$ and a vertex $v \in V$, we denote by $N_G(v)$ the set of neighbors of v, that is $N_G(v) = \{u : \{v, u\} \in E\}$. We denote by $N_G[v]$ the close neighborhood of V, that is $N_G[v] = N_G(v) \cup \{v\}$. Define $N^l_G(v) = \{u :\ u \text{ has distance at most } l \text{ from } v\}$, with $1 \leqslant l \leqslant 2$. Given a set of vertices $X \subseteq V$ and l, with $1 \leqslant l \leqslant 2$, define $N^l_G(X) = \bigcup_{u \in X} N^l_G(u)$. We may omit the subscript G when it is clear from the context. Now, we give the definition of s-club, which is fundamental for the paper.

Definition 1 *Given a graph $G = (V, E)$, and a subset $V' \subseteq V$, $G[V']$ is an s-club if it has diameter at most s.*

Notice that an s-club must be a connected graph. We present now the formal definition of the Minimum s-Club Cover problem we are interested in.

Problem 1 Minimum s-Club Cover *(*Min s-Club Cover*)*
Input: *a graph $G = (V, E)$ and an integer $s \geqslant 2$.*
Output: *a minimum cardinality collection $\mathcal{S} = \{V_1, \dots, V_h\}$ such that, for each i with $1 \leqslant i \leqslant h$, $V_i \subseteq V$, $G[V_i]$ is an s-club, and, for each vertex $v \in V$, there exists a set V_j, with $1 \leqslant j \leqslant h$, such that $v \in V_j$.*

We denote by s-Club Cover(h), with $1 \leqslant h \leqslant |V|$, the decision version of Min s-Club Cover that asks whether there exists a cover of G consisting of at most h s-clubs.

Notice that in Minimum Clique Partition we can assume that the cliques that cover a graph $G = (V, E)$ partition V, hence the cliques are vertex disjoint. Indeed, it can be shown that there exist h cliques that cover a graph if and only if there exist h cliques that partition the vertices of a graph. Obviously, h cliques that partition the vertices of a graph cover also the graph. On the other hard, if there exists h cliques that cover the vertices of a graph, we can compute h cliques that partition the graph: if two cliques shares a vertex, we can remove it from one of the cliques.

We cannot make the assumption that covering and partitioning a graph is essentially the same problem for s-clubs. Indeed, in a solution of Min s-Club Cover, a vertex may be covered by more than one s-club, in order to have a cover consisting of the minimum number of s-clubs. Consider the example of Fig. 1. The two 2-clubs induced by $\{v_1, v_2, v_3, v_4, v_5\}$ and $\{v_1, v_6, v_7, v_8, v_9\}$ cover G, and both these 2-clubs contain vertex v_1. However, if we ask for a partition of G, we need at least three 2-clubs (for example the 2-clubs induced by $\{v_1, v_2, v_3, v_4, v_5\}$, $\{v_6, v_7\}$ and $\{v_8, v_9\}$). This difference between Minimum Clique Partition and Min s-Club Cover is due to the fact that, while being a clique is a hereditary

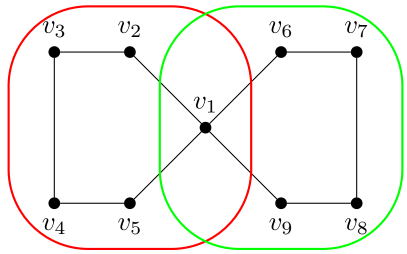

Figure 1: A graph and a cover consisting of two 2-clubs (induced by the vertices in the ovals). Notice that the 2-clubs of this cover must both contain vertex v_1. If v_1 is contained only in one 2-club, for example in the 2-club induced by $\{v_1, v_2, v_3, v_4, v_5\}$, then two 2-clubs are needed to cover $\{v_6, v_7, v_8, v_9\}$, since the subgraph induced by $\{v_6, v_7, v_8, v_9\}$ is not a 2-club (v_6 and v_9 have distance 3 in this subgraph).

property, this is not the case for being an s-club. If a graph G is an s-club, then a subgraph of G may not be an s-club (for example a star is a 2-club, but the subgraph obtained by removing its center is not anymore a 2-club).

A problem related to Min s-Club Cover, is that of partitioning a graph $G = (V, E)$ into the minimum number of s-clubs, denoted by Min s-Club Partition. Notice that, unlike the case of cliques, while a solution of Min s-Club Partition is also a solution of Min s-Club Cover, the opposite is not true. For example, the cover of Fig. 1 consisting of two 2-clubs is not a solution of Min s-Club Partition. Moreover, an optimal solution of Min s-Club Partition on the example of Fig. 1 consists of three 2-clubs. An optimal solution of Min s-Club Cover on a graph G contains at most the same number of s-clubs of an optimal solution of Min s-Club Partition on G.

3 Computational Complexity

In this section we investigate the computational complexity of 2-Club Cover and 3-Club Cover. We show that 2-Club Cover(3), that is deciding whether there exists a cover of a graph G with three 2-clubs, and 3-Club Cover(2), that is deciding whether there exists a cover of a graph G with two 3-clubs, are NP-complete.

3.1 2-Club Cover(3) is NP-complete

In this section we show that 2-Club Cover(3) is NP-complete by giving a reduction from the Clique Partition(3) problem, that is the problem of computing whether there exists a partition of a graph $G^p = (V^p, E^p)$ in three cliques. Consider an instance $G^p = (V^p, E^p)$ of Clique Partition(3), we construct an instance $G = (V, E)$ of 2-Club Cover(3) (see Fig. 2). The vertex set V is defined as follows:

$$V = \{w_i : v_i \in V^p\} \cup \{w_{i,j} : \{v_i, v_j\} \in E^p \wedge i < j\}\}$$

The set E of edges is defined as follows:

$$E = \{\{w_i, w_{i,j}\}, \{w_i, w_{h,i}\} : v_i \in V^p, w_i, w_{i,j}, w_{h,i} \in V\} \cup$$
$$\{\{w_{i,j}, w_{i,l}\}, \{w_{i,j}, w_{h,i}\}, \{w_{h,i}, w_{z,i}\} : w_{i,j}, w_{i,l}, w_{h,i}, w_{z,i} \in V\}$$

Before giving the main result of this section, we prove a property of G.

Lemma 1 *Let $G^p = (V^p, E^p)$ be an instance of* Clique Partition(3) *and let $G = (V, E)$ be the corresponding instance of* 2-Club Cover(3)*. Then, given two vertices $v_i, v_j \in V^p$ and the corresponding vertices $w_i, w_j \in V$:*

- *if $\{v_i, v_j\} \in E^p$, then $d_G(w_i, w_j) = 2$*
- *if $\{v_i, v_j\} \notin E^p$, then $d_G(w_i, w_j) \geqslant 3$*

Proof: Notice that $N_G(w_i) = \{w_{i,z} : \{v_i, v_z\} \in E^p \wedge i < z\} \cup \{w_{h,i} : \{v_i, v_h\} \in E^p \wedge h < i\}$. It follows that $w_j \in N_G^2(w_i)$ if and only if there exists a vertex $w_{i,j}$ (or $w_{j,i}$), which is adjacent to both w_i and w_j. But then, by construction, $w_j \in N_G^2(w_i)$ if and only if $\{v_i, v_j\} \in E^p$. □

We are now able to prove the main properties of the reduction.

Lemma 2 *Let $G^p = (V^p, E^p)$ be a graph input of* Clique Partition(3) *and let $G = (V, E)$ be the corresponding instance of* 2-Club Cover(3)*. Then, given a solution of* Clique Partition(3) *on $G^p = (V^p, E^p)$, we can compute in polynomial time a solution of* 2-Club Cover(3) *on $G = (V, E)$.*

Proof: Consider a solution of Clique Partition(3) on $G^p = (V^p, E^p)$, and let V_1^p, V_2^p, $V_3^p \subseteq V^p$ be the sets of vertices of G^p that partition V^p. We define a solution of 2-Club Cover(3) on $G = (V, E)$ as follows. For each d, with $1 \leqslant d \leqslant 3$, define

$$V_d = \{w_j \in V : v_j \in V_d^p\} \cup \{w_{i,j} : v_i \in V_d^p\}$$

We show that each $G[V_d]$, with $1 \leqslant d \leqslant 3$, is a 2-club. Consider two vertices $w_i, w_j \in V_d$, with $1 \leqslant i < j \leqslant |V|$. Since they correspond to two vertices $v_i, v_j \in V^p$ that belong to a clique of G^p, it follows that $\{v_i, v_j\} \in E^p$ and $w_{i,j} \in V_d$. Thus $d_{G[V_d]}(w_i, w_j) = 2$. Now, consider the vertices $w_i \in V_d$, with $1 \leqslant i \leqslant |V|$, and $w_{h,z} \in V_d$, with $1 \leqslant h < z \leqslant |V|$. If $i = h$ or $i = z$, assume w.l.o.g. $i = h$, then by construction $d_{G[V_d]}(w_i, w_{i,z}) = 1$. Assume that $i \neq h$ and $i \neq z$ (assume w.l.o.g. that $i < h < z$), since $w_{h,z} \in V_d$, it follows that $w_h \in V_d$. Since $w_i, w_h \in V_d$, it follows that $w_{i,h} \in V_d$. By construction, there exist edges $\{w_{i,h}, w_{h,z}\}$, $\{w_i, w_{i,h}\}$ in E^p, thus implying that $d_{G[V_d]}(w_i, w_{h,z}) = 2$. Finally, consider two vertices $w_{i,j}, w_{h,z} \in V_d$, with $1 \leqslant i < j \leqslant |V|$ and $1 \leqslant h < z \leqslant |V|$. Then, by construction, $w_i \in V_d$ and $w_h \in V_d$. But then, $w_{i,h}$ belongs to V_d, and, by construction, $\{w_{i,j}, w_{i,h}\} \in E$ and $\{w_{h,z}, w_{i,h}\} \in E$. It follows that $d_{G[V_d]}(w_{i,j}, w_{h,z}) = 2$.

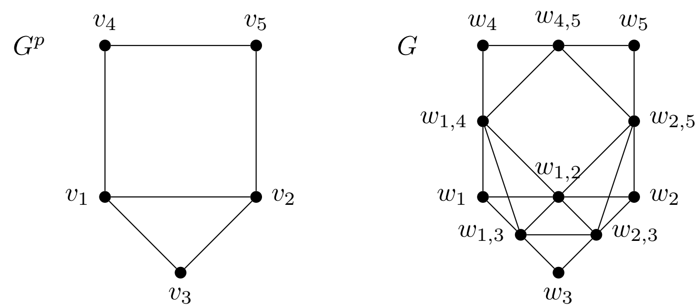

Figure 2: An example of a graph G^p input of Clique Partition(3) and the corresponding graph G input of 2-Club Cover(3).

We conclude the proof observing that, by construction, since V_1^p, V_2^p, V_3^p partition V^p, it holds that $V = V_1 \cup V_2 \cup V_3$, thus $G[V_1]$, $G[V_2]$, $G[V_3]$ covers G. □

Based on Lemma 1, we can prove the following result.

Lemma 3 *Let $G^p = (V^p, E^p)$ be a graph input of* Clique Partition(3) *and let $G = (V, E)$ be the corresponding instance of* 2-Club Cover(3)*. Then, given a solution of* 2-Club Cover(3) *on $G = (V, E)$, we can compute in polynomial time a solution of* Clique Partition(3) *on $G^p = (V^p, E^p)$.*

Proof: Consider a solution of 2-Club Cover(3) on $G = (V, E)$ consisting of three 2-clubs $G[V_1]$, $G[V_2]$, $G[V_3]$. Consider a 2-club $G[V_d]$, with $1 \leqslant d \leqslant 3$. We define three cliques $G^p[V_1^p]$, $G^p[V_2^p]$, $G^p[V_3^p]$ in G^p as follows. For each d, with $1 \leqslant d \leqslant 3$, V_d^p is defined as:

$$V_d^p = \{v_i : w_i \in V_d\}$$

Next, we show that $G[V_d^p]$, with $1 \leqslant d \leqslant 3$, is indeed a clique. By Lemma 1 if $w_i, w_j \in V_d$ then it holds $\{v_i, v_j\} \in E$, thus by construction $\{v_i, v_j\} \in E^p$ and $G[V_d^p]$ is a clique in G^p. Moreover, since $V_1 \cup V_2 \cup V_3 = V$, then $V_1^p \cup V_2^p \cup V_3^p = V^p$. Notice that V_1^p, V_2^p, V_3^p may not be disjoint, but, starting from (V_1^p, V_2^p, V_3^p), it is easy to compute in polynomial time a partition of G^p in three cliques (since being a clique is a hereditary property). □

We are now able to prove the main result of this section.

Theorem 1 2-Club Cover(3) *is NP-complete.*

Proof: By Lemma 2 and Lemma 3 and from the NP-hardness of Clique Partition(3) [17], it follows that 2-Club Cover(3) is NP-hard. The membership to NP follows easily from the fact that, given three 2-clubs of G, it can be checked in polynomial time whether they are indeed 2-clubs and whether they cover all vertices of G. □

3.2 3-Club Cover(2) is NP-complete

In this section we show that 3-Club Cover(2) is NP-complete by giving a reduction from a variant of Sat called 5-Opposite-Sat. Recall that a literal is positive if it is a non-negated variable, while it is negative if it is a negated variable.

Problem 2 5-Opposite-Satisfiability *(*5-Opposite-Sat*)*
Input: *a collection of clauses $\mathcal{C} = \{C_1, \dots, C_p\}$ over the set of variables $X = \{x_1, \dots, x_q\}$, where each $C_i \in \mathcal{C}$, with $1 \leqslant i \leqslant p$, contains exactly five literals and does not contain both a variable and its negation.*
Output: *a truth assignment f to the variables in X such that each clause C_i, with $1 \leqslant i \leqslant p$, contains a positive and a negative literal satisfied by f.*

A clause C_i is opposite-satisfied by a truth assignment f to the variables X if there exist a positive literal and a negative literal in C_i that are both satisfied by f. Notice that we assume that there exist at least one positive literal and at least one negative literal in each clause C_i, with $1 \leqslant i \leqslant p$, otherwise C_i cannot be opposite-satisfied. Moreover, we assume that each variable in an instance of 5-Opposite-Sat appears both as a positive literal and a negative literal in the instance. Notice that if this is not the case, for example a variable appears only as a positive literal, we can assign a true value to the variable, as defining an assignment to false does not contribute to opposite-satisfy any clause. First, we show that 5-Opposite-Sat is NP-complete, which may be of independent interest.

Theorem 2 5-Opposite-Sat *is NP-complete.*

Proof: We reduce from 3-Sat, where given a set X_3 of variables and a set $\mathcal{C}_3$ of clauses, which are a disjunction of 3 literals (a variable or the negation of a variable), we want to find an assignment to the variables such that all clauses are satisfied. Moreover, we assume that each clause in $\mathcal{C}_3$ does not contain a positive variable x and its negation $\overline{x}$, since such a clause is obviously satisfied by any assignment. The same property holds also for the instance of 5-Opposite-Sat we construct.

Consider an instance $(X_3, \mathcal{C}_3)$ of 3-Sat, we construct an instance $(X, \mathcal{C})$ of 5-Opposite-Sat as follows. Define $X = X_3 \cup X_N$, where $X_3 \cap X_N = \emptyset$ and X_N is defined as follows:

$$X_N = \{x_{C,i,1}, x_{C,i,2} : C_i \in \mathcal{C}_3\}$$

Consider $C_i \in \mathcal{C}_3 = (l_{i,1} \vee l_{i,2} \vee l_{i,3})$, where $l_{i,p}$, with $1 \leqslant p \leqslant 3$ is a literal, that is a variable (a positive literal) or a negated variable (a negative literal), we define two clauses $C_{i,1}$ and $C_{i,2}$ as follows:

- $C_{i,1} = l_{i,1} \vee l_{i,2} \vee l_{i,3} \vee x_{C,i,1} \vee \overline{x_{C,i,2}}$
- $C_{i,2} = l_{i,1} \vee l_{i,2} \vee l_{i,3} \vee \overline{x_{C,i,1}} \vee x_{C,i,2}$

The set $\mathcal{C}$ of clauses is defined as follows:

$$\mathcal{C} = \{C_{i,1}, C_{i,2} : C_i \in \mathcal{C}_3\}$$

We claim that $(X_3, \mathcal{C}_3)$ is satisfiable if and only if $(X, \mathcal{C})$ is opposite-satisfiable.

Assume that $(X_3, \mathcal{C}_3)$ is satisfiable and let f be an assignment to the variables on X_3 that satisfies $\mathcal{C}_3$. Consider a clause C_i in $\mathcal{C}_3$, with $1 \leqslant i \leqslant |\mathcal{C}_3|$. Since it is satisfied by f, it follows that there exists a literal $l_{i,p}$ of C_i, with $1 \leqslant p \leqslant 3$, that is satisfied by f. Define an assignment f' on X that is identical to f on X_3 and, if $l_{i,p}$ is positive, then assigns value false to both $x_{C,i,1}$ and $x_{C,i,2}$, if $l_{i,p}$ is negative, then assigns value true to both $x_{C,i,1}$ and $x_{C,i,2}$. It follows that both $C_{i,1}$ and $C_{i,2}$ are opposite-satisfied by f'.

Assume that $(X, \mathcal{C})$ is opposite-satisfied by an assignment f'. Consider two clauses $C_{i,1}$ and $C_{i,2}$, with $1 \leqslant i \leqslant |\mathcal{C}|$, that are opposite-satisfied by f', we claim that there exists at least one literal of $C_{i,1}$ and $C_{i,2}$ not in X_N which is satisfied. Assume this is not the case, then, if $C_{i,1}$ is opposite-satisfied, it follows that $x_{C,i,1}$ is true and $x_{C,i,2}$ is false, thus implying that $C_{i,2}$ is not opposite-satisfied. Then, an assignment f that is identical to f' restricted to X_3 satisfies each clause in $\mathcal{C}$.

Now, since 3-Sat is NP-complete [17], it follows that 5-Opposite-Sat is NP-hard. The membership to NP follows from the observation that, given an assignment to the variables on X, we can check in polynomial-time whether each clause in $\mathcal{C}$ is opposite-satisfied or not. $\square$

Let us now give the construction of the reduction from 5-Opposite-Sat to 3-Club Cover(2). Consider an instance of 5-Opposite-Sat consisting of a set $\mathcal{C}$ of clauses $C_1, \dots, C_p$ over set $X = \{x_1, \dots, x_q\}$ of variables. We assume that it is not possible to opposite-satisfy all the clauses by setting at most two variables to true or to false (this can be easily checked in polynomial-time).

Before giving the details, we present an overview of the reduction. Given an instance $(X, \mathcal{C})$ of 5-Opposite-Sat, for each positive literal x_i, with $1 \leqslant i \leqslant q$, we define vertices $x^T_{i,1}$, $x^T_{i,2}$ and for each negative literal $\overline{x_i}$, with $1 \leqslant i \leqslant q$, we define a vertex x^F_i. Moreover, for each clause $C_j \in \mathcal{C}$, with $1 \leqslant j \leqslant p$, we define a vertex $v_{C,j}$. We define other vertices to ensure that some vertices have distance not greater than three and to force the membership to one of the two 3-clubs of the solution (see Lemma 4). The construction implies that for each i with $1 \leqslant i \leqslant q$, $x^T_{i,1}$ and x^F_i belong to different 3-clubs (see Lemma 5); this corresponds to a truth assignment to the variables in X. Then, we are able to show that each vertex $v_{C,j}$ belongs to the same 3-club of a vertex $x^T_{i,1}$, with $1 \leqslant i \leqslant q$, and of a vertex x^F_h, with $1 \leqslant h \leqslant q$, adjacent to $v_{C,j}$ (see Lemma 7); these vertices correspond to a positive literal x_i and a negative literal $\overline{x_h}$, respectively, that are satisfied by a truth assignment, hence C_j is opposite-satisfied.

Now, we give the details of the reduction. Let $(X, \mathcal{C})$ be an instance of 5-Opposite-Sat, we construct an instance $G = (V, E)$ of 3-Club Cover(2) as follows (see Fig. 3). The vertex set V is defined as follows:

$$V = \{r, r', r_T, r'_T, r^*_T, r_F, r'_F\} \cup \{x^T_{i,1}, x^T_{i,2}, x^F_i : x_i \in X\} \cup \{v_{C,j} : C_j \in \mathcal{C}\} \cup \{y_1, y_2, y\}$$

The edge set E is defined as follows:

$$E = \{\{r, r'\}, \{\{r', r_T\}, \{r', r^*_T\}\{r', r_F\}\} \cup \{\{r_T, x^T_{i,1}\} : x_i \in X\}$$
$$\cup \{\{r_F, x^F_i\} : x_i \in X\} \cup \{\{r'_T, x^T_{i,1}\} : x_i \in X\} \cup \{\{r'_F, x^F_i\} : x_i \in X\} \cup$$
$$\{\{x^T_{i,1}, x^T_{i,2}\} : x_i \in X\} \cup \{\{r^*_T, x^T_{i,2}\}, \{y_1, x^T_{i,2}\} : x_i \in X\} \cup$$
$$\{\{x^T_{i,2}, x^F_j\} : x_i, x_j \in X, i \neq j\} \cup \{\{x^T_{i,1}, v_{C,j}\} : x_i \in C_j\} \cup \{\{x^F_i, v_{C,j}\} : \overline{x_i} \in C_j\} \cup$$
$$\{\{v_{C,j}, y\} : C_j \in \mathcal{C}\} \cup \{\{y, y_2\}, \{y_1, y_2\}, \{y_1, r'_T\}, \{y_1, r'_F\}\}$$

We start by proving some properties of the graph G.

Lemma 4 *Consider an instance* $(\mathcal{C}, X)$ *of* 5-Opposite-Sat *and let* $G = (V, E)$ *be the corresponding instance of* 3-Club Cover(2)*. Then, (1)* $d_G(r', y) > 3$*, (2)* $d_G(r, y) > 3$*, (3)* $d_G(r, v_{C,j}) > 3$*, for each* j *with* $1 \leqslant j \leqslant p$*, and (4)* $d_G(r, r'_F) > 3$*,* $d_G(r, r'_T) > 3$*.*

Proof: We start by proving (1). Notice that any path from r' to y must pass through r_T, r^*_T or r_F. Each of r_T, r^*_T or r_F is adjacent to vertices $x^T_{i,1}$, $x^T_{i,2}$ and x^F_i, with $1 \leqslant i \leqslant q$ (in addition to r'), and none of these vertices is adjacent to y, thus concluding that $d_G(r', y) > 3$. Moreover, observe that for each vertex $v_{C,j}$, with $1 \leqslant j \leqslant p$, there exists a vertex $x^T_{i,1}$, with $1 \leqslant i \leqslant q$, or x^F_h, with $1 \leqslant h \leqslant q$, that is adjacent to $v_{C,j}$, with $1 \leqslant j \leqslant p$, thus $d_G(r', v_{C_j}) = 3$, for each j with $1 \leqslant j \leqslant p$. As a consequence of (1), it follows that (2) holds, that is $d_G(r, y) > 3$. Since $d_G(r', v_{C_j}) = 3$, for each j with $1 \leqslant j \leqslant p$, it holds (3) $d_G(r, v_{C,j}) > 3$.

Finally, we prove (4). Notice that $N^2_G(r) = \{r', r^*_T, r_T, r_F\}$ and that none of the vertices in $N^2_G(r)$ is adjacent to r'_F and r'_T, thus $d_G(r, r'_F) > 3$. □

Consider two sets $V_1 \subseteq V$ and $V_2 \subseteq V$, such that $G[V_1]$ and $G[V_2]$ are two 3-clubs of G that cover G. As a consequence of Lemma 4, it follows that r and r' are in exactly one of $G[V_1]$, $G[V_2]$, w.l.o.g. $G[V_1]$, while r'_T, r'_F, y and $v_{C,j}$, for each j with $1 \leqslant j \leqslant p$, belong to $G[V_2]$ and not to $G[V_1]$.

Next, we show a crucial property of the graph G built by the reduction.

Lemma 5 *Given an instance* $(\mathcal{C}, X)$ *of* 5-Opposite-Sat*, let* $G = (V, E)$ *be the corresponding instance of* 3-Club Cover(2)*. Then, for each* i *with* $1 \leqslant i \leqslant q$*,* $d_G(x^T_{i,1}, x^F_i) > 3$*.*

Proof: Consider a path π of minimum length that connects $x^T_{i,1}$ and x^F_i, with $1 \leqslant i \leqslant q$. First, notice that, by construction, the path π after $x^T_{i,1}$ must pass through one of these vertices: r_T, r'_T, $x^T_{i,2}$ or $v_{C,j}$, with $1 \leqslant j \leqslant p$.

We consider the first case, that is the path π after $x^T_{i,1}$ passes through r_T. Now, the next vertex in π is either r' or $x^T_{h,1}$, with $1 \leqslant h \leqslant q$. Since both r' and $x^T_{h,1}$ are not adjacent to x^F_i, it follows that in this case the path π has length greater than three.

We consider the second case, that is the path π after $x^T_{i,1}$ passes through r'_T. Now, after r'_T, π passes through either y_1 or $x^T_{h,1}$, with $1 \leqslant h \leqslant q$. Since both

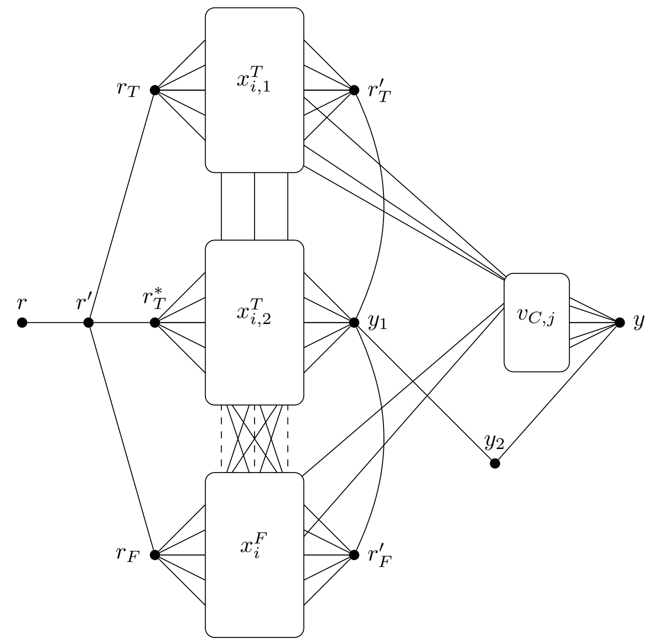

Figure 3: Schematic construction for the reduction from 5-Opposite-Sat to 3-Club Cover(2).

y_1 and $x^T_{h,1}$ are not adjacent to x^F_i, it follows that in this case the path π has length greater than three.

We consider the third case, that is the path after $x^T_{i,1}$ passes through $x^T_{i,2}$. Now, the next vertex of π is either r^*_T or y_1 or x^F_h, with $1 \leqslant h \leqslant q$ and $h \neq i$. Since r^*_T, y_1 and x^F_h are not adjacent to x^F_i, it follows that in this case the path π has length greater than three.

We consider the last case, that is the path after $x^T_{i,1}$ passes through $v_{C,j}$, with $1 \leqslant j \leqslant p$. We have assumed that x_i and $\overline{x_i}$ do not belong to the same clause, thus by construction x^F_i is not incident in $v_{C,j}$. It follows that after $v_{C,j}$, the path π must pass through either y or $x^T_{h,1}$, with $1 \leqslant h \leqslant q$, or x^F_z, $1 \leqslant z \leqslant q$ and $z \neq i$. Once again, since y, $x^T_{h,1}$ and x^F_z are not adjacent to x^F_i, it follows that also in this case the path π has length greater than three, thus concluding the proof. □

Now, we are able to prove the main results of this section.

Lemma 6 *Given an instance $(\mathcal{C}, X)$ of* 5-Opposite-Sat*, let $G = (V, E)$ be the corresponding instance of* 3-Club Cover(2)*. Then, given a truth assignment that opposite-satisfies $\mathcal{C}$, we can compute in polynomial-time two 3-clubs that cover G.*

Proof: Consider a truth assignment f on the set X of variables that opposite-satisfies $\mathcal{C}$. In the following we construct two 3-clubs $G[V_1]$ and $G[V_2]$ that cover

G. The two sets V_1, V_2 are defined as follows:

$$V_1 = \{r, r', r_T, r^*_T, r_F\} \cup \{x^T_{i,1}, x^T_{i,2} : f(x_i) = false\} \cup \{x^F_i, : f(x_i) = true\}$$

$$V_2 = \{r'_T, r'_F, y, y_1, y_2\} \cup \{x^T_{i,1}, x^T_{i,2} : f(x_i) = true\} \cup \{x^F_i : f(x_i) = false\} \cup$$
$$\{v_{C,j} : 1 \leqslant j \leqslant p\}$$

Next, we show that $G[V_1]$ and $G[V_2]$ are indeed two 3-clubs that cover G. First, notice that $V_1 \cup V_2 = V$, hence $G[V_1]$ and $G[V_2]$ cover G. Next, we show that both $G[V_1]$ and $G[V_2]$ are indeed 3-clubs.

Let us first consider $G[V_1]$. By construction, $d_{G[V_1]}(r, x^T_{i,1}) = 3$ and $d_{G[V_1]}(r, x^T_{i,2}) = 3$, for each i with $1 \leqslant i \leqslant i \leqslant q$, and $d_{G[V_1]}(r, x^F_i) = 3$, for each i with $1 \leqslant i \leqslant i \leqslant q$. Moreover, $d_{G[V_1]}(r', x^T_{i,1}) = 2$ and $d_{G[V_1]}(r', x^T_{i,2}) = 2$, for each i with $1 \leqslant i \leqslant q$, and $d_{G[V_1]}(r', x^F_i) = 2$, for each i with $1 \leqslant i \leqslant i \leqslant q$. As a consequence, it holds that r_T, r'_T and r_F have distance at most three in $G[V_1]$ from each vertex $x^T_{i,1}$, from each vertex $x^T_{i,2}$, and from each vertex x^F_i. Since r, r_T, r^*_T and r_F are in $N(r')$, it follows that r, r', r_T, r^*_T and r_F are at distance at most 2 in $G[V_1]$. Hence, we focus on vertices $x^T_{i,1}$, with $1 \leqslant i \leqslant q$, $x^T_{h,2}$, with $1 \leqslant h \leqslant q$ and x^F_j, with $1 \leqslant j \leqslant q$. Since there exists a path that passes trough $x^T_{i,1}$, r_T, $x^T_{h,1}$ and $x^T_{h,2}$, vertices $x^T_{i,1}$, $x^T_{h,1}$ are at distance at most two in $G[V_1]$, while $x^T_{i,1}$, $x^T_{h,2}$ are at distance at most three in $G[V_1]$ (if $i = h$ they are at distance one). Vertices $x^T_{h,2}$ and x^F_j are at distance one in $G[V_1]$, since by construction $h \neq j$ and $\{x^T_{h,2}, x^F_j\} \in E$. Finally, $x^T_{i,1}$ and x^F_j are at distance two in $G[V_1]$, since there exists a path that passes trough $x^T_{i,1}$, $x^T_{i,2}$ and x^F_j in $G[V_1]$, as $i \neq j$. It follows that $G[V_1]$ is a 3-club.

We now consider $G[V_2]$. We recall that, for each i with $1 \leqslant i \leqslant q$, if $x^T_{i,1}$, $x^T_{i,2} \in V_2$, then $x^F_i \in V_1$. Furthermore, we recall that we assume that each x_i appears as a positive and a negative literal in the instance of 5-Opposite-Sat, thus each vertex $x^T_{i,1}$, with $1 \leqslant i \leqslant q$, and each vertex x^F_h, with $1 \leqslant h \leqslant q$, are connected to some $V_{C,j}$, with $1 \leqslant j \leqslant p$.

First, notice that vertex y is at distance at most three in $G[V_2]$ from each vertex of V_2, since it has distance one in $G[V_2]$ from each vertex $v_{C,j}$, with $1 \leqslant j \leqslant p$, thus distance two from $x^T_{i,1}$, with $1 \leqslant i \leqslant q$, and x^F_h, with $1 \leqslant h \leqslant q$, and three from $x^T_{i,2}$, with $1 \leqslant i \leqslant q$, r'_T and r'_F. Since y is adjacent to y_2, it has distance one from y_2 and two from y_1.

Now, consider a vertex $v_{C,j}$, with $1 \leqslant j \leqslant p$. Since f opposite-satisfies $\mathcal{C}$, it follows that there exist two vertices in V_2, $x^T_{i,1}$, with $1 \leqslant i \leqslant q$, and x^F_z, with $1 \leqslant z \leqslant q$, which are connected to $v_{C,j}$. It follows that $v_{C,j}$ has distance 2 in $G[V_2]$ from r'_T and from r'_F, and at most 3 from each $x^T_{h,1} \in V_2$, with $1 \leqslant h \leqslant q$, and from each $x^F_z \in V_2$, with $1 \leqslant z \leqslant q$. Furthermore, notice that, since $v_{C,j}$ is adjacent to x^F_z and x^F_z is adjacent to each $x^T_{h,2} \in V_2$, with $1 \leqslant h \leqslant q$ and $h \neq z$, then $v_{C,j}$ has distance at most two in $G[V_2]$ from each $x^T_{h,2} \in V_2$. Finally, since

$v_{C,j}$ is adjacent to y, it has distance two and three respectively, from y_2 and y_1, in $G[V_2]$.

Consider a vertex $x^T_{i,1} \in V_2$, with $1 \leqslant i \leqslant q$. We have already shown that it has distance at most three in $G[V_2]$ from any $v_{C,j}$, with $1 \leqslant j \leqslant p$, and two from y. Since $x^T_{i,1}$ is adjacent to r'_T, it has distance at most two from each other vertex $x^T_{h,1}$, with $1 \leqslant h \leqslant q$, and three from each other vertex $x^T_{h,2}$ of $G[V_2]$. Moreover, it has distance two from y_1 and three from y_2 and r'_F. Since $x^T_{i,2}$ is adjacent to every vertex $x^F_z \in V_2$, with $1 \leqslant z \leqslant q$, as $z \neq i$, it follows that $x^T_{h,1}$ has distance at most two from every vertex $x^F_z \in V_2$.

Consider a vertex $x^T_{i,2} \in V_2$, with $1 \leqslant i \leqslant q$. We have already shown that it has distance at most two from each $v_{C,j}$ in $G[V_2]$. Since it is connected to $x^T_{i,1}$, it has distance three from y and two from r'_T in $G[V_2]$. By construction $x^T_{i,2}$ is adjacent to every vertex $x^F_z \in V_2$, with $1 \leqslant z \leqslant q$, hence $x^T_{i,2}$ has distance at most two from r'_F in $G[V_2]$. Moreover, $x^T_{i,2}$ has distance two from each vertex $x^T_{h,2}$ in $G[V_2]$, with $1 \leqslant i \leqslant q$, since by construction they are both adjacent to y_1. Since $x^T_{i,2}$ is adjacent to y_1, thus it has distance at most two from y_2 in $G[V_2]$.

Consider a vertex x^F_h, with $1 \leqslant h \leqslant q$. It has distance one from r'_F in $G[V_2]$, and thus distance two from y_1 and three from y_2 in $G[V_2]$. Moreover, x^F_h is adjacent to each $x^T_{i,2} \in V_2$, with $1 \leqslant i \leqslant q$, thus it has distance two from each $x^T_{i,1}$ and distance three from r'_T in $G[V_2]$. Since by construction there exists at least one $v_{C,j}$, with $1 \leqslant j \leqslant p$, adjacent to x^F_h, thus x^F_h has distance two from y and three from each $v_{C,z}$ in $G[V_2]$.

Finally, we consider vertices r'_T, r'_F, y_1 and y_2. Notice that it suffices to show that these vertices have pairwise distance at most three in $G[V_2]$, since we have previously shown that any other vertex of V_2 has distance at most three from these vertices in $G[V_2]$. Since $r'_T, r'_F, y_2 \in N(y_1)$, they are all at distance at most two. It follows that $G[V_2]$ is a 3-club, thus concluding the proof. □

Lemma 7 *Given an instance $(\mathcal{C}, X)$ of* 5-Opposite-Sat*, let $G = (V, E)$ be the corresponding instance of* 3-Club Cover(2)*. Then, given two 3-clubs that cover G, we can compute in polynomial time a truth assignment that opposite-satisfies $\mathcal{C}$.*

Proof: Consider two 3-clubs $G[V_1]$, $G[V_2]$, with $V_1, V_2 \subseteq V$, that cover G. First, notice that by Lemma 4 we assume that $r, r' \in V_1 \setminus V_2$, while $y, r'_T, r'_F \in V_2 \setminus V_1$ and $v_{C,j} \in V_2 \setminus V_1$, for each j with $1 \leqslant j \leqslant p$. Moreover, by Lemma 5 it follows that for each i with $1 \leqslant i \leqslant q$, $x^T_{i,1}$ and x^F_i do not belong to the same 3-club, that is exactly one belongs to V_1 and exactly one belongs to V_2.

By construction, each path of length at most three from a vertex $v_{C,j}$, with $1 \leqslant j \leqslant p$, to r'_F must pass through some x^F_h, with $1 \leqslant h \leqslant q$. Similarly, each path of length at most three from a vertex $v_{C,j}$, with $1 \leqslant j \leqslant p$, to r'_T must pass through some $x^T_{i,1}$. Assume that $v_{C,j}$, with $1 \leqslant j \leqslant p$, is not adjacent to a vertex $x^T_{i,1} \in V_2$, with $1 \leqslant i \leqslant q$ ($x^F_h \in V_2$, with $1 \leqslant h \leqslant p$ respectively). It follows that $v_{C,j}$ is only adjacent to y and to vertices x^F_w, with $1 \leqslant w \leqslant q$ ($x^T_{u,1}$, with

$1 \leqslant u \leqslant q$, respectively) in $G[V_2]$. In the first case, notice that y is adjacent only to $v_{C,z}$, with $1 \leqslant z \leqslant p$, and y_2, none of which is adjacent to r'_T (r'_F, respectively), thus implying that this path from $v_{C,j}$ to r'_T (to r'_F, respectively) has length at least 4. In the second case, x^F_w ($x^T_{u,1}$, respectively) is adjacent to r'_F, r_F, $v_{C,j}$ and $x^T_{i,2}$ (r'_T, r_T, $v_{C,j}$, $x^T_{u,2}$, respectively), none of which is adjacent to r'_T (r'_F, respectively), implying that also in this case the path from $v_{C,j}$ to r'_T (to r'_F, respectively) has length at least 4. Since $r'_T, r'_F, v_{C,j} \in V_2$, it follows that, for each $v_{C,j}$, the set V_2 contains a vertex $x^T_{i,1}$, with $1 \leqslant i \leqslant q$, and a vertex x^F_h, with $1 \leqslant h \leqslant q$, connected to $v_{C,j}$.

By Lemma 5 exactly one of $x^T_{i,1}$, x^F_i belongs to V_2, thus we can construct a truth assignment f as follows: $f(x_i) :=$ true, if $x^T_{i,1} \in V_2$, $f(x_i) :=$ false, if $x^F_i \in V_2$. The assignment f opposite-satisfies each clause of $\mathcal{C}$, since each $v_{C,j}$ is connected to a vertex $x^T_{i,1}$, for some i with $1 \leqslant i \leqslant q$, and a vertex x^F_h, for some h with $1 \leqslant h \leqslant q$.

□

We can now state the main result of this section.

Theorem 3 3-Club Cover(2) *is NP-complete.*

Proof: By Lemma 6 and Lemma 7, and from the NP-hardness of 5-Opposite-Sat (see Theorem 2), it follows that 3-Club Cover(2) is NP-hard. The membership in NP follows easily from the fact that, given two 3-clubs, it can be checked in polynomial time whether are 3-clubs and cover all vertices of G. □

4 Hardness of Approximation

In this section we consider the approximation complexity of Min 2-Club Cover and Min 3-Club Cover and we prove that Min 2-Club Cover is not approximable within factor $O(|V|^{1/2-\varepsilon})$, for each $\varepsilon > 0$, and that Min 3-Club Cover is not approximable within factor $O(|V|^{1-\varepsilon})$, for each $\varepsilon > 0$, unless P = NP.

4.1 Hardness of Approximation of Min 2-Club Cover

The proof for Min 2-Club Cover is obtained with a reduction very similar to that of Section 3.1. We present a preserving-factor reduction from Minimum Clique Partition to Min 2-Club Cover. Let $G^p = (V^p, E^p)$ be a graph input of Minimum Clique Partition, we build in polynomial time a corresponding instance $G = (V, E)$ of Min 2-Club Cover as in Section 3.1. In what follows we prove the following results that are useful for the reduction.

Lemma 8 *Let $G^p = (V^p, E^p)$ be a graph input of* Minimum Clique Partition *and let $G = (V, E)$ be the corresponding instance of* Min 2-Club Cover*. Then, given a solution of* Minimum Clique Partition *on $G^p = (V^p, E^p)$ consisting of k cliques, we can compute in polynomial time a solution of* Min 2-Club Cover *on $G = (V, E)$ consisting of k 2-clubs.*

Proof: Consider a solution of Minimum Clique Partition on $G^p = (V^p, E^p)$ where $\{V_1^p, V_2^p, \dots, V_k^p\}$ is the set of k cliques that partition V^P. We define a solution of Min 2-Club Cover on $G = (V, E)$ consisting of k 2-clubs as follows. For each $d, 1 \leqslant d \leqslant k$, let

$$V_d = \{w_j \in V : v_j \in V_d^p\} \cup \{w_{i,j} : v_i \in V_d^p \wedge i < j\}$$

As for the proof of Lemma 2, it follows that for each d, $G[V_d]$ is a 2-club. Furthermore, $G[V_1], \dots, G[V_k]$ cover each vertex of V, as each $v_i \in V^p$ is covered by one of the cliques $V_1^p, V_2^p \dots V_k^p$. □

Lemma 9 *Let $G^p = (V^p, E^p)$ be a graph input of* Minimum Clique Partition *and let $G = (V, E)$ be the corresponding instance of* Min 2-Club Cover*. Then, given a solution of* Min 2-Club Cover *on $G = (V, E)$ consisting of k 2-clubs, we can compute in polynomial time a solution of* Minimum Clique Partition *on $G^p = (V^p, E^p)$ with k cliques.*

Proof: Consider the 2-clubs $G[V_1], \dots, G[V_k]$ that cover G. As for the proof of Lemma 3, the result follows from the fact that by Lemma 1, given $w_i, w_j \in V_d$, for each d with $1 \leqslant d \leqslant k$, it holds that $\{v_i, v_j\} \in E$. As a consequence, we can define a solution of Minimum Clique Partition on $G^p = (V^p, E^p)$ consisting of k cliques as follows, for each $d, 1 \leqslant d \leqslant k$:

$$V_d^p = \{v_i : w_i \in V_d\}$$

□

Theorem 4 *Unless P = NP,* Min 2-Club Cover *is not approximable within factor $O(|V|^{1/2-\varepsilon})$, for each $\varepsilon > 0$.*

Proof:

The inapproximability of Min 2-Club Cover follows from Lemma 8 and Lemma 9, and from the inapproximability of Minimum Clique Partition, which is known to be inapproximable within factor $O(|V^p|^{1-\varepsilon'})$ [28] (where $G^p = (V^p, E^p)$ is an instance of Minimum Clique Partition). Hence Min 2-Club Cover is not approximable within factor $O(|V^p|^{1-\varepsilon'})$, for each $\varepsilon' > 0$, unless P = NP. By the definition of $G = (V, E)$, it holds $|V| = |V^p| + |E^p| \leqslant |V^p|^2$ hence, for each $\varepsilon > 0$, Min 2-Club Cover is not approximable within factor $O(|V|^{1/2-\varepsilon})$, unless P = NP. □

4.2 Hardness of Approximation of Min 3-Club Cover

We show that Min 3-Club Cover is not approximable within factor $O(|V|^{1-\varepsilon})$, for each $\varepsilon > 0$, unless P = NP, by giving a preserving-factor reduction from Minimum Clique Partition.

Consider an instance $G^p = (V^p, E^p)$ of Minimum Clique Partition, we construct an instance $G = (V, E)$ of Min 3-Club Cover by adding a pendant vertex connected to each vertex of V^p. Formally, $V = \{u_i, w_i : v_i \in V^p\}$, $E = \{\{u_i, w_i\} : 1 \leqslant i \leqslant |V^p|\} \cup \{\{u_i, u_j\} : \{v_i, v_j\} \in E^p\}\}$.

We prove now the main properties of the reduction.

Lemma 10 *Let $G^p = (V^p, E^p)$ be an instance of* Minimum Clique Partition *and let $G = (V, E)$ be the corresponding instance of* Min 3-Club Cover. *Then, given a solution of* Minimum Clique Partition *on $G^p = (V^p, E^p)$ consisting of k cliques, we can compute in polynomial time a solution of* Min 3-Club Cover *on $G = (V, E)$ consisting of k 3-clubs.*

Proof: Consider a solution of Minimum Clique Partition on $G^p = (V^p, E^p)$, consisting of the cliques $\{G^p[V_{c,1}], G^p[V_{c,2}], \dots, G^p[V_{c,k}]\}$. Then, for each i, with $1 \leqslant h \leqslant k$, define the following subset $V_h \subseteq V$:

$$V_h = \{u_j, w_j \in V : v_j \in V_h^p\}$$

Since $V_1^p, V_2^p \dots V_k^p$ partition V^p, it follows that $V_1, V_2 \dots V_k$ partition (hence cover) G. Now, we show that each $G[V_h]$, with $1 \leqslant h \leqslant k$, is a 3-club. First, notice that since $G[V_h^p]$, is a clique, then the set $\{u_j : u_j \in V_h\}$ induces a clique in G. Then, it follows that, for each $u_i, w_j, w_z \in V_h$, $d_{G[V_h]}(u_i, w_j) \leqslant 2$ and $d_{G[V_h]}(w_j, w_z) \leqslant 3$, thus concluding the proof. □

Lemma 11 *Let $G^p = (V^p, E^p)$ be a graph input of* Minimum Clique Partition *and let $G = (V, E)$ be the corresponding instance of* Min 3-Club Cover. *Then, given a solution of* Min 3-Club Cover *on $G = (V, E)$ consisting of k 3-clubs, we can compute in polynomial time a solution of* Minimum Clique Partition *on $G^p = (V^p, E^p)$ consisting of k cliques.*

Proof: Consider the k 3-clubs $G[V_1], \dots, G[V_k]$ that cover G. First, we show that for each $V_h, 1 \leqslant h, \leqslant k$, and for each $w_i, w_j \in V_h$, with $1 \leqslant i, j \leqslant |V^p|$, it holds that $u_i, u_j \in V_h$. Indeed, notice that $N(w_i) = \{u_i\}$ and $N(w_j) = \{u_j\}$, and by the definition of a 3-club we must have $d_{G[V_h]}(w_i, w_j) \leqslant 3$, it follows that $u_i, u_j \in V_h$. Hence, we can define a set of cliques of G^p. For each V_h, with $1 \leqslant h \leqslant k$, define a set V_h^p:

$$V_h^p = \{v_i : w_i \in V_h\}$$

Notice that each V_h^p, $1 \leqslant h \leqslant k$, induces a clique in G^p, as by construction if $v_i, v_j \in V_h^p$, then $w_i, w_j \in V_h$, and this implies $\{v_i, v_j\} \in E^p$. Notice that the cliques $V_1^p, \dots, V_k^p$ may overlap, but starting from $V_1^p, \dots, V_k^p$, we can easily compute in polynomial time a clique partition of G^p consisting of at most k cliques. □

Lemma 10 and Lemma 11 imply the following result.

Theorem 5 Min 3-Club Cover *is not approximable within factor $O(|V|^{1-\varepsilon})$, for each $\varepsilon > 0$, unless P = NP.*

5 An Approximation Algorithm for Min 2-Club Cover

In this section, we present an approximation algorithm for Min 2-Club Cover that achieves an approximation factor of $2|V|^{1/2} \log^{3/2} |V|$. Notice that, due to

Theorem 4, the approximation factor is almost tight. We start by describing the approximation algorithm, then we present the analysis of the approximation factor.

Algorithm 1: Club-Cover-Approx

Data: a graph G
Result: a cover $\mathcal{S}$ of G
$V' := V$; /* V' is the set of uncovered vertices of G, initialized to V */
$\mathcal{S} := \emptyset$;
while $V' \neq \emptyset$ **do**
  Let v be a vertex of V such that $|N[v] \cap V'|$ is maximum;
  Add $N[v]$ to $\mathcal{S}$;
  $V' := V' \setminus N[v]$;

Club-Cover-Approx is similar to the textbook greedy approximation algorithm for Minimum Dominating Set and Minimum Set Cover. While there exists an uncovered vertex of G, the Club-Cover-Approx algorithm greedily defines a 2-club induced by the set $N[v]$ of vertices, with $v \in V$, such that $N[v]$ covers the maximum number of uncovered vertices (notice that some of the vertices of $N[v]$ may already be covered). While for Minimum Dominating Set the choice of each iteration is optimal, here the choice is suboptimal. Notice that indeed computing a maximum 2-club is NP-hard.

Clearly the algorithm returns a feasible solution for Min 2-Club Cover, as each set $N[v]$ picked by the algorithm is a 2-club and, by construction, each vertex of V is covered. Next, we show the approximation factor yielded by the Club-Cover-Approx algorithm for Min 2-Club Cover.

First, consider the set V_D of vertices $v \in V$ picked by the Club-Cover-Approx algorithm, so that $N[v]$ is added to $\mathcal{S}$. Notice that $|V_D| = |\mathcal{S}|$ and that V_D is a dominating set of G, since, at each step, the vertex v picked by the algorithm dominates each vertex in $N[v]$, and each vertex in V is covered by the algorithm, so it belongs to some $N[v]$, with $v \in V_D$.

Let D be a minimum dominating set of the input graph G. By the property of the greedy approximation algorithm for Minimum Dominating Set, the set V_D has the following property [16]:

$$|V_D| \leqslant |D| \log |V| \tag{1}$$

The size of a minimum dominating set in graphs of diameter bounded by 2 (hence 2-clubs) has been considered in [8], where the following result is proven.

Lemma 12 ([8]) *Let $H = (V_H, E_H)$ be a 2-club, then H has a dominating set of size at most $1 + \sqrt{|V_H| + \ln(|V_H|)}$.*

The approximation factor $2|V|^{1/2} \log^{3/2} |V|$ for Club-Cover-Approx is obtained by combining Lemma 12 and Equation 1.

Theorem 6 *Let OPT be an optimal solution of* Min 2-Club Cover*, then Club-Cover-Approx returns a solution having at most $2|V|^{1/2} \log^{3/2} |V||OPT|$ 2-clubs.*

Proof: Let D be a minimum dominating set of G and let OPT be an optimal solution of Min 2-Club Cover. We start by proving that $|D| \leqslant 2|OPT||V|^{1/2}\log^{1/2}|V|$. For each 2-club $G[C]$, with $C \subseteq V$, that belongs to OPT, by Lemma 12 there exists a dominating set D_C of size at most $1 + \sqrt{|C| + \ln(|C|)} \leqslant 2\sqrt{|C| + \ln(|C|)}$. Since $|C| \leqslant |V|$, it follows that each 2-club $G[C]$ that belongs to OPT has a dominating set of size at most $2\sqrt{|V| + \ln(|V|)}$. Consider, now, $D' = \bigcup_{C \in OPT} D_C$. It follows that D' is a dominating set of G, since the 2-clubs in OPT covers G. Since D' contains $|OPT|$ sets D_C and $|D_C| \leqslant 2\sqrt{|V| + \ln(|V|)}$, for each $G[C] \in OPT$, it follows that $|D'| \leqslant 2|OPT|\sqrt{|V| + \ln(|V|)}$. Since D is a minimum dominating set, it follows that

$$|D| \leqslant |D'| \leqslant 2|OPT|(\sqrt{|V| + \ln(|V|)}).$$

By Equation 1, it holds $|V_D| \leqslant 2|D|\log|V|$ thus $|V_D| \leqslant 2|V|^{1/2}\ln^{1/2}|V|\log|V||OPT| \leqslant 2|V|^{1/2}\log^{3/2}|V||OPT|$. □

Notice that, starting from a solution $\mathcal{S}$ of Algorithm 1, we can compute in polynomial time an approximated solution for the Min 2-Club Partition problem on input G having factor $2|V|^{1/2}\log^{3/2}|V||OPT'|$, where OPT' is an optimal solution of Min 2-Club Partition on G. Indeed, first observe that $|OPT'| \geqslant |OPT|$, as observed in Section 2. Recall that $\mathcal{S}$ consists of 2-clubs $N[v]$, with $v \in V_D \subseteq V$. Then, starting from $\mathcal{S}$, compute a solution $\mathcal{S}'$ of Min 2-Club Partition by greedily assigning the shared vertices to exactly one 2-club of $\mathcal{S}'$ such that if there exists a 2-club $N[u] \in \mathcal{S}$, then there exists a 2-club $C_u \in \mathcal{S}'$ with $C_u \subseteq N[u]$. Notice that, each vertex $u \in V_D$ is part only of the 2-club $C_u \subseteq N[u]$ and it is not assigned to any other 2-club of $\mathcal{S}'$. $\mathcal{S}'$ is a solution of Min 2-Club Partition on input G and contains as many 2-clubs as $\mathcal{S}$. Thus

$$|\mathcal{S}'| = |\mathcal{S}| \leqslant 2|V|^{1/2}\log^{3/2}|V||OPT| \leqslant 2|V|^{1/2}\log^{3/2}|V||OPT'|.$$

6 Conclusion

There are some interesting directions for the problem of covering a graph with s-clubs. From the computational complexity point of view, the main open problem is whether 2-Club Cover(2) is NP-complete or is in P. Moreover, it would be interesting to study the computational/parameterized complexity of the problem in specific graph classes, as done for Minimum Clique Partition [5, 6, 24, 10]. For example Minimum Clique Partition is polynomial time solvable for graphs of bounded clique-width [11]. Finally, from the approximation complexity point of view, there is a small gap between the inapproximabily result and the approximation factor for Min 2-Club Cover, an open problem is reducing this gap.

Acknowledgements

We thank the anonymous referee for useful comments that improved the paper presentation.

References

[1] R. D. Alba. A graph-theoretic definition of a sociometric clique. *The Journal of Mathematical Sociology*, 3:113–126, 1973.

[2] Y. Asahiro, Y. Doi, E. Miyano, K. Samizo, and H. Shimizu. Optimal Approximation Algorithms for Maximum Distance-Bounded Subgraph Problems. *Algorithmica*, Jul 2017. doi:10.1007/s00453-017-0344-y.

[3] B. Balasundaram, S. Butenko, and S. Trukhanov. Novel Approaches for Analyzing Biological Networks. *J. Comb. Optim.*, 10(1):23–39, 2005. doi:10.1007/s10878-005-1857-x.

[4] J. Bourjolly, G. Laporte, and G. Pesant. An exact algorithm for the maximum k-club problem in an undirected graph. *European Journal of Operational Research*, 138(1):21–28, 2002. doi:10.1016/S0377-2217(01)00133-3.

[5] M. R. Cerioli, L. Faria, T. O. Ferreira, C. A. J. Martinhon, F. Protti, and B. A. Reed. Partition into cliques for cubic graphs: Planar case, complexity and approximation. *Discrete Applied Mathematics*, 156(12):2270–2278, 2008. doi:10.1016/j.dam.2007.10.015.

[6] M. R. Cerioli, L. Faria, T. O. Ferreira, and F. Protti. A note on maximum independent sets and minimum clique partitions in unit disk graphs and penny graphs: complexity and approximation. *RAIRO - Theor. Inf. and Applic.*, 45(3):331–346, 2011. doi:10.1051/ita/2011106.

[7] M. Chang, L. Hung, C. Lin, and P. Su. Finding large k-clubs in undirected graphs. *Computing*, 95(9):739–758, 2013. doi:10.1007/s00607-012-0263-3.

[8] W. J. Desormeaux, T. W. Haynes, M. A. Henning, and A. Yeo. Total domination in graphs with diameter 2. *Journal of Graph Theory*, 75(1):91–103, 2014. doi:10.1002/jgt.21725.

[9] R. Dondi, G. Mauri, F. Sikora, and I. Zoppis. Covering with clubs: Complexity and approximability. In C. S. Iliopoulos, H. W. Leong, and W. Sung, editors, *Combinatorial Algorithms - 29th International Workshop, IWOCA 2018, Singapore, July 16-19, 2018, Proceedings*, volume 10979 of *Lecture Notes in Computer Science*, pages 153–164. Springer, 2018. doi:10.1007/978-3-319-94667-2_13.

[10] A. Dumitrescu and J. Pach. Minimum Clique Partition in Unit Disk Graphs. *Graphs and Combinatorics*, 27(3):399–411, 2011. doi:10.1007/s00373-011-1026-1.

[11] W. Espelage, F. Gurski, and E. Wanke. How to Solve NP-hard Graph Problems on Clique-Width Bounded Graphs in Polynomial Time. In

A. Brandstädt and V. B. Le, editors, *Graph-Theoretic Concepts in Computer Science, 27th International Workshop, WG 2001, Boltenhagen, Germany, June 14-16, 2001, Proceedings*, volume 2204 of *Lecture Notes in Computer Science*, pages 117–128. Springer, 2001. doi:10.1007/3-540-45477-2_12.

[12] M. R. Garey and D. S. Johnson. Computers and Intractability: A Guide to the Theory of NP-Completeness. 1979.

[13] M. R. Garey, D. S. Johnson, and L. J. Stockmeyer. Some Simplified NP-Complete Graph Problems. *Theor. Comput. Sci.*, 1(3):237–267, 1976. doi:10.1016/0304-3975(76)90059-1.

[14] P. A. Golovach, P. Heggernes, D. Kratsch, and A. Rafiey. Finding clubs in graph classes. *Discrete Applied Mathematics*, 174:57–65, 2014. doi:10.1016/j.dam.2014.04.016.

[15] S. Hartung, C. Komusiewicz, and A. Nichterlein. Parameterized Algorithmics and Computational Experiments for Finding 2-Clubs. *J. Graph Algorithms Appl.*, 19(1):155–190, 2015. doi:10.7155/jgaa.00352.

[16] D. S. Johnson. Approximation Algorithms for Combinatorial Problems. *J. Comput. Syst. Sci.*, 9(3):256–278, 1974. doi:10.1016/S0022-0000(74)80044-9.

[17] R. M. Karp. Reducibility among combinatorial problems. In R. E. Miller and J. W. Thatcher, editors, *Proceedings of a symposium on the Complexity of Computer Computations, held March 20-22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York.*, The IBM Research Symposia Series, pages 85–103. Plenum Press, New York, 1972.

[18] C. Komusiewicz. Multivariate Algorithmics for Finding Cohesive Subnetworks. *Algorithms*, 9(1):21, 2016. doi:10.3390/a9010021.

[19] C. Komusiewicz and M. Sorge. An algorithmic framework for fixed-cardinality optimization in sparse graphs applied to dense subgraph problems. *Discrete Applied Mathematics*, 193:145–161, 2015. doi:10.1016/j.dam.2015.04.029.

[20] S. Laan, M. Marx, and R. J. Mokken. Close communities in social networks: boroughs and 2-clubs. *Social Netw. Analys. Mining*, 6(1):20:1–20:16, 2016. doi:10.1007/s13278-016-0326-0.

[21] R. Mokken. Cliques, clubs and clans. *Quality & Quantity: International Journal of Methodology*, 13(2):161–173, 1979.

[22] R. J. Mokken, E. M. Heemskerk, and S. Laan. Close communication and 2-clubs in corporate networks: Europe 2010. *Social Netw. Analys. Mining*, 6(1):40:1–40:19, 2016. doi:10.1007/s13278-016-0345-x.

[23] S. Pasupuleti. Detection of protein complexes in protein interaction networks using n-clubs. In E. Marchiori and J. H. Moore, editors, *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics, 6th European Conference, EvoBIO 2008, Naples, Italy, March 26-28, 2008. Proceedings*, volume 4973 of *Lecture Notes in Computer Science*, pages 153–164. Springer, 2008. doi:10.1007/978-3-540-78757-0_14.

[24] I. A. Pirwani and M. R. Salavatipour. A weakly robust PTAS for minimum clique partition in unit disk graphs. *Algorithmica*, 62(3-4):1050–1072, 2012. doi:10.1007/s00453-011-9503-8.

[25] A. Schäfer, C. Komusiewicz, H. Moser, and R. Niedermeier. Parameterized computational complexity of finding small-diameter subgraphs. *Optimization Letters*, 6(5):883–891, 2012. doi:10.1007/s11590-011-0311-5.

[26] M. Sozio and A. Gionis. The community-search problem and how to plan a successful cocktail party. In B. Rao, B. Krishnapuram, A. Tomkins, and Q. Yang, editors, *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, July 25-28, 2010*, pages 939–948. ACM, 2010. doi:10.1145/1835804.1835923.

[27] I. Zoppis, R. Dondi, E. Santoro, G. Castelnuovo, F. Sicurello, and G. Mauri. Optimizing social interaction - A computational approach to support patient engagement. In R. Zwiggelaar, H. Gamboa, A. L. N. Fred, and S. B. i Badia, editors, *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018) - Volume 5: HEALTHINF, Funchal, Madeira, Portugal, January 19-21, 2018.*, pages 651–657. SciTePress, 2018. doi:10.5220/0006730606510657.

[28] D. Zuckerman. Linear Degree Extractors and the Inapproximability of Max Clique and Chromatic Number. *Theory of Computing*, 3(1):103–128, 2007. doi:10.4086/toc.2007.v003a006.