Algorithmica (2023) 85:992–1028
https://doi.org/10.1007/s00453-022-01062-3

# On the Tractability of Covering a Graph with 2-Clubs

**Riccardo Dondi[1] · Manuel Lafond[2]**

Received: 15 June 2020 / Accepted: 4 November 2022 / Published online: 21 November 2022


**Abstract**

Covering a graph with cohesive subgraphs is a classical problem in theoretical computer science, for example when the cohesive subgraph model considered is a clique. In this paper, we consider as a model of cohesive subgraph the 2-clubs, which are induced subgraphs of diameter at most 2. We prove new complexity results on the Min 2-Club Cover problem, a variant recently introduced in the literature which asks to cover the vertices of a graph with a minimum number of 2-clubs. First, we answer an open question on the decision version of Min 2-Club Cover that asks if it is possible to cover a graph with at most two 2-clubs, and we prove that it is W[1]-hard when parameterized by the distance to a 2-club. Then, we consider the complexity of Min 2-Club Cover on some graph classes. We prove that Min 2-Club Cover remains NP-hard on subcubic planar graphs, W[2]-hard on bipartite graphs when parameterized by the number of 2-clubs in a solution, and fixed-parameter tractable on graphs having bounded treewidth.

**Keywords** Graph algorithm · Cohesive subgraphs · 2-Clubs · Parameterized complexity

## 1 Introduction

Covering a graph with cohesive subgraphs, in particular cliques, is a relevant problem in theoretical computer science with many practical applications. Two classical problems in this direction are the Minimum Clique Cover problem and the Minimum

A preliminary version of the paper appeared in [12].

✉ Riccardo Dondi
riccardo.dondi@unibg.it

Manuel Lafond
manuel.lafond@USherbrooke.ca

1 Università degli Studi di Bergamo, Bergamo, Italy

2 Université de Sherbrooke, Sherbrooke, QC, Canada

Clique Partition problem [20], which are well-known to be NP-hard [26]. The first problem asks for the minimum number of cliques in a graph that cover all its edges, while the second problem asks for the minimum number of cliques in a graph that cover all its vertices. Notice that while this latter problem asks to cover all the vertices of a graph with cliques, we can always assume that the cliques partition the set of vertices. Indeed, if a vertex belongs to more than one clique, we can remove it from all the cliques except for one.

Covering the vertices of a graph with minimum number of vertices is a fundamental problem in graph mining, for decomposing a graph into cohesive modules and identify communities, with applications for example in computational biology [28] or in the analysis of transportation network [15]. Notice that Minimum Clique Partition is related to Graph Coloring, since a partition into cliques of the vertices of a graph corresponds to a coloring of the complement of the graph.

Minimum Clique Partition is known to be NP-hard even in restricted cases when the input graph is planar and cubic [7], in unit disk graphs [8], while admitting a PTAS for this graph class [16, 40]. Moreover, Minimum Clique Cover and Minimum Clique Partition are not approximable within a factor of $|V|^{1-\varepsilon}$ for every $\varepsilon > 0$, unless P = NP [46]. As for parameterized complexity, Minimum Clique Partition is unlikely to be in the class XP when parameterized by the number of cliques in the solution, as deciding if it is possible to color a graph with three colors is an NP-complete problem [19]. On the other hand, Minimum Clique Cover is fixed-parameter tractable when parameterized by the number of cliques in the solution, [22, 36] and the fastest parameterized algorithm has time complexity $O^*(2^{2^k})$ and it is based on finding a kernel of at most $2^k$ vertices for the problem [22].

These two problems are based on the clique model, that is a subgraph whose vertices are all pairwise connected, and ask for cliques that cover the input graph. Because the clique model is often considered too strict, other definitions of cohesive graphs have been considered in the literature, some of them called *relaxed cliques* [29], and rather ask for subgraphs that are "close" to a clique. For example, while each pair of distinct vertices in a clique are at distance exactly one, an $s$-*club* relaxes this constraint and is defined as an induced subgraph of diameter at most $s$, that is its vertices are at distance at most $s$ from each other in the subgraph. A different but related model, called $s$-clique, is defined as a subgraph whose vertices are at distance at most $s$ in the input graph, but not necessarily in the induced subgraph. Another alternative to cliques are $s$-plexes, where a subgraph is an $s$-plex if the minimum degree of a vertex in it is at least the size of the subgraph minus $s$. The minimum $s$-plex partition problem is studied in [23], the problem of editing edges to obtain an $s$-plex partition is studied in [24], and [43] asks to find $k$ $s$-plexes that cover a maximum number of vertices.

In this paper, we focus on the $s$-club model, which have several applications. In [38] the analysis of protein interactions is based on clustering a network with minimum number of $s$-clubs. A similar approach has been considered in [6] to analyze social networks. The $s$-club model has also been applied to edit a graph into disjoint clusters ($s$-clubs) [11, 18, 32]. A 1-club is a clique, so a natural step towards generalizing cliques using distances is to study the $s = 2$ case, especially given that 2-clubs have applications in social network analysis and bioinformatics [1, 4, 31, 34, 35, 44]. Hence, we mainly concentrate our efforts on 2-clubs.

Finding 2-clubs and, more generally $s$-clubs, of maximum size, a problem known as Maximum s-Club, has been extensively studied in the literature. Maximum s-Club is NP-hard, for each $s \geq 1$ [5]. Furthermore, the decision version of the problem that asks whether there exists an $s$-club larger than a given size in a graph of diameter $s+1$ is NP-complete, for each $s \geq 1$ [4].

Maximum s-Club has also been studied in the parameterzied complexity framework. The problem is fixed-parameter tractable when parameterized by the size of an $s$-club [9, 30, 42]; the fastest parameterized algorithm has running time $O(|V|(|V|+|E|)+|V|((k-2)^k \cdot k! \cdot k^3))$ [42]. Moreover the problem has been studied for structural parameters in chordal graphs and weakly chordal graphs [21, 25]. As for the approximation complexity, Maximum s-Club on an input graph $G=(V,E)$ is approximable within factor $|V|^{1/2}$, for every $s \geq 2$ [2] and not approximable within factor $|V|^{1/2-\varepsilon}$, for each $\varepsilon > 0$ and $s \geq 2$, unless P = NP [2].

Recently, the relaxation approach of $s$-clubs has been applied to the problem of covering a graph with $s$-clubs instead of the classical approach that asks for covering a graph with cliques. More precisely, the Min s-Club Cover problem asks for a minimum collection $\{C_1, \ldots, C_h\}$ of subsets of vertices (possibly not disjoint) whose union contains every vertex, and such that every $C_i$, $1 \leq i \leq h$, is an $s$-club. This problem has been considered in [13], in particular for $s = 2, 3$. The decision version of the problem is NP-complete when it asks whether it is possible to cover a graph with two 3-clubs, and whether is possible to cover a graph with three 2-clubs [13]. Min 3-Club Cover on an input graph $G=(V,E)$ has been shown to be not approximable within factor $|V|^{1-\varepsilon}$, for each $\varepsilon > 0$, while Min 2-Club Cover on an input graph $G=(V,E)$ is approximable within factor $O(|V|^{1/2}\log^{3/2}|V|)$ and not approximable within factor $|V|^{1/2-\varepsilon}$ [13].

Another combinatorial problem recently introduced that considers $s$-club as a model of cohesive subgraph asks for a set of at most $r$ disjoint $s$-clubs, each one of size at least $t \geq 2$, that covers the maximum number of vertices of a graph [14, 45]. Notice that in this case the $s$-clubs must be disjoint and are not constrained to cover the whole graph. This problem is NP-hard [14, 45] and fixed-parameter tractable when parameterized by the number of covered vertices [14].

In this paper, we present results on the complexity of Min 2-Club Cover. In Sect. 3 we answer an open question on the decision version of Min 2-Club Cover that asks if it is possible to cover a graph with at most two 2-clubs, and we prove that it is not only NP-hard, but W[1]-hard even when parameterized by the parameter “distance to 2-club”. Notice that, in contrast, the decision problem that asks if it is possible to cover a graph with two cliques is in P. Our hardness result is obtained showing the W[1]-hardness when parameterized by $k$ of an intermediate problem, called Steiner-2-Club (that may be of independent interest). Then, we consider the complexity of Min 2-Club Cover on some graph classes. In Sect. 4 we prove that Min 2-Club Cover is NP-hard on subcubic planar graphs. In Sect. 5 we prove that Min 2-Club Cover on a bipartite graph $G=(V,E)$ is W[2]-hard when parameterized by the number of 2-clubs in a solution and not approximable within factor $\Omega(\log(|V|))$. Finally, we prove in Sect. 6 that Min 2-Club Cover is fixed-parameter tractable on graphs having bounded treewidth. We start in Sect. 2 by giving some definitions and by defining formally the Min 2-Club Cover problem.

## 2 Preliminaries

Given a graph $G = (V, E)$ and a subset $W \subseteq V$, we denote by $G[W]$ the subgraph of $G$ induced by $W$. Given two disjoint subsets $X, Y \subseteq V$, we say that $X$ and $Y$ are *fully adjacent* if, for every $x \in X$, $y \in Y$, it holds that $xy \in E$. Given two vertices $u, v \in V$, the distance between $u$ and $v$ in $G$, denoted by $d_G(u, v)$, is the number of edges on a shortest path from $u$ to $v$. 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 $N_G[v] = N_G(v) \cup \{v\}$. If $G$ is understood, we may drop the $G$ subscript. For a vertex $v$ of $G$, let $N^2(v) = N(v) \cup \bigcup_{u \in N(v)} N(u)$, i.e. the neighbors of $v$ plus the neighbors of neighbors of $v$. We also use $N^2[v] = N^2(v) \cup \{v\}$ (notice that $N^2[v] = N^2(v)$ unless $v$ is an isolated vertex). Given a set $V' \subseteq V$, define $N(V') = \{u : \{v, u\} \in E, v \in V'\} \setminus V'$.

**Definition 1** Given a graph $G = (V, E)$, a subset $V' \subseteq V$, such that $G[V']$ has diameter at most 2, is a 2-club.

Notice that a 2-club must be connected, and that $d_{G[V']}(u, v)$ might differ from $d_G(u, v)$.

Now we present the definition of the problem we are interested in, called Minimum 2-Club Cover.

***Problem 1*** Minimum 2-Club Cover (Min 2-Club Cover)
**Input:** A graph $G = (V, E)$.
**Output:** A minimum cardinality collection $\mathcal{C} = \{V_1, \ldots, V_h\}$ such that, for each $i$ with $1 \leq i \leq h$, $V_i \subseteq V$, $V_i$ is a 2-club, and, for each vertex $v \in V$, there exists a set $V_j \in \mathcal{C}$ such that $v \in V_j$.

Notice that the 2-clubs in $\mathcal{C} = \{V_1, \ldots, V_h\}$ do not have to be disjoint. We denote by 2-Club Cover(h), with $1 \leq h \leq |V|$, the decision version of Min 2-Club Cover that asks whether there exists a cover of $G$ consisting of at most $h$ 2-clubs.

We present the definitions of nice tree decomposition of a graph [27], that will be useful in Sect. 6.

**Definition 2** Given a graph $G = (V, E)$, a nice tree decomposition of $G$ is a rooted tree $T = (B, E_B)$ (we denote $|B| = l$), where each vertex $B_i \in B$, $1 \leq i \leq l$, is a bag (that is $B_i \subseteq V$), with $|B_i| \leq \delta + 1$, such that:

1. $\bigcup_{i=1}^{l} B_i = V$
2. For every $\{u, v\} \in E$, there is a bag $B_j \in B$, with $1 \leq j \leq l$, such that $u, v \in B_j$
3. The bags of $T$ containing a vertex $u \in V$ induce a subtree of $T$.
4. Each $B_i \in B$ can be:

   (a) An *introduce vertex*: $B_i$ has a single child $B_j$, with $B_i = B_j \cup \{u\}$, where $u \in V$

(b) A *forget vertex*: $B_i$ has a single child $B_j$, with $B_i = B_j \setminus \{u\}$, where $u \in V$
(c) A *join vertex*: $B_i$ has exactly two children $B_l$, $B_r$ with $B_i = B_l = B_r$.

Each leaf-bag is associated with a single vertex of $V$.

## 3 W[1]-hardness of 2-Club Cover(2) for Parameter Distance to 2-Club

In this section, we show that the 2-Club Cover(2) problem, i.e. deciding if a graph can be covered by two 2-clubs, is W[1]-hard for the parameter "distance to 2-club", which is the number of vertices to be removed from the input graph $G = (V, E)$ such that the resulting graph is a 2-club. Note that Max 2-Club is fixed-parameter tractable for this parameter [42], in fact, Max s-club is FPT in the parameter "distance to $s$-club" for all $s \geq 1$). This result is given by introducing an intermediate problem, called the Steiner-2-Club. We first show that Steiner-2-Club is W[1]-hard, even in a restricted case, then we give a parameterized reduction from this restriction of Steiner-2-Club to 2-Club Cover(2) for the parameter distance to 2-club, thus showing that also this latter problem is W[1]-hard.

We start by introducing the Steiner-2-Club problem.

***Problem 2*** Steiner-2-Club
**Input:** A graph $G_s = (V_s, E_s)$, and a set $X_s \subseteq V_s$.
**Output:** Does there exist a 2-club in $G_s$ that contains every vertex of $X_s$?

We call $X_s$ the set of *terminal vertices*. We show that Steiner-2-Club is W[1]-hard for parameter $|X_s|$, by a parameter-preserving reduction from Multicolored Clique. Next, we recall the definition of the Multicolored Clique problem.

***Problem 3*** Multicolored Clique
**Input:** A graph $G_c = (V_c, E_c)$, where $V_c$ is partitioned into $k$ independent sets $V_{c,1}, \ldots, V_{c,k}$ (hereafter called the *color classes*).
**Output:** Does there exist a clique $V'_c \subseteq V_c$ such that $|V'_c| = k$ and for each $1 \leq i \leq k$, $|V'_c \cap V_{c,i}| = 1$?

It is well-known that Multicolored Clique is W[1]-hard for parameter $k$ [17].

Our proof holds on a restriction of Steiner-2-Club, called Restricted Steiner-2-Club, where the set $X_s$ is an independent set, $|X_s| > 4$, and each vertex in $V_s \setminus X_s$ has at most 2 neighbors in $X_s$. We start by giving a hardness result for Restricted Steiner-2-Club.

**Theorem 3** *The* Restricted Steiner-2-Club *problem is $W[1]$-hard with respect to the number of terminal vertices $|X_S|$.*

***Proof*** Let $G_c = (V_c, E_c)$ be an instance of Multicolored Clique, where $V_c$ is partitioned into color classes $V_{c,1}, \ldots, V_{c,k}$. We construct a corresponding instance $(G = (V_s, E_s), X_s)$ of Restricted Steiner-2-Club, where $|X_s| = k + 1$, as follows (see an example in Fig. 1).

Define the set $X_s$ of terminal vertices as follows:

$$X_s = \{x_0\} \cup \{x_i : V_{c,i} \text{ is a color class of } G_c\}$$

where $x_0$ is a special dummy vertex.

The set $V_s \backslash X_s$ of non terminal vertices is defined as:

$$V_s \backslash X_s = \bigcup_{v \in V_c} W_v$$

where $W_v$ is defined as follows:

$$W_v = \{w_{v,i} : 0 \leq i \leq k\}$$

Formally, we then define the edge set $E_s = E_s^1 \cup E_s^2 \cup E_s^3 \cup E_s^4$ where:

$$\begin{aligned}
E_s^1 &= \{\{w_{v,i}, w_{v,j}\} : v \in V_c, 0 \leq i < j \leq k\} \\
E_s^2 &= \{\{x_i, w_{v,i}\} : v \in V_c, 0 \leq i \leq k\} \\
E_s^3 &= \{\{x_i, w_{v,j}\} : v \in V_{c,i}, 1 \leq i \leq k, 0 \leq j \leq k\} \\
E_s^4 &= \{\{w_{u,i}, w_{v,i}\} : \{u, v\} \in E_c, 1 \leq i \leq k\}.
\end{aligned}$$

In words, the edges of $G_s$ are as follows: (1) each $W_v$ is a clique; (2) for each $i \in \{0, 1, \ldots, k\}$ and each $v \in V_c$, we add an edge between $x_i$ and $w_{v,i}$ because they share $i$ in their subscript; (3) for each $i \in \{1, \ldots, k\}$ and each vertex $v$ of color class $i$, we add all possible edges between $x_i$ and $W_v$; and (4) for $\{u, v\} \in E_c$ and each $i \in \{1, \ldots, k\}$, we and an edge between $w_{u,i}$ and $w_{v,i}$, i.e. there is a matching between $W_u$ and $W_v$ based on the non-zero $i$ subscripts. Notice that there is no edge $\{w_{u,0}, w_{v,0}\}$, with $\{u, v\} \in E_c$.

Also note that $G_s = (V_s, E_s)$ is an instance of Restricted Steiner-2-Club, since $X_s$ is an independent set and each vertex $w_{v,i}$, with $v \in V_{c,j}$, $0 \leq i \leq k$ and $1 \leq j \leq k$, is connected to at most two vertices of $X_s$, namely $x_i$ and $x_j$. We will use that fact a few times in the proof.

We now show that $G_c$ has a multicolored clique of size $k$ if and only if $G_s$ has a 2-club containing $X_s$.

($\Rightarrow$) Suppose that $G_c$ has a multicolored clique $v_1, \ldots, v_k$, where we assume that $v_i \in V_{c,i}$, $1 \leq i \leq k$, i.e. each $v_i$ is of color $i$. We claim that $C := X_s \cup W_{v_1} \cup \ldots \cup W_{v_k}$ is a 2-club. Consider two distinct vertices $y$ and $z$ of $C$. We show that $y$ and $z$ are at distance at most 2 in $G_s[C]$. There are three possible cases for vertices $y$ and $z$.

1. $y, z \in X_s$. Suppose that $y = x_i$ and $z = x_j$ for some $i, j \in \{0, \ldots, k\}$. If $i = 0$ and $j > 0$, then recall that $W_{v_j}$ is included in $C$, where $v_j$ is the vertex of color $j$ in the multicolored clique. Then $w_{v_j,0} \in W_{v_j}$ is a common neighbor of $x_0$ and $x_j$ in $C$ since $\{x_0, w_{v_j,0}\} \in E_s^2$ and $\{x_j, w_{v_j,0}\} \in E_s^3$. The case $j = 0$ is similar. If $i, j > 0$, then $W_{v_i}$ and $W_{v_j}$ are both included in $C$. In this case, $w_{v_i,j} \in W_{v_i} \subseteq C$ is a common neighbor of $x_i$ and $x_j$ since $\{x_i, w_{v_i,j}\} \in E_s^3$ amd $\{x_j, w_{v_i,j}\} \in E_s^2$.
2. $y \in X_s, z \in W_{v_j}$ for some $j \in \{1, \ldots, k\}$. Then $y = x_i$ and $z = w_{v_j,t}$ for some $i, t \in \{0, \ldots, k\}$. If $t \neq i$, then consider the vertex $w_{v_j,i} \in W_{v_j} \backslash \{w_{v_j,t}\}$. We have $\{x_i, w_{v_j,i}\} \in E_s^2$ and $\{w_{v_j,t}, w_{v_j,i}\} \in E_s^1$, and so $w_{v_j,i}$ is a common neighbor of

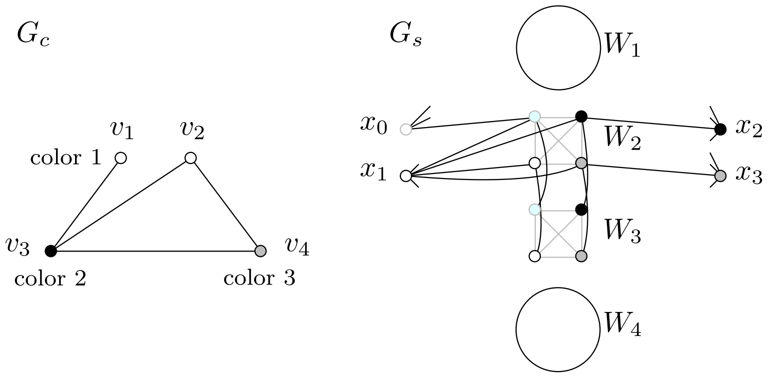


**Fig. 1** An illustration of the reduction. Left: a graph $G_c$ with vertices partitioned into 3 colors (1 is white, 2 is black, 3 is gray). Right: the corresponding graph $G_s$. For clarity, only the cliques $W_2$ and $W_3$ are drawn and their edges are grayed out (the $E_s^1$ edges). The color of the clique vertices corresponds to the second subscript of the vertex (for instance, $w_{2,3}$ is gray since it corresponds to color 3, and $w_{2,0}$ is represented with a gray stroke). The same color code is used for the $x_i$'s, since each $x_i$ corresponds to color $i$. Also, for the $x_i$'s we only show their incidents edges with an endpoint in $W_2$. Note that $x_1$ has all edges into $W_2$ since $v_2$ is of color 1 (the edges of $E_s^3$), and the other $x_i$'s have only one edge shared with $W_2$ (the edges of $E_s^2$). There are edges between $W_2$ and $W_3$ because $v_2v_3 \in E(G_c)$ (the edges of $E_s^4$). Not shown are the edges between $W_1$ and $W_3$, between $W_2$ and $W_4$, and between $W_3$ and $W_4$

$y = x_i$ and $z = w_{v_j,t}$ . If instead $t = i$, then $y = x_i$ and $z = w_{v_j,i}$ share an edge in $E_s^2$.

3. $y \in W_{v_r}, z \in W_{v_t}$ for some $r, t$ with $1 \leq r, t \leq k$. If $r = t$, then $y$ and $z$ are in the same clique $W_{v_r} = W_{v_t}$, thus they have distance one in $G_s[C]$. Hence consider the case that $r \neq t$, and let $y = w_{v_r,i}$ and $z = w_{v_t,j}$ for some $i, j \in \{0, \ldots k\}$. If $i = j = 0$, then $x_0 \in X_s \subseteq C$ is a common neighbor of $y = w_{v_r,0}$ and $z = w_{v_t,0}$ because of the $E_s^2$ edges. Assume that one of $i, j$ is not 0. Without loss of generality, we suppose that $j \neq 0$. Note that $\{v_r, v_t\} \in E_c$. Thus if $i = j > 0$, because of the $E_s^4$ edges, there is an edge between $y = w_{v_r,i}$ and $z = w_{v_t,j} = w_{v_t,i}$. So assume that $i \neq j$. Because $j > 0$, there exists an edge $\{w_{v_r,j}, w_{v_t,j}\} \in E_s^4$ and an edge $\{w_{v_r,j}, w_{v_r,i}\} \in E_s^1$. Then $y = w_{v_r,i}$ and $z = w_{v_t,j}$ are at distance at most 2 in $G_s[C]$.

This shows that every two of vertices in $G_s[C]$ are at distance at most 2, and therefore that $C$ is a 2-club.

($\Leftarrow$) Suppose that there is a 2-club $C$ in $G$ with $X_s \subseteq C$. We first claim that for each color class $i$ with $1 \leq i \leq k$, there exists a vertex $v_i \in V_{c,i}$ such that $w_{v_i,0} \in C$. Indeed, consider vertices $x_0, x_i \in C$, with $1 \leq i \leq k$. By construction $\{x_0, x_i\} \notin E_s$, hence there must exist a vertex $u \in C$ which is a neighbor of both $x_0$ and $x_i$ in $C$. Note that only $E_s^2$ specifies a set of neighbors for $x_0$, and that only vertices of the form $w_{v,0}$ are neighbors of $x_0$, where $v \in V_c$. Moreover, the definitions of $E_s^2$ and $E_s^3$ imply that the only vertices of the form $w_{v,0}$ that can be a neighbor of $x_i$ are those where $v \in V_{c,i}$. It follows that $u$ can only belong to some clique $W_v$ such that $v \in V_{c,i}$ and $u = w_{v,0}$. Since this is true for every $i \in \{1, \ldots, k\}$, our claim holds.

Now, for each $i$, with $1 \leq i \leq k$, choose any vertex $v_i \in V_{c,i}$ such that $w_{v_i,0} \in C$ (our previous claim implies that such a $v_i$ always exists). We claim that $\{v_1, v_2, \ldots, v_k\}$ is a clique of $G_c$.

To prove this, fix any color class $i$ with $1 \leq i \leq k$. Let $j \neq i$ be any other color class, with $1 \leq j \leq k$. Note that by the construction of $E_s^2$ and $E_s^3$ , $w_{v_i,0}$ and $x_j$ do not share an edge since $i \neq j$ and $j > 0$. Since $w_{v_i,0}$ and $x_j$ are both in $C$, they must have a common neighbor in $G[C]$. Consider such a common neighbor $z$ of $w_{v_i,0}$ and $x_j$. The set of neighbors of $w_{v_i,0}$ in $G_s$ is $\{x_0, x_i\} \cup (W_{v_i} \backslash \{w_{v_i,0}\})$, so $z$ must be in $W_{v_i}$. Since $v_i$ is of color $i \neq j$, the only neighbor of $x_j$ in $W_{v_i}$ is $w_{v_i,j}$ (because of $E_s^2$). Therefore, $w_{v_i,j} \in C$ for each $j \neq i$. Since this holds for every $i$, we have that , for each distinct $i, j$ with $1 \leq i, j \leq k$, $w_{v_i,j} \in C$. Combined with the fact that $w_{v_i,0} \in C$, this implies that $W_{v_1}, \ldots, W_{v_k}$ are each entirely contained in $C$.

We now argue that $v_i, v_j$ share an edge for any two distinct $i, j$, with $1 \leq i, j \leq k$. Let $h \notin \{i, j\}$ with $1 \leq h \leq k$. We know that $w_{v_i,h} \in C$. Consider the common neighbor $z'$ of $w_{v_i,h}$ and $w_{v_j,0}$ in $C$ (which must exist). The neighbors of $w_{v_j,0}$ are $\{x_0, x_j\} \cup (W_{v_j} \backslash \{w_{v_j,0}\})$, so $z'$ must be in $W_{v_j}$ (because the neighbors of $w_{v_i,h}$ in $X_s$ are $x_i$ and $x_h$, which are distinct from $x_0, x_j$ ). The edge set $E_s^4$ implies that the only possible neighbor of $w_{v_i,h}$ in $W_{v_j}$ is $w_{v_j,h}$, and the edge $\{w_{v_i,h}, w_{v_j,h}\}$ exists in $G_s$ if and only if $\{v_i, v_j\} \in E_c$. Since this holds for any $i, j$ pair, this shows that $\{v_1, \ldots, v_k\}$ is a clique. □

We now prove the hardness of 2-Club Cover(2).

**Theorem 4** *The 2-Club Cover(2) problem is W[1]-hard for the parameter distance to 2-club.*

***Proof*** Let $(G_s = (V_s, E_s), X_s)$ be an instance of Restricted Steiner-2-Club, where $k = |X_s|$ and $V_s = \{v_1, \ldots, v_n\}$. Without loss of generality, we will assume that $X_s = \{v_{n-k+1}, \ldots, v_n\}$. It follows from Theorem 3 that Restricted Steiner-2-Club is W[1]-hard when parameterized by $k$. Recall that in Restricted Steiner-2-Club $|X_s| = k > 4$.

Starting from $(G_s = (V_s, E_s), X_s)$, we construct an instance $G = (V, E)$ of 2-Club Cover(2), where $V = H \uplus W \uplus Y \uplus Z$ (here $\uplus$ means disjoint union). See Fig. 2 for an illustration of the graph $G$. First, we define the sets $H, W, Y, Z$ and the edges of the subgraphs $G[H]$, $G[W]$, $G[Y]$ and $G[Z]$, then the remaining edges of $G$. The subgraph $G[H] = (H, E_H)$ is a copy of $G_s$, and is defined as follows:

$$H = \{h_i : v_i \in V_s\} \quad E_H = \{\{h_i, h_j\} : \{v_i, v_j\} \in E_s\},$$

Moreover, define $H_X \subseteq H$ as follows

$$H_X = \{h_i \in H : v_i \in X_s\}.$$

Notice that, by construction, since $X_s$ is an independent set, it follows that $H_X$ is an independent set in $G$.

The subgraph $G[W] = (W, E_W)$ is a complete graph containing a vertex for each two vertices $v_i, v_j$ in $V'_s$, where $V'_s = V_s \backslash X_s$, with $1 \leq i < j \leq n - k$, defined as

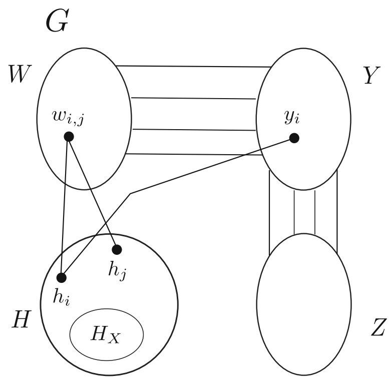


**Fig. 2** The structure of the graph $G$ built by the reduction. $W$, $Y$, $Z$ are cliques, while $G[H]$ is isomorphic to $G_s$. Multiple lines between two sets represent that they are fully adjacent. An example of edges between $W$ and $H \setminus H_X$ and an example of edges between $Y$ and $H \setminus H_X$ are given

follows:

$$W = \{w_{i,j} : v_i, v_j \in V'_s\} \quad E_W = \{\{w_{i,j}, w_{h,l}\} : w_{i,j}, w_{h,l} \in W\}.$$

The subgraph $G[Y] = (Y, E_Y)$ is also complete and has a vertex for each $v_i \in V'_s$. It is defined as follows:

$$Y = \{y_i : v_i \in V'_s\} \quad E_Y = \{\{y_i, y_j\} : y_i, y_j \in Y\}.$$

The subgraph $G[Z] = (Z, E_Z)$ is yet another complete graph, which contains $k$ vertices.

$$Z = \{z_i : 1 \leq i \leq k\} \quad E_Z = \{\{z_i, z_j\} : z_i, z_j \in Z\}.$$

Finally, we define the edges in $E$ between two vertices that belong to different sets in $H$, $W$, $Y$ and $Z$.

1. $W$ and $Y$ are fully adjacent;
2. $Y$ and $Z$ are fully adjacent;
3. Each vertex $w_{i,j}$ of $W$ shares an edge with vertices $h_i$ and $h_j$ of $H$. More precisely, for each distinct $v_i, v_j \in V'_s$, $\{w_{i,j}, h_i\}, \{w_{i,j}, h_j\} \in E$.
4. Each vertex $y_i$ of $W$ shares an edge with the vertex $h_i$ of $H$. More precisely, for each $v_i \in V'_s$, $\{h_i, y_i\} \in E$.

Notice that, by construction, $W \cup Y$ and $Y \cup Z$ are cliques. Also notice that there are no edges between $H$ and $Z$.

We first prove that $G = (V, E)$ has a distance to 2-club of exactly $k$. First note that a vertex of $H_X$ and a vertex of $Z$ are at distance three in $G$, since there is no edge between $H$ and $Z$, and also because vertices of $H_X$ and $Z$ do not share any common neighbor in $G$. It follows that to obtain a 2-club from $G$, either all the vertices of $H_X$ or all the vertices of $Z$ have to be removed from $G$. This implies a distance of at least $k$ from a 2-club, since $|H_X| = |Z| = k$.

Next we prove in the following claim that $V \setminus H_X$ is a 2-club.

***Claim*** (1). $V \backslash H_X$ is a 2-club of $G$.

***Proof*** We prove that two vertices of $V \backslash H_X$ are at distance at most two in $G[V \backslash H_X]$. First, recall that $W$, $Y$ and $Z$ are cliques of $G$, hence the distance between two vertices of each of these subsets have distance at most one in $G[V \backslash H_X]$. Thus it is sufficient to argue that each vertex of $H \backslash H_X$ is at distance at most 2 from any other vertex. Consider the remaining cases:

- Any two vertices $w_{i,j}$, $y_h$, with $w_{i,j} \in W$ and $y_h \in Y$, are adjacent and any two vertices $y_h, z_l$, with $y_h \in Y$ and $z_l \in Z$ are adjacent. It then follows that any two vertices $w_{i,j} \in W, z_l \in Z$ are at distance 2 in $G[V \backslash H_X]$.
- Given two vertices $h_i, h_j \in H \backslash H_X$, with $i < j$, there exists a vertex $w_{i,j} \in W$ which is adjacent to $h_i$ and $h_j$. Hence $h_i$ and $h_j$ have distance at most two in $G[V \backslash H_X]$.
- Consider vertices $h_i \in H \backslash H_X$ and $w_{j,l} \in W$, then $h_i$ and $w_{j,l}$ are either adjacent (if $i = j$ or $i = l$), or there exists a vertex $w_{i,p}$ or $w_{p,i}$ that is adjacent to both $h_i$ and $w_{j,l}$. Hence they have distance at most 2 in $G[V \backslash H_X]$.
- Consider vertices $h_i \in H \backslash H_X$ and $y_t \in Y$, then $h_i$ and $y_t$ are either adjacent (when $i = t$) or there exists a vertex $y_i$ that is adjacent to both $h_i$ and $y_t$. Hence they have distance at most 2 in $G[V \backslash H_X]$.
- Consider vertices $h_i \in H \backslash H_X$ and $z_u \in Z$, then there exists a vertex $y_i$ which is adjacent to both $h_i$ and $z_u$. Hence they have distance 2 in $G[V \backslash H_X]$. □

Thus we have shown that $V \backslash H_X$ is a 2-club in $G$ and that $G$ has distance at most $|H_X| = k$ from a 2-club. It follows that $G$ has distance from 2-club exactly $k$.

In order to complete the proof, we have to show that there exists a solution of Restricted Steiner-2-Club on instance $(G_s, X_s)$ if and only $G$ can be covered by two 2-clubs.

First assume that Restricted Steiner-2-Club on instance $(G_s, X_s)$ admits a 2-club $C_s$ containing $X_s$. Then, we claim that $V \backslash H_X$ and $C = \{h_i \in H : v_i \in C_s\}$ are a solution of 2-Club Cover(2) on instance $G$, that is they are two 2-clubs of $G$ and cover every vertex of $V$. First notice that, since $X_s \subseteq C_s$, then $H_X \subseteq C$ and thus $C \cup (V \backslash H_X) = V$ as desired. It remains to show that $C$ and $V \backslash H_X$ are 2-clubs of $G$. By Claim 1, we already know that $V \backslash H_X$ is a 2-club of $G$. Moreover, since $G[H]$ is isomorphic to $G_s$ and $C_s$ is a 2-club of $G_s$, $C$ is also 2-club of $G$.

Conversely, suppose that $G = (V, E)$ can be covered by two 2-clubs $C_1$ and $C_2$. First, recall that vertices of $H_X$ and vertices of $Z$ are at distance 3 from each other. It follows that one of these 2-clubs, say $C_1$, satisfies $H_X \subseteq C_1$, while the other, in our case $C_2$, satisfies $Z \subseteq C_2$. We claim that $(W \cup Y) \cap C_1 = \emptyset$. Assume that there exists a vertex $w_{i,j} \in W \cap C_1$, where $v_i, v_j \in V_s'$ are the vertices of $G_s$ corresponding to $w_{i,j}$. Since $H_X \subseteq C_1$ and $H_X$ has only neighbors in $H \backslash H_X$, it must be that any vertex $h_l \in H_X$ has a common neighbor with $w_{i,j}$ in $G[C_1]$. Consider a common neighbor $r$ of $w_{i,j}$ and $h_l$ in $G[C_1]$. Then $r \in H \backslash H_X$. It follows that $r = h_i$ or $r = h_j$, since the only vertices of $H \backslash H_X$ adjacent to $w_{i,j}$ are $h_i$ or $h_j$. This holds for each $h_l \in H_X$, thus $H_X \subseteq N(h_i) \cup N(h_j)$. Because $G_s$ is a restricted instance, $v_i, v_j \in X_s$ have at most two neighbors in $V_s \backslash X_s$, therefore $h_i, h_j \in H \backslash H_X$ have at most two neighbors in $H_X$. Since $H_X \subseteq N(h_i) \cup N(h_j)$, we have $|H_X| \leq 4$, while $|H_X| = |X_s| > 4$ by assumption. This is a contradiction, thus there is no vertex in $W \cap C_1$.

Assume that there exists a vertex $y_i \in Y \cap C_1$, where $v_i \in V_s'$ is the vertex of $G_s$ corresponding to $y_i$. By construction, the common neighbor of each $h_j \in H_X$ and vertex $y_i \in Y$ is $h_i \in H \setminus H_X$. This implies that $H_X \subseteq N(h_i)$, again reaching a contradiction, since $G_s$ is a restricted instance and hence, by construction, $h_i$ has at most 2 neighbors in $H_X$, while $|H_X| > 4$. We can conclude that there is no vertex $y_i \in C_1$.

Our arguments imply that $(W \cup Y \cup Z) \cap C_1 = \emptyset$ and thus $C_1 \subseteq H$. Define a 2-club $C_s \subseteq V_s$ of $G_s$ as follows: $C_s = \{v_i : h_i \in C_1\}$. Since $C_1$ is a 2-club of $G$, and $G[H]$ is isomorphic to $G_s$, it follows that $C_s$ is a 2-club of $G_s$. Moreover, $H_X \subseteq C_1$, implying that $X_s \subseteq C_s$. Thus $C_s$ is a solution of Restricted Steiner-2-Club, implying that 2-Club Cover(2) is W[1]-hard when parameterized by distance to a 2-club. □

## 4 Hardness of Min 2-Club Cover in Subcubic Planar Graphs

In this section we prove that Min 2-Club Cover is NP-hard even if the input graph is connected, has maximum degree 3 (i.e. a subcubic graph) and it is planar. We present a reduction from the Minimum Clique Partition problem on planar subcubic graphs (we denote this restriction by Min Subcubic Planar Clique Partition). which is known to be NP-hard [7].

***Problem 4*** (Min Subcubic Planar Clique Partition)
**Input:** A planar subcubic graph $G_P = (V_P, E_P)$.
**Output:** A partition of $V_P$ into a minimum number of cliques of $G_P$.

We first prove that subcubic graphs have a specific type of matching,[1] which will be useful for our reduction. Moreover, a *triangle* in a graph is a clique of size 3.

**Lemma 5** *Let $G_P = (V_P, E_P)$ be a connected subcubic graph that is not isomorphic to $K_4$. Then there is a matching $F_P \subseteq E_P$ in $G_P$ that can be computed in polynomial time, with the following properties:*

(i) *every triangle of $G_P$ contains exactly one edge of $F_P$;*
(ii) *every edge of $F_P$ is contained in some triangle of $G_P$.*

***Proof*** First observe that an edge $\{u, v\} \in E_P$ can belong to at most 2 distinct triangles, as otherwise $u$ and $v$ would have degree more than 3, since $u$ and $v$ must have a distinct neighbor in every distinct triangle. Also note that a vertex of $G_P$, since we have assumed that $G_P$ is not a $K_4$, can belong to at most two distinct triangles. To see this, assume that $u \in V_P$ belongs to two distinct triangles $T_1, T_2$. Since $u$ has degree at most 3, $T_1$ and $T_2$ must share an edge. It follows that $u$ has degree 3, and we let its neighbors be $v, w, z$. Assume that $u$ belongs to a third triangle $T_3$. Then either this triangle contains only vertices in $v, w, z$, thus making $\{u, v, w, z\}$ a $K_4$ or it contains a vertex $y \notin \{u, v, w, z\}$. Since $y$ is in a triangle with $u$, $y \in N(u)$, thus $u$ would have degree greater than three.

Next, we show how to construct the the set $F_P$ explicitly, and we will show after that it indeed a matching, and that it satisfies all required conditions. Starting with $F_P = \emptyset$, apply the following two steps:

[1] Recall that a matching is a set of edges that share no endpoint.

1. Add to $F_P$ every edge that belongs to 2 triangles;
2. Let $\mathcal{T}_P$ be the set of triangles with no edge in $F_P$ after the previous step. Then, for every triangle $T_P \in \mathcal{T}_P$, choose one arbitrary edge of $T_P$ and add it to $F_P$.

It is clear that every edge of $F_P$ is in a triangle of $G_P$, and it is easy to see that $F_P$ can be constructed in polynomial time. Let us argue that $F_P$ is a matching. Suppose for contradiction that two distinct edges $\{x, y\}, \{y, z\} \in F_P$ with a common endpoint (that is $y$) are added in Step 1. Then $\{x, y\}$ belongs to two triangles formed by vertices $\{x, y, w\}$ and $\{x, y, w'\}$ for some $w, w' \in V_P$. But $y$ has neighbors $\{x, z, w, w'\}$ and is of degree at most three, which implies that $w = z$ or $w' = z$ (since $x \neq z, w, w'$ and $w \neq w'$). Let assume w.l.o.g. that $w' = z$. Now, $\{y, z\} \in E_P$ also belongs to two triangles, since it is added by Step 1, one of which is $\{x, y, z\}$ and the other $\{y, z, r\}$ for some $r \in V_P$, $r \neq x$. If $r = w$, then $G_P$ is a $K_4$ formed by $\{x, y, z, w\}$. If $r \neq w$, then $y$ has four neighbors $\{x, z, w, r\}$, all distinct, which is a contradiction.

Suppose instead that an edge $\{x, y\} \in E_P$ included in $F_P$ at Step 1 shares a vertex with an edge $\{y, z\} \in E_P$ included at Step 2. Then $y$ belongs to 3 distinct triangles, two from Step 1 and one from Step 2, which is not possible.

Finally, suppose that $\{x, y\} \in E_P$ and $\{y, z\} \in E_P$ are adjacent edges both included in $F_P$ in Step 2. Assume that $\{x, y\}$ was added to $F_P$ because of triangle $\{x, y, w\}$, and that $\{y, z\}$ was added to $F_P$ because of another triangle $\{y, z, w'\}$. If $w = w'$, then the edge $\{y, w\}$ belongs to these two triangles. In this case, $\{y, w\}$ would have been added in Step 1 and $\{x, y\}$ would not have been added in Step 2 because of $\{x, y, w\}$ (since this triangle would be covered by $\{y, w\}$). If $w \neq w'$, then $y$ has four neighbors $\{x, z, w, w'\}$, a contradiction. This shows that $F_P$ is a matching.

It remains to show that every triangle has an edge in $F_P$. If a triangle $T_P$ contains an edge $\{x, y\}$ such that $\{x, y\}$ is in two triangles, then $T_P$ will be covered in Step 1. If $T_P$ contains no such edge, one of its edges will be added in Step 2. This concludes the proof. □

We are now ready to describe our reduction. Informally, an instance $G$ of Min 2-Club Cover, is constructed starting from $G_P = (V_P, E_P)$ by subdividing every edge of $E_P \setminus F_P$, and, for every vertex obtained by the subdivision of an edge, by connecting it to a new dangling path of length two.

Next, we define the graph $G$ formally. Given a instance $G_P = (V_P, E_P)$ of Min Subcubic Planar Clique Partition, where $V_P = \{u_1, \dots, u_n\}$, we first compute a matching $F_P$ of $G_P$ that satisfies the requirements of Lemma 5. Then, define $G = (V, E)$, an instance of Min 2-Club Cover, where $V = V' \cup V_1 \cup V_B$ as follows. First, define $V' = \{v_i : u_i \in V_P\}$.

For each edge $\{u_i, u_j\} \in E_P \setminus F_P$, with $1 \leq i < j \leq n$, define:

$$V_1 = \{v_{i,j,1} : \{u_i, u_j\} \in E_P \setminus F_P\} \quad V_B = \{v_{i,j,2}, v_{i,j,3} : \{u_i, u_j\} \in E_P \setminus F_P\}.$$

Next, we define the edge set $E$ of $G$

$$\begin{aligned} E = &\{\{v_i, v_j\} : v_i, v_j \in V', \{u_i, u_j\} \in F_P\} \\ &\cup \{\{v_i, v_{i,j,1}\}, \{v_j, v_{i,j,1}\} : v_i, v_j \in V', v_{i,j,1} \in V_1, \{u_i, u_j\} \in E_P \setminus F_P\} \end{aligned}$$

$$\cup \{\{v_{i,j,t}, v_{i,j,t+1}\} : v_{i,j,t}, v_{i,j,t+1} \in V, t \in \{1,2\}\}.$$

Notice that $G$ has maximum degree three, since $G_P$ has maximum degree three. Indeed, the vertices in $V'$ have the same degree as the corresponding vertices in $G_P$, those in $V_1$ have degree exactly three and those in $V_B$ degree at most two.

Next we show that, since $G_P$ is planar, then also $G$ is planar. Informally, given a planar embedding of $G$, one can easily subdivide the edges of $G$ (the $V_1$ vertices) without changing the embedding, then successively attach vertices of degree one (the $V_B$ vertices) on this embedding.

To be more formal, recall that a graph is planar if and only if it does not contain a subgraph that is a subdivision of a $K_5$ (a clique of size 5) or a $K_{3,3}$ (a biclique of size 3). Indeed, the vertices of $V_B$ cannot belong to a subdivision of a $K_5$ or a $K_{3,3}$, since they don't belong to a cycle of $G$. Hence, it is sufficient to consider the subgraph $G[V' \cup V_1]$. Notice that the vertices in $V_1$ have degree two in $G[V' \cup V_1]$. But then, if $G[V' \cup V_1]$ contains a subdivision of a $K_5$ or a $K_{3,3}$, the same property holds for $G_P$, since the vertices of $V_1$ are obtained by subdiving edges of $G_P$, a contradiction to the planarity of $G_P$.

For the remainder of this section, set $q = |E_P| - |F_P|$, that is $q$ is the number of edges of $G_P$ that were subdivided in the construction of $G$.

**Lemma 6** *Given a planar cubic graph $G_P$ instance of* Min Subcubic Planar Clique Partition, *consider the corresponding instance $G$ of* Min 2-Club Cover. *If there exists a clique partition $\mathcal{C} = \{C_{P,1}, \ldots, C_{P,k}\}$ of $G_P$ with $k$ cliques, then there exists a solution of* Min 2-Club Cover *on instance $G$ consisting of $q + k$ 2-clubs.*

***Proof*** Recall that $G_P$ is a subcubic graph. Note that if $\mathcal{C} = \{C_{P,1}, \ldots, C_{P,k}\}$ is a clique partition of $G_P$, then each $C_{P,i}$, with $1 \leq i \leq k$, is either a triangle, two adjacent vertices or a singleton vertex of $G_P$, since we have assumed that $G_P$ is not a $K_4$. For each $C_{P,i} \in \mathcal{C}$, with $1 \leq i \leq k$, we define a corresponding 2-club $C_i$ in $G$. If $C_{P,i} = \{u_j\}$, with $1 \leq j \leq n$, that is it is a singleton, then define $C_i = \{v_j\}$, with $v_j \in V'$. Consider the case that $C_{P,i} = \{u_j, u_l\}$, with $1 \leq j, l \leq n$, i.e. $C_{P,i}$ is an edge of $G_P$. If $\{u_j, u_l\} \in F_P$, then $C_i = \{v_j, v_l\}$. If $\{u_j, u_l\} \in E_P \setminus F_P$, then $C_i = \{v_j, v_l, v_{i,l,1}\}$.

If $C_{P,i} = \{u_j, u_l, u_z\}$, then $C_{P,i}$ is a triangle in $G_P$. By construction, the matching $F_P$ contains an edge connecting two vertices of $v_j$, $v_l$, $v_z$. Thus, in $G$ there exists a cycle $D$ of length 5 that contains $v_j$, $v_l$, $v_z$. Then $D$ is a 2-club of $G$ and we define $C_i = D$. Since each vertex of $G_P$ belongs to a clique of $\{C_{P,1}, \ldots, C_{P,k}\}$, the 2-clubs $C_1 \ldots, C_k$ cover every vertex in $V'$. The vertices of $V_1 \cup V_B$ are covered with $q$ 2-clubs as follows. For each vertex of $V_1$, define a 2-club $\{v_{i,j,1}, v_{i,j,2}, v_{i,j,3}\}$. It follows that $G$ admits a cover with at most $q + k$ 2-clubs. □

**Lemma 7** *Given a graph $G_P$ instance of* Min Subcubic Planar Clique Partition, *consider the corresponding graph $G$ instance of* Min 2-Club Cover. *Then, any 2-club covering of $G$ contains strictly more than $q$ 2-clubs. Moreover, if there exists a solution $\mathcal{C} = \{C_1, \ldots, C_{q+k}\}$ of* Min 2-Club Cover *on instance $G$, for some $k \geq 1$, there exists a clique partition of $G_P$ with at most $k$ cliques.*

***Proof*** First, notice that the set $V_B$ contains $q$ vertices of degree 1, each of which must be covered by a distinct 2-club. Moreover in $G$, the distance between any such degree 1 vertex of $V_B$ and any vertex of $V'$ is at least 3. Therefore, any solution of Min 2-Club Cover on instance $G$ contains at least $q$ 2-clubs that do not contain any vertex of $V'$, which proves the first part of the lemma.

Now, let $\mathcal{C} = \{C_1, \ldots, C_{q+k}\}$ be a solution of Min 2-Club Cover on instance $G$. It follows that there are at most $k$ 2-clubs $D_1, \ldots, D_h$ of $\mathcal{G}$, $h \leq k$, that are used to cover the vertices of $V'$. For each such $D_i$, with $1 \leq i \leq h$, containing at least one member of $V'$, define a subgraph $C_{P,i}$ of $G_P$ as follows:

$$C_{P,i} = \{u_j : v_j \in D_i \cap V'\}.$$

We claim that each $C_{P,i}$, with $1 \leq i \leq h$, is a clique of $G_P$. To prove this claim, we show that every distinct $u_j, u_l \in C_{P,i}$, with $1 \leq j, l \leq n$, are connected by an edge in $G_P$. Consider the vertices $v_j, v_l \in V'$ corresponding to $u_j, u_l$. If $\{v_j, v_l\} \in E$, then $\{u_j, u_l\} \in E_P$, and our claim holds. Assume that $\{v_j, v_l\} \notin E$. Then there exists a vertex $z \in V$ such that $z \in D_i$ and $z$ is adjacent to both $v_j$ and $v_l$, because $v_j, v_l$ are at distance 2 in $G[D_i]$. If $z \in V_1$, by construction $z = v_{j,l,1}$, assuming w.l.o.g. $j < l$, then it follows that $\{u_j, u_l\} \in E_P$. So, suppose that $z \notin V_1$. Notice that by construction $z \notin V_B$, since the vertices of $V_B$ are not adjacent to vertices of $V'$. Then, $z = v_y \in V'$, with $1 \leq y \leq n$, where $v_y$ corresponds to vertex $u_y \in V_P$. It follows that $\{v_j, v_y\}, \{v_l, v_y\} \in E$ and that $\{u_j, u_y\}, \{u_l, u_y\} \in E_P$. By construction, since $v_{j,y,1}$ nor $v_{l,y,1}$ exist in $V_1$, it follows that $\{u_j, u_y\}, \{u_l, u_y\} \in F_P$. Since the edges in $F_P$ form a matching, this is a contradiction. We thus conclude that $\{u_j, u_l\} \in E_P$, and that $C_{P,i}$ is a clique, for each $i$ with $1 \leq i \leq h$.

It remains to show that a clique partition of $G_P$ of size at most $h$ can be obtained from $C_{P,1}, \ldots, C_{P,h}$. Notice that, since $D_1, \ldots, D_h$ cover $V'$, then by construction $C_{P,1}, \ldots, C_{P,h}$ cover $V_P$. It is easy to see that if two cliques $C_{P,i}$, $C_{P,j}$, with $1 \leq i < j \leq h$, share a vertex, we can remove the vertices from one of the two. We can repeat this procedure until we obtain a partition of $V_P$. This concludes the proof. □

From Lemma 6, Lemma 7 and from the NP-hardness of Min Subcubic Planar Clique Partition [7], we can conclude that Min 2-Club Cover is NP-hard on planar subcubic graphs.

**Theorem 8** *Min 2-Club Cover is NP-hard on planar subcubic graphs.*

## 5 Hardness of Min 2-Club Cover on Bipartite Graphs

We show that Min 2-Club Cover, on bipartite graphs, is (1) W[2]-hard when parameterized by $h$ (the number of 2-clubs in a solution of Min 2-Club Cover) and (2) not approximable within factor $\Omega(\log |V|)$ unless $P = NP$. We give a reduction from Minimum Set Cover to Min 2-Club Cover on bipartite graphs. Next, we recall the definition of Minimum Set Cover.

***Problem 5*** Minimum Set Cover (Minimum Set Cover)

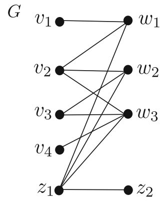


**Fig. 3** An example of the reduction from Minimum Set Cover to Min 2-Club Cover. $G$ is the bipartite graph that corresponds to an instance of Minimum Set Cover that consists of $U = \{u_1, u_2, u_3, u_4\}$ and three sets $S_1 = \{u_1, u_2\}$, $S_2 = \{u_2, u_3\}$, $S_3 = \{u_2, u_3, u_4\}$

**Input:** A set $U = \{u_1, \ldots, u_n\}$ of $n$ elements and a collection $\mathcal{S} = \{S_1, \ldots, S_m\}$ of sets, where $S_i \subseteq U$, with $1 \leq i \leq m$
**Output:** A minimum cardinality collection $\mathcal{S}' \subseteq \mathcal{S}$ such that for each element $u_i \in U$, with $1 \leq i \leq n$, there exists a set of $\mathcal{S}'$ containing $u_i$.

Minimum Set Cover is W[2]-hard when parameterized by the size of a cover [39].

**Theorem 9** *Min 2-Club Cover is W[2]-hard on bipartite graphs when parameterized by the number of 2-clubs in the cover.*

***Proof*** We describe the reduction from Minimum Set Cover to the Min 2 - Club Cover problem on bipartite graphs. Given an instance $(U, \mathcal{S})$ of Minimum Set Cover, in the following we define a bipartite graph $G = (V, E)$, which is an instance of Min 2-Club Cover, where $V = V_1 \uplus V_2$ (for an example see Fig. 3):

$$V_1 = \{v_i : u_i \in U, 1 \leq i \leq n\} \cup \{z_1\} \quad V_2 = \{w_j : S_j \in \mathcal{S}, 1 \leq j \leq m\} \cup \{z_2\}$$
$$E = \{\{v_i, w_j\} : u_i \in S_j, 1 \leq i \leq n, 1 \leq j \leq m\}$$
$$\cup \{\{z_1, w_j\} : 1 \leq j \leq m\}\} \cup \{z_1, z_2\}.$$

The graph $G$ is bipartite, as there is no edge connecting two vertices of $V_1$ or two vertices of $V_2$. Next, we prove the main results on which the reduction is based.

***Claim*** 9.1. Let $(U, \mathcal{S})$ be an instance of Minimum Set Cover and let $G = (V, E)$ be the corresponding instance of Min 2-Club Cover. Given a solution of Minimum Set Cover of size $z$, then a solution $\mathcal{C}$ of Min 2-Club Cover of size $z + 1$ can be computed in polynomial time.

***Proof*** First, consider a solution $\mathcal{S}'$ of Minimum Set Cover consisting of $z$ sets, we define a solution $\mathcal{C}$ of Min 2-Club Cover consisting of $z + 1$ 2-clubs as follows. For each $S_i$ in $\mathcal{S}'$, for some $i$ with $1 \leq i \leq m$, then the 2-club $N[w_i]$ belongs to $\mathcal{C}$; moreover the 2-club $N[z_1]$ belongs to $\mathcal{C}$.

We claim that each vertex of $G$ is covered by $\mathcal{C}$. First, notice that $N[z_1]$ covers each vertex $w_i$, with $1 \leq i \leq m$, and vertices $z_1, z_2$. Since $\mathcal{S}'$ covers each element of $U$, it follows by construction that each vertex $v_j$, with $1 \leq j \leq n$, belongs to a 2-club in $\mathcal{C}$. Finally, by construction, $\mathcal{C}$ contains $z + 1$ 2-clubs. □

***Claim*** 9.2 Let $(U, \mathcal{S})$ be an instance of Minimum Set Cover and let $G = (V, E)$ be the corresponding instance of Min 2-Club Cover as described above. Given a solution of Min 2-Club Cover of size $h$, with $h \geq 2$, a set cover of $(U, \mathcal{S})$ consisting of at most $h-1$ sets can be computed in polynomial time.

***Proof*** Consider a solution $\mathcal{C}$ of Min 2-Club Cover of size $h$, with $h \geq 2$. First, notice that $N^2[z_2] = \{z_1, z_2\} \cup \{w_j : S_j \in \mathcal{S}\}$ and that a 2-club containing $z_2$ must be a subset of $N^2[z_2]$. Since $N[z_1] = N^2[z_2]$ and $z_2$ must be covered, it follows that we can assume that $N[z_1]$ is a 2-club of $\mathcal{C}$. Note that $N[z_1]$ covers all the vertices in $\{z_1\} \cup \{z_2\} \cup \{w_j : S_j \in \mathcal{S}\}$.

Note that, for each $v_i \in V_1$, with $1 \leq i \leq n$, and each $w_j \in V_2$, with $1 \leq j \leq m$, such that $u_i \notin S_j$, we have $d_G(v_i, w_j) \geq 3$, as $N(v_i) = \{w_t : u_i \in S_t\}$, while $N(w_j) = \{v_p : u_p \in S_j\}$. As a consequence, each 2-club that contains a vertex $v_i \in V_1$, with $1 \leq i \leq n$, does not contain any $w_j \in V_2$, with $1 \leq j \leq m$, such that $u_i \notin S_j$. Next, starting from $\mathcal{C}$, we compute in polynomial time a solution $\mathcal{C}'$ of Min 2-Club Cover on instance $G$ such that (1) $\mathcal{C}'$ contains at most as many 2-clubs as $\mathcal{C}$ and (2) each 2-club of $\mathcal{C}' \setminus \{N[z_1]\}$ contains exactly one vertex $w_j \in V_2$, with $1 \leq j \leq m$. Assume that there exists a 2-club $X$ of $\mathcal{C} \setminus \{N[z_1]\}$ containing vertices $w_{j_1}, w_{j_2}$, $1 \leq j_1, j_2 \leq m$. Notice that, for each vertex $v_i \in X$, $1 \leq i \leq n$, we have shown that $u_i \in S_{j_1}, S_{j_2}$. Thus we can remove $w_{j_2}$ from $X$, and similarly each vertex of $(X \cap V_2) \setminus \{w_{j_1}\}$ since $X \setminus ((X \cap V_2) \setminus \{w_{j_1}\})$ is a 2-club of $G$ and each vertex of $(X \cap V_2) \setminus \{w_{j_1}\}$ is covered by the 2-club $N[z_1]$ of $\mathcal{C}$. Hence $X$ contains exactly one vertex of $V_2 \setminus \{z_2\}$. By repeating this procedure, we obtain a set $\mathcal{C}'$ of 2-clubs of $G$ that, as $\mathcal{C}$, covers $U$, such that (1) each 2-club of $\mathcal{C}'$ is a subset of $N[w_{j_1}]$, for some $w_{j_1} \in V_2$, (2) $|\mathcal{C}'| \leq |\mathcal{C}|$. Indeed, notice that by construction $\mathcal{C}'$ contains at most one 2-club for each 2-club of $\mathcal{C}$; furthermore, note that if a 2-club of $\mathcal{C}'$ does not contain vertices $w_j \in V_2$, with $1 \leq j \leq m$, it follows that it can cover at most one vertex $v_i$, with $1 \leq i \leq n$, thus we can replace this 2-club with a 2-club $N[w_j]$, with $1 \leq j \leq m$, such that $u_i \in S_j$.

Now, starting from $\mathcal{C}'$, we can define a solution $\mathcal{S}'$ of Minimum Set Cover consisting of the following sets:

$$\{S_j : w_j \text{ belongs to a 2-club of } \mathcal{C}' \setminus \{N[z_1]\}, 1 \leq j \leq m\}.$$

Since each vertex $v_i$, $1 \leq i \leq n$, is covered by some 2-club in $\mathcal{C}' \setminus \{N[z_1]\}$ containing exactly one vertex $w_j \in V_2$, it follows that $\mathcal{S}'$ covers every element in $U$. Finally, $\mathcal{S}'$ contains at most $h-1$ sets. □

From Claim 9.1, Claim 9.2 and from the W[2]-hardness of Minimum Set Cover [39] when parameterized by $h$, we can conclude that Min 2-Club Cover is W[2]-hard on bipartite graphs. □

As a consequence of Claim 9.1, Claim 9.2 we can prove also a bound on the approximation of Min 2-Club Cover on bipartite graphs.

**Corollary 10** *Min 2-Club Cover is not approximable within factor* $\Omega(\log(|V|))$ *on bipartite graphs unless* $P = NP$.

***Proof*** It follows from Claim 9.1 and Claim 9.2 that the reduction described is also an approximation preserving reduction [3]. Since Minimum Set Cover is not approximable within factor $\Omega(\log n)$ , even when $n$ and $m$ are polynomially related [33, 37], unless $P = NP$, it follows that Min 2-Club Cover is not approximable within factor $\Omega(\log n)$. By definition of graph $G = (V, E)$, $V = V_1 \uplus V_2$, where $|V_1| = n + 1$ and $|V_2| = m + 1$, thus $|V_1| + |V_2| = m + n + 2$. Since $n$ and $m$ are polynomially related, it follows that Min 2-Club Cover is not approximable within factor $\Omega(\log |V|)$, unless $P = NP$. □

## 6 An FPT Algorithm for Min 2-Club Cover on Graphs of Bounded Treewidth

In this section we show that Min 2-Club Cover is fixed parameter tractable when parameterized by the treewidth $\delta$ of the input graph $G$.

Let us note that the graph property of "being a 2-club" is expressible in Monadic Second Order logic (MSO) [41]. If it was possible to also express the Min 2-Club Cover problem in MSO, it would be fixed-parameter tractable in $\delta$ by Courcelle's theorem [10]. However, this seems difficult to achieve, since the number of 2-clubs in an optimal cover could be close to $n$. This makes it difficult to express in an MSO formula of bounded size, since the latter would need to specify that the property of "being a 2-club" applies to $\Theta(n)$ subsets of vertices. We therefore present a tree decomposition dynamic programming algorithm.

First, we present the algorithm, then we prove its correctness.

### 6.1 A Dynamic Programming Algorithm

From now on, we will assume that we are given a nice tree decomposition $T = (B, E_B)$ of $G$ (see Definition 2). We will further assume that the width of $T$ is $\delta$, so that every bag $B_i \in B$ has at most $\delta + 1$ vertices. We start by introducing some definitions related to $T = (B, E_B)$. We denote by $T_i$, with $1 \leq i \leq l$, the subtree of $T$ rooted at $B_i$, and we denote by $V(T_i)$ the vertices contained in at least one bag of $T_i$.

Given a 2-club $X$ of $G$ such that $X \cap V(T_i) \neq \emptyset$, with $1 \leq i \leq l$, $X \cap T(V_i)$ is called a *partial* 2-*club*. Notice that all the vertices of a partial 2-club have distance at most 2 in $G[X]$ but not necessarily in $G[X \cap V(T_i)]$. We prove now a property of partial 2-clubs.

**Lemma 11** *Given a partial* 2*-club* $X$*, of* $V(T_i)$*, with* $1 \leq i \leq l$*, then two vertices* $u, v \in X \cap (V(T_i) \setminus B_i)$ *have distance at most* 2 *in* $G[X \cap V(T_i)]$.

***Proof*** Consider vertices $u, v \in X \cap (V(T_i) \setminus B_i)$. Since $u, v \in V(T_i) \setminus B_i$, the third property of a nice tree decomposition implies that $N(u) \subseteq V(T_i)$ and $N(v) \subseteq V(T_i)$. Since $u, v$ in a 2-club of $G$, then $N(u) \cup N(v) \subseteq V(T_i)$, thus concluding the proof. □

As a consequence of Lemma 11, it follows that if $X \subseteq V(T_i)$ does not contain vertices of $B_i$, it is indeed a 2-club of $G[V(T_i)]$.

In order to bound the information we store in our dynamic programming tables, we will need the notion of a *succinct* partial 2-club.

**Definition 12** Let $B_i$ be a bag of $T$. A *succinct partial 2-club at* $B_i$ is an object $P$ that defines the following three components:

- $P[B_i]$ is a subset of $B_i$;
- given $u, v \in P[B_i]$, $P[u, v]$ is a value in $\{0, 1, 2, +\infty\}$;
- $P[out]$ is a subset of $2^{P[B_i]}$, the powerset of $P[B_i]$.

Roughly speaking, the goal of a succinct partial 2-club is to capture all the information of a partial 2-club, but without storing the actual vertices of $V(T_i)\backslash B_i$. The set $P[B_i]$ represents the subset of $B_i$ in the partial 2-club, $P[u, v]$ represents distances between $B_i$ vertices in the partial 2-club, and $P[out]$ represents all possible neighborhoods of vertices of $V(T_i)\backslash B_i$ in $B_i$ (see below).

More concretely, we present the following definition.

**Definition 13** Consider a solution $\mathcal{S}$ of Min 2-Club Cover on $G$ and a 2-club $X$ of $\mathcal{S}$. For a given bag $B_i$, let $P_X$ be a succinct partial 2-club at $B_i$. We say that $P_X$ *describes* $X$ if all of the following holds:

- $P_X[B_i] = X \cap B_i$;
- given $u, v \in X \cap B_i$,

$$P_X[u, v] = \begin{cases} 0 & \text{if } u = v \\ 1 & \text{if } d_{G[X \cap V(T_i)]}(u, v) = 1 \\ 2 & \text{if } d_{G[X \cap V(T_i)]}(u, v) = 2 \\ +\infty & \text{otherwise} \end{cases}$$

- $Z \in P_X[out]$ if and only if there is a vertex $z \in X \cap (V(T_i)\backslash B_i)$ such that $N(z) \cap P_X[B_i] = Z$.

In other words, $Z \in P_X[out]$ whenever there is some vertex $v$ whose neighborhood in $X \cap B_i$ is precisely $Z$.

Two succinct partial 2-clubs at $B_i$, say $P$ and $Q$, are equal if $P[B_i] = Q[B_i]$, $P[u, v] = Q[u, v]$ for all $u, v \in P[B_i]$ and $P[out] = Q[out]$. We will have to guess the succinct partial 2-clubs of a solution, and the following bound on the number of succinct partial 2-clubs will be useful.

**Lemma 14** *Let $B_i$ be a bag of $T$. Then there are at most $2^{4 \cdot 2^{\delta+1}}$ distinct succinct partial 2-clubs at $B_i$.*

***Proof*** Let $P$ be a succinct partial 2-club at $B_i$. There are $2^{\delta+1}$ possible values for $P[B_i]$. For $u, v \in P[B_i]$, there are 4 possible values for $P[u, v]$, and there are at most $(\delta + 1)^2$ pairs on which $P[u, v]$ is defined, and so there are at most $4^{(\delta+1)^2}$ ways to define the set of $P[u, v]$ entries. The number of distinct subsets in $P[out]$ is $2^{\delta+1}$, and each subset can be present or not. Thus there are at most $2^{2^{\delta+1}}$ ways to define the $P[out]$ entries.

Combining the possibilities, the number of distinct succinct partial 2-clubs is bounded by $2^{\delta+1} 4^{(\delta+1)^2} 2^{2^{\delta+1}} \leq 2^{4 \cdot 2^{\delta+1}}$. □

Our algorithm is somewhat technical, so we discuss the main intuition before delving into the details. For each subtree $T_i$, we want to know if it is possible to cover

$V(T_i)$ with $h$ partial 2-clubs, with $1 \leq h \leq n$ (since $n$ is an upper bound on the number of required partial 2-clubs). For technical reasons, we will allow not covering some $B_i$ vertices yet, and rather ask if $A_i \cup (V(T_i)\setminus B_i)$ can be covered with $h$ partial 2-clubs, where we ask this question for every $A_i \subseteq B_i$.

We distinguish two types of partial 2-clubs: those that are *complete*, in the sense that they are actually 2-clubs and are part of a global solution, and those that are *incomplete*, in the sense that they still need vertices from $V\setminus V(T_i)$ in a global solution (the notion of complete and incomplete 2-clubs is merely conceptual and not used in the upcoming formal framework).

For each bag $B_i$, we must store information on the incomplete partial 2-clubs for the parent of $B_i$. They will be completed as we go up the tree decomposition. We do not need to store the complete 2-clubs, as nothing needs to be added to them. Actually, it suffices to store only the incomplete partial 2-clubs that have vertices in $V(T_i)\setminus B_i$. The information that turns out to be necessary and sufficient for such an incomplete partial 2-club $X$ is all contained in its succinct representation $P_X$. These will tell us whether we can add a new vertex of $G$ in an introduce vertex of the given tree decomposition, or if we can merge two incomplete 2-clubs in a join vertex of the given tree decomposition.

Obviously, the partial 2-clubs, complete or incomplete, of an optimal solution are unknown, so we make a guess by storing every possible combination of succinct partial 2-clubs at each bag $B_i$. One important difficulty is that in a 2-club cover $\mathcal{S}$ of $G$, there may be many 2-clubs of $\mathcal{S}$ whose succinct representations at $B_i$ are equal. Therefore, there seems to be no upper bound on the number of partial, incomplete 2-clubs we need to store for the upper levels of the tree decomposition. However, in order to attain an FPT algorithm, we need to limit this number by a function of $\delta$. The following is a first step towards achieving this.

**Lemma 15** *Let $\mathcal{S}$ be an optimal solution of Min 2-Club Cover on instance $G$ and let $B_i$, $1 \leq i \leq l$, be a bag of $T$. Then there are at most $\delta + 1$ 2-clubs of $\mathcal{S}$ that have vertices in both $V(T_i)\setminus B_i$ and $V\setminus V(T_i)$.*

***Proof*** Let $\mathcal{Z} \subseteq \mathcal{S}$ be the subset of 2-clubs such that $Z \in \mathcal{Z}$ if and only if $Z \cap (V(T_i)\setminus B_i) \neq \emptyset$ and $Z\setminus V(T_i) \neq \emptyset$. Let $Z \in \mathcal{Z}$. Then for any $u \in Z\setminus V(T_i)$, $u$ must have a neighbor in $B_i$, as otherwise $u$ could not be at distance 2 from a vertex in $V(T_i)\setminus B_i$. Similarly, any vertex $v \in V(T_i)\setminus B_i$ must have a neighbor in $B_i$. Therefore, $\{\{u\} \cup N(u) : u \in B_i\}$ is a set of 2-clubs that covers the same vertices as the 2-clubs of $\mathcal{Z}$ that have neighbors in $V(T_i)\setminus B_i$. By the optimality of $\mathcal{S}$, we may thus assume that $\mathcal{Z}$ has at most $\delta + 1$ such 2-clubs. □

Thanks to Lemma 15, we have a bound of $\delta + 1$ on the number of partial 2-clubs that intersect with both the lower and upper levels of a bag $B_i$ in the tree decomposition.

Note that the above does not consider the number of partial, incomplete 2-clubs that contain only vertices in $B_i$ and $V\setminus V(T_i)$ (and nothing from $V(T_i)\setminus B_i$). There are examples in which this number is not bounded by a function of only $\delta$. However, we will not have to store those.

We now introduce the main definition that will be used to formalize the above intuitions and compute an optimal set of 2-clubs along the tree decomposition.

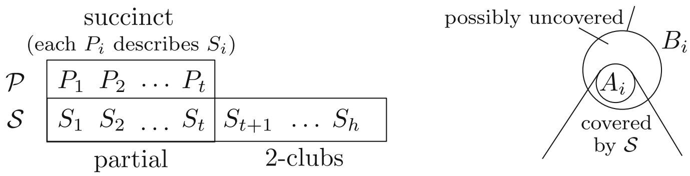


**Fig. 4** The main components behind Definition 16. $C[\mathcal{P}, A_i, h] = 1$ only when a set $\mathcal{S}$ as shown exists

**Definition 16** Let $\mathcal{P} = \{P_1, \ldots, P_t\}$ be a multi-set of succinct partial 2-clubs at $B_i$, and let $A_i \subseteq B_i$. Define a function $C[\mathcal{P}, A_i, h]$ in the range $\{0, 1\}$ that takes value 1 if and only if there exists a multi-set $\mathcal{S} = \{S_1, \ldots, S_h\}$ of $h \geq t$ partial 2-clubs, some of which are possibly empty, such that all the following conditions are satisfied:

1. for any $j$ with $1 \leq j \leq t$, $P_j$ describes $S_j$;
2. for any $j$ with $1 \leq j \leq t$, $S_j \cap (V(T_i)\setminus B_i) \neq \emptyset$;
3. $S_{t+1}, \ldots, S_h$ are 2-clubs of $G$;
4. $A_i \cup (V(T_i)\setminus B_i) \subseteq S_1 \cup S_2 \cup \ldots \cup S_h$.

Definition 16 is crucial for our purposes. In our treewidth-based dynamic programming table, we will store the succinct partial 2-clubs that satisfy all properties of the definition, as these contain exactly the information needed to compute the minimum 2-club cover. Figure 4 illustrates the components of the definition. In what follows, we will refer to the $i$-th condition of the definition, where $i \in \{1, 2, 3, 4\}$, as Definition 16.$i$. Intuitively speaking, Definitions 16.1 and 16.2 say that $\mathcal{P}$ contains the information on the incomplete partial 2-clubs of a solution that have vertices below and above $B_i$. Definition 16.3 says that only the first $t$ partial 2-clubs are incomplete, and the others are 2-clubs that do not need additional vertices. Definition 16.4 says that $\mathcal{S}$ must cover $V(T_i)\setminus B_i$, plus the $A_i$ subset. Note that this set $B_i\setminus A_i$ of uncovered leaves, we assume that it will be covered later (this is needed for technical reasons regarding join vertices).

The entries of $\mathcal{P}$ represent incomplete partial 2-clubs that contain vertices in both $V(T_i)\setminus B_i$ and $V(G)\setminus V(T_i)$. As a consequence of Lemma 15, later on we will be able to limit $|\mathcal{P}|$ to $\delta + 1$.

Now, we present a property of the bag at the root of the tree decomposition.

**Lemma 17** *Let $B_R$ be the bag at the root of the tree decomposition, then there exists a set of $h$ 2-clubs (non-partial) that covers $V$ if and only if $C[\emptyset, B_R, h] = 1$.*

***Proof*** Suppose that $C[\emptyset, B_R, h] = 1$. Then since $t = 0$, Definition 16.3 ensures that there are $h$ 2-clubs $S_1, \ldots, S_h$ of $G$ that, by Definition 16.4, cover all of $B_R \cup (V(T_R)\setminus B_R) = V(T_R) = V(G)$ are covered (since here $A_i = B_R$). Conversely, if there exists a set of $h$ 2-clubs $S_1, \ldots, S_h$ that cover $V(G)$, then Definition 16.1 and Definition 16.2 are vacuously satisfied, and it is easy to verify that the cover satisfies the remaining two elements of Definition 16, and so $C[\emptyset, B_R, h] = 1$. □

Next, we describe the recurrence to compute $C[\mathcal{P}, A_i, h]$, with three cases depending on whether the bag $B_i$ is a leaf, an introduce vertex, a forget vertex or a join vertex.

### 6.1.1 Leaf Case

When $B_i$ is a leaf of the tree decomposition and $B_i = \{u\}$, we put:

- $C[\emptyset, \emptyset, h] = 1$ for any $h$ with $0 \leq h \leq n$ since there is nothing to cover, and we can use $h$ empty partial 2-clubs to do so;
- $C[\emptyset, \{u\}, h] = 1$ for any $h$ with $1 \leq h \leq n$ since we can cover $u$ with the complete 2-club $\{u\}$, and have $h - 1$ empty 2-clubs;
- $C[\mathcal{P}, A_i, h] = 0$ if none of the above conditions are met. In particular, $\mathcal{P}$ must be empty since there cannot exist a partial 2-club with elements in $V(T_i)\backslash B_i$, as required by Definition 16.4.

### 6.1.2 Introduce Vertex

Let $B_i$ be an introduce vertex with child $B_j$, where $B_i = B_j \cup \{u\}$. Figure 5 shows how an entry $C[\mathcal{Q}, A_j, h']$ at $B_j$ can be used to determine whether $C[\mathcal{P}, A_i, h] = 1$.

Put $C[\mathcal{P}, A_i, h] = 1$ if and only if there exists an integer $h'$, a multi-set of succinct partial 2-clubs $\mathcal{Q}$ at $B_j$, and $A_j \subseteq B_j$ such that $C[\mathcal{Q}, A_j, h'] = 1$, and if there exists an ordering of the elements of $\mathcal{P}$ and $\mathcal{Q}$ so that $\mathcal{P} = \{P_1, \ldots, P_t\}$, $\mathcal{Q} = \{Q_1, \ldots, Q_s\}$ with $s \geq t$, and there exists an integer $b \leq t$, such that all of the following holds:

- (entries 1 to $b$ at $B_j$ remained the same)
  for each $k$ with $1 \leq k \leq b$, $P_k$ and $Q_k$ are equal.
- (we add $u$ to entries $b + 1$ to $t$)
  for each $k$ with $b + 1 \leq k \leq t$,
  - $P_k[B_i] = Q_k[B_j] \cup \{u\}$;
  - for each $v, w \in Q_k[B_j]$, let $d = 2$ if $\{u, v\}, \{u, w\} \in E(G)$, and $d = \infty$ otherwise. Then $P_k[v, w] = \min(d, Q_k[v, w])$;
  - for each $v \in Q_k[B_j]$, let $d$ be the distance between $u$ and $v$ in $G[P_k[B_i]]$ if this distance is at most 2, or let $d = \infty$ otherwise. Then $P_k[u, v] = d$;
  - $P_k[out] = Q_k[out]$. Moreover, for each $Z \in Q_k[out]$, $u$ has at least one neighbor in $Z$ (otherwise, $u$ cannot be at distance 2 from the vertices with neighborhood $Z$).
- (we added $u$ to entries $t + 1$ to $s$, they are now complete)
  for each $k$ with $t + 1, \ldots, s$, then adding $u$ to $Q_k$ makes it a complete 2-club. That is, for each $v, w \in Q_k[B_j]$, either $Q_k[v, w] \leq 2$ or $\{u, v\}, \{u, w\} \in E(G)$; for each $v \in Q_k[B_j]$, $d_{G[P_k[B_i]]}(v, u) \leq 2$?; and for each $Z \in Q_k[out]$, $u$ has a neighbor in $Z$.

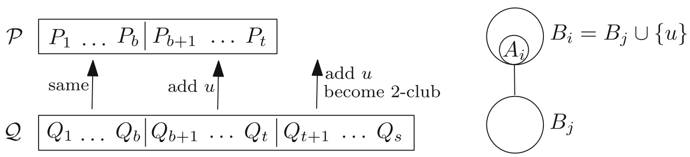


**Fig. 5** Idea behind introduce vertices

- (all $A_i$ vertices are covered)
  there exists a set of 2-clubs $R_1, \ldots, R_p$ in $G[B_i]$, each containing $u$, such that $A_i \subseteq A_j \cup (\bigcup_{k=1}^{t} P_k[B_i]) \cup (\bigcup_{k=t+1}^{s} (Q_k[B_j] \cup \{u\})) \cup (\bigcup_{k=1}^{p} R_k)$;
- $h = h' + p$, where $p$ is defined in the previous condition.

### 6.1.3 Forget Vertex

Let $B_i$ be a forget vertex and let $B_j$ be the only child of $B_i$, with $B_i = B_j \setminus \{u\}$ (Fig. 6).

Put $C[\mathcal{P}, A_i, h] = 1$ if and only if there exists a multi-set of succinct partial 2-clubs $\mathcal{Q}$ at $B_j$ such that $C[\mathcal{Q}, A_j, h'] = 1$, and if there exists an ordering of the elements of $\mathcal{P}$ and $\mathcal{Q}$ so that $\mathcal{P} = \{P_1, \ldots, P_t\}$, $\mathcal{Q} = \{Q_1, \ldots, Q_s\}$ with $s \leq t$ such that all of the following holds:

- for each $k$ with $1 \leq k \leq s$, if $u \notin Q[B_j]$, then $P_k$ and $Q_k$ are equal;
- for each $k$ with $1 \leq k \leq s$, if $u \in Q[B_j]$, then
  - $P_k[B_i] = Q_k[B_j] \setminus \{u\}$;
  - for each $v \in P_k[B_i]$, $Q_k[u, v] \leq 2$ (if not, $u$ and $v$ can never have distance 2 or less, even if we add new vertices);
  - for each $v, w \in P_k[B_i]$, we have $P_k[v, w] = Q_k[v, w]$;
  - Let $Q_k[out] = \{Z_1, \ldots, Z_l\}$. Then $P_k[out] = \{Z_1 \setminus \{u\}, \ldots, Z_l \setminus \{u\}\} \cup \{N(u) \cap P_k[B_i]\}$.
- for each $k$ with $s + 1 \leq k \leq t$, $P_k[B_i] \cup \{u\}$ is a partial 2-club, and $P_k$ describes $P_k[B_i] \cup \{u\}$;
- if $s = t$, then $A_j = A_i \cup \{u\}$. Otherwise, $A_j = A_i \setminus (P_{s+1}[B_i] \cup \ldots \cup P_t[B_i] \cup \{u\})$;
- $h = h' + (t - s)$.

### 6.1.4 Join Vertex

Let $B_i$ be a join vertex and let $B_l$, $B_r$ the left and right child, respectively, of $B_i$ (Fig. 7). Recall that $B_i = B_l = B_r$.

Put $C[\mathcal{P}, A_i, h] = 1$ if and only if there exist integers $h_l, h_r$, a set of succinct partial 2-clubs $\mathcal{L}$ at $B_l$, a set of succinct partial 2-clubs $\mathcal{R}$ at $B_r$, and subsets $A_l, A_r \subseteq B_i$ such that $C[\mathcal{L}, A_l, h_l] = C[\mathcal{R}, A_r, h_r] = 1$, and there exists an ordering $\mathcal{P} = \{P_1, \ldots, P_t\}$, $\mathcal{L} = \{L_1, \ldots, L_s\}$ and $\mathcal{R} = \{R_1, \ldots, R_q\}$, and integers $a, b$ with $0 \leq a \leq b \leq \min(s, q)$ such that:

- $t = q - a + s - b$;

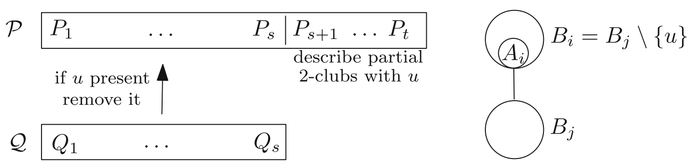


**Fig. 6** Idea behind forget vertices

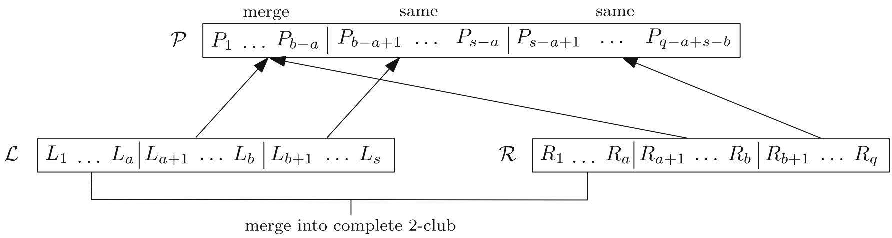


**Fig. 7** Idea behind join vertices

- for each $k$ with $1 \leq k \leq a$, $L_k$ and $R_k$ can be merged to form a complete 2-club. That is, the following holds:
  - $L_k[B_l] = R_k[B_r]$;
  - for each $u, v \in L_k[B_l] = R_k[B_r]$, we have $\min(L_k[u, v], R_k[u, v]) \leq 2$;
  - for each $Z_l \in L_k[out]$ and $Z_r \in R_k[out]$, we must have $Z_l \cap Z_r \neq \emptyset$ (to ensure that the vertices with neighborhoods $Z_l$ and $Z_r$ can be put in the same 2-club).
- for each $k$ with $a + 1 \leq k \leq b$, $L_k$ and $R_k$ are merged into an incomplete 2-club. That is, the following holds:
  - $P_{k-a}[B_i] = L_k[B_l] = R_k[B_r]$;
  - for each $u, v \in P_{k-a}[B_i]$, we have $P_{k-a}[u, v] = \min(L_k[u, v], R_k[u, v])$;
  - $P_{k-a}[out] = L_k[out] \cup R_k[out]$. Moreover, for each $Z_l \in L_k[out]$ and $Z_r \in R_k[out]$, we must have $Z_l \cap Z_r \neq \emptyset$ (as in the previous case, to ensure that the vertices with neighborhoods $Z_l$ and $Z_r$ can be put in the same 2-club).
- (the other entries are copied into $\mathcal{P}$)
  for each $k$ with $b + 1 \leq k \leq s$, $P_{k-a}$ and $L_k$ are equal, and for each $k$ with $b + 1 \leq k \leq q$, $P_{k-a+(s-b)}$ and $R_k$ are equal.
- $h = h_l + h_r - b$.
- $A_i = A_l \cup A_r$.

### 6.2 Correctness Proof

Next, we prove the correctness of the dynamic programming algorithm described in Sect. 6.1

**Lemma 18** *Consider a nice tree decomposition $(T, B)$ of a graph $G = (V, E)$ instance of Min 2-Club Cover, and let $B_i$ be a vertex of $T$, with $1 \leq i \leq l$. Given a set $\mathcal{P}$ of succinct partial 2-clubs at $B_i$, $A_i \subseteq B_i$ and $h \in \mathbb{N}$, then $C[\mathcal{P}, A_i, h] = 1$ if and only if there exists a set of $h$ partial 2-clubs $\mathcal{S}$ such that Definition 16 holds for $\mathcal{S}, \mathcal{P}, A_i$ and $h$.*

***Proof*** We prove the lemma by induction on the structure of $T$.

As a base is, suppose that $B_i$ is a leaf, with $B_i = \{u\}$. The correctness easily follows from the description of the base case given in the recurrence.

We now consider the inductive step. Given an internal vertex $B_i$ of the tree decomposition $T$, we assume that the lemma holds for each child of $B_i$ and we prove that the lemma holds for $B_i$.
($\Longrightarrow$) Assume that $C[\mathcal{P}, A_i, h] = 1$. We show that there exists a collection $\mathcal{S}$ of $h$ partial 2-clubs such that satisfies Definition 16.

We distinguish three cases, depending on the fact that $B_i$ is an introduce vertex, a forget vertex or a join vertex.

### 6.2.1 Introduce Vertex

Assume that $B_i$ is an introduce vertex, having child $B_j$, with $u \in B_i \setminus B_j$. By the definition of the recurrence, there exist a set of succinct partial 2-clubs $\mathcal{Q}$ at $B_j$, $A_j \subseteq B_j$ and $h'$ such that $C[\mathcal{Q}, A_j, h'] = 1$. Moreover, we may apply a labeling $\mathcal{P} = \{P_1, \ldots, P_t\}$, $\mathcal{Q} = \{Q_1, \ldots, Q_s\}$, and there exists an integer $b$ such that the recurrence is satisfied. By induction, there exists a set $\mathcal{S}' = \{S'_1, \ldots, S'_{h'}\}$ of $h' = h - p$ partial 2-clubs of $V(T_j)$, where $p$ is defined as in the recurrence, that satisfies Definition 16. Now, compute a set $\mathcal{S} = \{S_1, \ldots, S_h\}$ of $h$ partial 2-clubs of $V(T_i)$ starting from $\mathcal{S}'$ as follows. We show that Definition 16.1 holds while presenting the construction of $\mathcal{S}$. Consider an integer $k$ and the following cases:
**Case 1** $1 \le k \le b$.
Put $S_k = S'_k$. By the induction hypothesis, $Q_k$ describes $S'_k$. By the recurrence, $P_k$ is equal to $Q_k$, so it correctly describes $S_k$, so Definition 16.1 is satisfied.
**Case 2** $b + 1 \le k \le t$.
Put $S_k = S'_k \cup \{u\}$. Let us first argue that $S_k$ is indeed a partial 2-club. We only need to ensure that $u$ is at distance at most 2 from vertices below $B_i$. Let $z \in S'_k \setminus B_j$, and let $Z$ be the neighbors of $z$ in $B_i$. By induction, $Z \in Q_k[out]$ since $Q_k$ describes $S'_k$. Moreover, the recurrence requires that $u$ has a neighbor in $Z$, ensuring that $u$ and $z$ have distance at most 2 in $S_k$. Thus under the assumption that $S'_k$ is a partial 2-club, $S_k$ is also a partial 2-club.

We now argue that $P_k$ describes $S_k$. By induction, $Q_k[B_j] = S'_k \cap B_j$. By the recurrence, $P_k[B_i] = Q_k[B_j] \cup \{u\} = S_k \cap B_i$.

Let $v, w \in Q_k[B_j]$. If the shortest path between $v$ and $w$ in $S_k$ has length at most 2, then this path is either the same as in $S'_k$, or it uses $u$. Hence putting $P_k[v, w] = \min(d, Q_k[v, w])$ as in the recurrence is correct. Now let $v \in Q_k[B_j]$. By the properties of a tree decomposition, $u$ has no neighbor in $V(T_i) \setminus B_i$, so if the distance between $u, v$ in $S_k$ is at most 2, the shortest path only uses vertices of $S_k \cap B_i = P_k[B_i]$. Thus putting $P_k[u, v] = d$ as in the recurrence is correct. Thus the $P_k[v, w]$ and $P_k[u, v]$ entries correspond to the distances between $B_i$ elements in $S_k$.

Now let $Z \in P_k[out]$. By the recurrence, $Z \in Q_k[out]$ as well. Having $Z \in P_k[out]$ is therefore correct since any $z \in V(T_i) \setminus B_i$ has the same neighborhood in either $S'_k \cap B_j$ or $S_k \cap B_i$. Consider some $Z \notin P_k[out]$. If $u \in Z$, this is appropriate since no $z \in V(T_i) \setminus B_i$ has $u$ as a neighbor. Otherwise, $Z \notin Q_k[out]$ as well, which is correct by induction. It follows that $P_k$ describes $S_k$, as desired.
**Case 3** $t + 1 \le k \le s$.

In this case, put $\mathcal{S}[t+k-b] = S'_k \cup \{u\}$. Since this partial 2-club does not correspond to any entry of $\mathcal{P}$, it must be an actual 2-club to satisfy Definition 16.3. It is easy to verify from the recurrence that $S'_k \cup \{u\}$ is indeed a 2-club.
We have shown that Definition 16.1 is satisfied with $\mathcal{P}$ and $\mathcal{S}$ so far.

To finish the construction of $\mathcal{S}$, add to $\mathcal{S}$ all of $S'_{s+1}, \ldots, S'_{h'}$, which are 2-clubs by induction. Also add $R_1, \ldots, R_p$ to $\mathcal{S}$ as they are described in the recurrence. Note that these $p$ 2-clubs ar the only ones in $\mathcal{S}$ not in $\mathcal{S}'$, so $|\mathcal{S}| = h' + p$, as desired.

Since each entry $S_k$, $1 \leq k \leq t$, is either $S'_k$ or $S'_k \cup \{u\}$, it follows by induction that Definition 16.2 holds (i.e. each $S_k$ has vertices in $V(T_i)\backslash B_i$). Definition 16.3 holds because after $S_t$, we only add 2-clubs (either those resulting from $S'_{b+1}, \ldots, S'_s$ by adding $u$, those in $S'_{s+1}, \ldots, S'_{h'}$ that were already 2-clubs, or $R_1, \ldots, R_p$ which are 2-clubs).

Finally, we must show that Definition 16.4 holds. This is because $\mathcal{S}'$ covers $A_j$ by induction, and if any element of $A_i \backslash A_j$ is not in $S_1, \ldots, S_s$, then by the recurrence, such an element will be covered by some 2-club in $R_1, \ldots, R_p$.

### 6.2.2 Forget Vertex

Assume that $B_i$ is a forget vertex, with child $B_j$, and $u \in B_j \backslash B_i$. By the definition of the recurrence, there exist a set of succinct partial 2-clubs $\mathcal{Q}$ satisfying $C[\mathcal{Q}, A_i, h'] = 1$. Moreover, we may apply a labeling $\mathcal{P} = \{P_1, \ldots, P_t\}$, $\mathcal{Q} = \{Q_1, \ldots, Q_s\}$ such that the recurrence is satisfied.

By the induction hypothesis, there exists a set $\mathcal{S}' = \{S'_1, \ldots, S'_{h'}\}$ of $h'$ partial 2-clubs of $V(T_j)$ that satisfies Definition 16 with respect to $\mathcal{Q}$.

We construct the set $\mathcal{S} = \{S_1, \ldots, S_h\}$ of partial 2-clubs at $B_i$ as follows:

(1) for $k$ with $1 \leq k \leq s$, put $S_k = S'_k$;
(2) for $k$ with $s+1 \leq k \leq t$, put $S_k = P_k[B_i] \cup \{u\}$;
(3) for $k$ with $t+1 \leq k \leq h'$, append $S'_k$ to $\mathcal{S}$ (i.e. put $S'_k$ among $S_{t+1}, \ldots, S_h$).

We show that $\mathcal{S}$ satisfies Definition 16 with respect to $\mathcal{P}$.

To see that Definition 16.1 holds, consider $k$ with $1 \leq k \leq s$. If $u \notin Q_k[B_j]$, then $Q_k$ describes $S'_k = S_k$ and $P_k$ describes $S_k$ since it is made equal to $Q_k$. If $u \in Q_k[B_j]$, then $Q_k$ describes $S'_k = S_k$. In that case, $P_k[B_i] = Q_k[B_j]\backslash\{u\} = S_k \cap B_i$. Let $v, w \in P_k[B_i]$. Since $S'_k = S_k$, putting $P_k[v, w] = Q_k[v, w] = d_{G[S'_k]}(v, w)$ correctly describes the $v, w$ distance. Next, consider $z \in S_k \backslash B_i$. If $z \neq u$, then by induction $N(z) \cap Q_k[B_j]$ is in $Q_k[out]$, and it follows from the recurrence that $N(z) \cap P_k[B_i]$ is in $P_k[out]$. If $z = u$, then $N(u) \cap P_k[B_i]$ is in $P_k[out]$ by the recurrence. Therefore, $P_1, \ldots, P_s$ describes the first $s$ entries of $\mathcal{S}$. As for $k$ with $s+1 \leq k \leq t$, $P_k$ describes $S_k$ since we explicitly put $S_k = P_k[B_i] \cup \{u\}$. Thus Definition 16.1 holds for $\mathcal{P}$ and $\mathcal{S}$.

Consider Definition 16.2. For $k$ with $1 \leq k \leq s$, by induction $S'_k = S_k$ has vertices in $V(T_j)\backslash B_i$, and thus in $V(T_i)\backslash B_i$. For $k$ with $s+1 \leq k \leq t$, $S_k = P_k[B_i] \cup \{u\}$, and thus $S_k$ has vertices in $V(T_i)\backslash B_i$ (since $u \notin B_i$ and $P_k[B_i] \neq \emptyset$). Therefore, Definition 16.2 holds for $\mathcal{P}$ and $\mathcal{S}$.

The elements $S_{t+1}, \ldots, S_h$ of $\mathcal{S}$ are obtained from $S'_{t+1}, \ldots, S_{h'}$, which are 2-clubs by induction. Therefore, Definition 16.3 holds for $\mathcal{P}$ and $\mathcal{S}$.

Finally, consider Definition 16.4. If $A_j = A_i \cup \{u\}$, then by assumption $\mathcal{S}'$ covers $A_i \cup \{u\} \cup (V(T_j)\backslash B_j)$, from which it follows that $\mathcal{S}$ covers $A_i \cup (V(T_i)\backslash B_i)$. If $A_j = A_i\backslash(P_{s+1}[B_i]\cup\ldots\cup P_t[B_i]\cup\{u\})$, then $\mathcal{S}'$ covers $V(T_j)\cup A_j$, and $S_{s+1},\ldots,S_t$ contain the remaining vertices (in particular, $u$). Therefore, Definition 16.4 is satisfied.

We deduce that there exists a set of partial 2-clubs $\mathcal{S}$ such that Definition 16 is satisfied with respect to $\mathcal{P}$ and $A_i$.

### 6.2.3 Join Vertex

Assume that $B_i$ is a join vertex, with children $B_l$ and $B_r$, where $B_i = B_l = B_r$. Assume that

$$C[\mathcal{L}, A_l, h_l] = C[\mathcal{R}, A_r, h_r] = 1,$$

for some set of succinct partial 2-clubs at $B_l$ and $B_r$, respectively, subsets $A_l, A_r \subseteq B_i = B_l = B_r$, and integers $h_l, h_r$, defined as in the recurrence. These exist, since $C[\mathcal{P}, A_i, h] = 1$. Let us write $\mathcal{P} = \{P_1,\ldots,P_t\}$, $\mathcal{L} = \{L_1,\ldots,L_s\}$ and $\mathcal{R} = \{R_1,\ldots,R_q\}$. Let $a$ and $b$ be integers defined as in the recurrence.

By the induction hypothesis, there exists $\mathcal{S}^l$ ($\mathcal{S}^r$, respectively) of $h_l$ ($h_r$, respectively) partial 2-clubs that covers vertices in $A_l \cup (T_l\backslash B_l)$ (in $A_r \cup (T_r\backslash B_r)$, respectively) and that satisfies Definition 16. Let us write $\mathcal{S}^l = \{S^l_1,\ldots,S^l_{h_l}\}$ and $\mathcal{S}^r = \{S^r_1,\ldots,S^r_{h_r}\}$, where the first $s$ elements of $\mathcal{S}^l$ are in correspondence with $\mathcal{L}$, and the first $q$ elements of $\mathcal{S}^r$ in correspondence with $\mathcal{R}$. Now, starting from $\mathcal{S}^l$ and $\mathcal{S}^r$ construct a set $\mathcal{S} = \{S_1,\ldots,S_h\}$ of $h = h_l + h_r - b$ partial 2-clubs as follows:

- for $k$ with $a+1 \le k \le b$, put $S_{k-a} = S^l_k \cup S^r_k$;
  We argue now that $P_{k-a}$ describes $S_{k-a}$ to satisfy Definition 16.1. By the recurrence and by induction, $P_{k-a} = R_k[B_r] = L_k[B_l] = S^l_l \cap B_l = S^r_k \cap B_r = S_{k-a} \cap B_i$, as desired. Consider distinct $u, v \in P_{k-a}[B_i]$. If $\{u, v\} \in E(G)$, then they have distance 1 in $S^l_k$ and, by induction, $L_k[u, v] = 1$. Clearly, $P_{k-a} = \min(L_k[u, v], R_k[u, v]) = 1 = d_{G[S_{k-a}]}(u, v)$. If $d_{G[S_{k-a}]}(u, v) = 2$, then $u, v$ share a common neighbor in $S^l_k$ or $S^r_k$, and $P_{k-a} = \min(L_k[u, v], R_k[u, v]) = 2$ describes $S_{k-a}$. If $d_{G[S_{k-a}]}(u, v) > 2$, then $\min(L_k[u, v], R_k[u, v])$ will be $\infty$, which is correct. Finally, let $u \in S_{k-a}\backslash B_i$. Then either $u \in S^l_k\backslash B_l$ or $u \in S^r_k\backslash B_r$. In either case, if $Z$ is the neighborhood of $u$ in $B_i$, then $Z \in L_k[out]$ or $Z \in R_k[out]$ since $B_i = B_l = B_r$. By the recurrence, $Z \in L_k[out] \cup R_k[out] = P_{k-a}[out]$. Thus, $P_{k-a}$ describes $S_{k-a}$.
  Moreover, $S_{k-a}$ satisfies Definition 16.2 because by assumption, $S^l_k$ and $S^r_k$ satisfy Definition 16.2 (i.e. they have vertices in $V(T_l)\backslash B_l$ and $V(T_r)\backslash B_r$, respectively.
  We must also show that $S_{k-a}$ is a partial 2-club. By assumption, $\mathcal{S}^l_k$ and $\mathcal{S}^r_k$ are partial 2-clubs, and so each $u, v \in S^l_k\backslash V(T_l)$ are at distance at most 2 in $S^l_k$ and each $u, v \in S^r_k\backslash V(T_r)$ are at distance at most 2 in $S^r_k$. Since $S_{k-a} = S^l_k \cup S^r_k$, these $u, v$ distances cannot increase, and so their distance is also at most 2 in $S_{k-a}$. Moreover, each $u \in S^l_k\backslash V(T_l)$ and each $u \in S^r_k\backslash V(T_r)$ is at distance at most 2 with each $v \in S_{k-a} \cap B_i$. Consider $v \in \mathcal{S}^l_k\backslash V(T_l)$ and $w \in \mathcal{S}^r_k\backslash V(T_r)$, and let $Z_l$

and $Z_r$ be their neighborhoods in $P_{k-a}[B_i]$, respectively. The recurrence requires $Z_l \cap Z_r \neq \emptyset$, and so $v$ and $w$ have distance at most 2 in $S_{k-a}$.
- for $k$ with $b+1 \leq k \leq s$, put $S_{k-a} = S_k^l$. Hence $S_{k-a}$ is a partial 2-club and, since by induction $L_k$ describes $S_k^l$ and $P_{k-a}$ is equal to $L_k$ in the recurrence, $P_{k-a}$ describes $S_{k-a}$, satisfying Definition 16.1. Moreover, $S_{k-a}$ satisfies Definition 16.2 since $S_k^l$ does, by induction.
- for $k$ with $b+1 \leq k \leq q$, put $S_{k-a+(s-b)} = S_k^r$. Hence $S_{k-a+(s-b)}$ is a partial 2-club and, since by induction $R_k$ describes $S_k^r$ and $P_{k-a+(s-b)}$ is equal to $R_k$ in the recurrence, $P_{k-a+(s-b)}$ describes $S_{k-a+(s-b)}$, satisfying Definition 16.1. Moreover, $S_{k-a}$ satisfies Definition 16.2 since $S_k^r$ does, by induction.
- for $k$ with $1 \leq k \leq a$, put $S_{t+k} = S_k^l \cup S_k^r$. Then $S_{t+k}$ must be a 2-club to satisfy Definition 16.3. One may check that the recurrence has all the conditions required on $L_k$ and $R_k$, which describe $S_k^l$ and $S_k^r$, respectively, for $S_k^l \cup S_k^r$ to be a 2-club.
- for $k$ with $s+1 \leq k \leq h_l$, add $S_k^l$, which is a 2-club, to $\mathcal{S}$. Thus Definition 16.3 is satisfied.
- for $k$ with $q+1 \leq k \leq h_r$, add $S_k^r$, which is a 2-club, to $\mathcal{S}$. Thus Definition 16.3 is satisfied.

We have argued that Definition 16.1, 16.2 and 16.3 are satisfied. Summing over the above cases, the number of partial 2-clubs in $\mathcal{S}$ is $b-a+s-b+q-b+a+h_l-s+h_r-q = h_l+h_r-b = h_l+h_r-b = h$, as desired. It remains to show that Definition 16.4 holds. We see that $A_i$ is covered since $A_i = A_l \cup A_r$ and, by assumption, $\mathcal{S}^l$ covers $A_l$, $\mathcal{S}^r$ covers $A_r$, and every vertex in a partial 2-club in $\mathcal{S}^l \cup \mathcal{S}^r$ is added in $\mathcal{S}$.

We conclude that Definition 16 holds for $\mathcal{S}$.

($\Longleftarrow$) Assume that there exists a set $\mathcal{S} = \{S_1, \ldots, S_h\}$ of $h$ partial 2-clubs that satisfies Definition 16 with respect to $\mathcal{P}$ and $A_i$. We prove that $C[\mathcal{P}, A_i, h] = 1$ according to the recurrence. We distinguish three cases depending on the fact that $B_i$ is an introduce vertex, a forget vertex or a join vertex. Let us write $\mathcal{P} = \{P_1, \ldots, P_t\}$. Since we may relabel elements of $\mathcal{P}$, we will often assume that the $P_k$'s are ordered conveniently for our purposes.

### 6.2.4 Introduce Vertex

Assume that $B_i$ is an introduce vertex and that $B_j$ is the child of $B_i$ in $T$, with $u \in B_i \backslash B_j$.

To show that $C[\mathcal{P}, A_i, h] = 1$, we construct a list $\mathcal{S}'$ of $h'$ partial 2-clubs, a set $\mathcal{Q}$ of succinct partial 2-clubs at $B_j$, and $A_j \subseteq B_j$ such that Definition 16 is satisfied. If we achieve this, by induction we know that $C[\mathcal{Q}, A_j, h'] = 1$. We also prove that $\mathcal{Q}$, $A_j$ and $h'$ satisfy all the conditions of the recurrence to have $C[\mathcal{P}, A_i, h] = 1$.

We assume that we have ordered $\mathcal{S}$ and $\mathcal{P}$ so that there exist integers $b$ and $s$, with $b \leq t \leq s$, satisfying:

(1) $P_1, \ldots, P_b$, and thus $S_1, \ldots, S_b$, do not contain $u$;
(2) $P_{b+1}, \ldots P_t$, and thus $S_{b+1}, \ldots, S_t$, contain $u$;
(3) $S_{t+1}, \ldots, S_s$ are 2-clubs that contain $u$ but are not subsets of $B_i$;
(4) $S_{s+1}, \ldots, S_{s+p}$ are 2-clubs that contain $u$ and are subsets of $B_i$;

(5) $S_{s+p+1}, \ldots, S_h$ are 2-clubs that do not contain $u$.

The reader may observe that every element of $\mathcal{P}$ and $\mathcal{S}$ fits somewhere in these cases. We define $A_j = A_i \backslash (\{u\} \cup S_{s+1} \cup \ldots \cup S_{s+p})$ and $h' = h - p$.

We now define $\mathcal{S}'$ and $\mathcal{Q}$ as follows:

- for each $k$ with $1 \leq k \leq b$: ($S_k$ does not contain $u$)
  Then put $S'_k = S_k$, and make $Q_k$ equal to $P_k$. Since by assumption $P_k$ describes $S_k$, we know that $Q_k$ describes $S'_k$. We also know that $S_k$ has vertices not in $B_i$, and so does $S'_k$. Thus Definitions 16.1 and 16.2 are satisfied by $Q_k$ and $S_k$. Moreover, $Q_k$ satisfies the recurrence.
- for each $k$ with $b + 1 \leq p \leq t$: ($S_k$ contains $u$)
  Then put $S'_k = S_k \backslash \{u\}$, and define $Q_k$ so that it describes $S'_k$ in order to satisfy Definition 16.1. Since $S'_k$ has vertices outside $B_i$, $S_k$ satisfies Definition 16.2. We want to show that $Q_k$ satisfies the recurrence.
  We have $Q_k[B_j] = S'_k \cap B_j = (S_k \cap B_i) \backslash \{u\} = P_k[B_i] \backslash \{u\}$ as in the recurrence. Let $v \in Q_k[B_j]$. Since $u$ has no neighbor in $V(T_i) \backslash B_i$, the distance between $u$ and $v$ in $S_k$ could be 3 or more, or uses only vertices in $P_k[B_i]$, so $P_k[u, v] = d$ as in the recurrence is correct. Let $v, w \in Q_k[B_j]$. The distance between $v$ and $w$ in $S_k$ is either the same as in $S'_k$, i.e. it is $Q_k[v, w]$, or the addition of $u$ changes this distance, in which case we take the shortest path in $G[P_k[B_i]]$. It follows that $P_k[v, w]$ is defined as in the recurrence.
  Finally, consider $P_k[out]$. Since a vertex $z \in V(T_i) \backslash B_i$ has the same neighborhood in either $B_i$ or $B_j$, it follows that $P_k[out] = Q_k[out]$, as in the recurrence. Therefore, $Q_k$ satisfies all the recurrence conditions.
- for each $k$ with $t + 1 \leq k \leq s$: ($S_k$ is a 2-club containing $u$ but is not a subset of $B_i$)
  Put $S'_k = S_k \backslash \{u\}$, and define $Q_k$ so that it describes $S'_k$ in order to satisfy Definition 16.1. Since $S_k$ is not a subset of $B_i$, it contains vertices in $V(T_i) \backslash B_i$. Then so does $S'_k$, and Definition 16.2 is satisfied.
  Since $S_k$ is a 2-club and $k > t$, it is easy to see in this case that all the conditions in the recurrence must be satisfied.
- for each $k$ with $s + 1 \leq k \leq s + p$: ($S_k$ contains $u$ and $S_k \subseteq B_i$)
  Define $\{R_1, \ldots, R_p\} = \{S_{s+1}, \ldots, S_{s+p}\}$ for later reference. These do not have any correspondent in $\mathcal{S}'$ or $\mathcal{Q}$.
- for each $k$ with $s + p + 1 \leq k \leq h$: ($S_k$ is a 2-club not containing $u$)
  Then append $S_k$ to $\mathcal{S}'$.

Note that $\mathcal{S}'$ has $h' = h - p$ partial 2-clubs since the only 2-clubs of $\mathcal{S}$ without a correspondent in $\mathcal{S}'$ are the $R_k$ 2-clubs. For the same reason, $\mathcal{S}'$ covers $A_j$ as we have defined it. Thus Definition 16.4 holds on $\mathcal{Q}$ and $\mathcal{S}'$. The above construction shows that $\mathcal{S}'$ and $\mathcal{Q}$ satisfy Definitions 16.1 and 16.2. It is also clear that Definition 16.3 is satisfied with $\mathcal{Q}$ and $\mathcal{S}'$. Therefore, $C[\mathcal{Q}, A_j, h'] = 1$.

The only requirement of the recurrence not demonstrated to hold is that concerning $A_i$, which must be a subset of

$$Y := A_j \cup \left( \bigcup_{k=1}^{t} P_k \right) \cup \left( \bigcup_{k=t+1}^{s} (Q_k[B_j] \cup \{u\}) \right) \cup \left( \bigcup_{k=1}^{p} R_k \right)$$

$$= A_j \cup \left( \bigcup_{k=1}^{b} Q_k[B_j] \right) \cup \left( \bigcup_{k=b+1}^{s} Q_k[B_j] \cup \{u\} \right) \cup \left( \bigcup_{k=1}^{p} R_k \right)$$

Assume that there exists $w \in A_i \backslash Y$. Then $w \notin A_j$ and $w \notin R_1, \ldots, R_p$. Recall that we defined $A_j = A_i \backslash (\{u\} \cup R_1 \cup \ldots \cup R_p)$. This implies that $w = u$. In turn, this implies that $b = t = s$ (otherwise, if $b < t$, there would be $P_{b+1}[B_i] = Q_{b+1}[B_j] \cup \{u\}$ in $Y$, and if $s > t$, there would be $Q_{t+1}[B_i] \cup \{u\}$ in $Y$, thereby covering $u$). This also implies that $p = 0$, i.e. there is no $R_k$ 2-club, as otherwise they would cover $u$. Thus the partial 2-clubs of $\mathcal{S}$ are $S_1, \ldots, S_b, S_{s+p+1}, \ldots, S_h$, none of which covers $w = u$. This contradicts the fact that $\mathcal{S}$ satisfies Definition 16.4, and thus $w$ cannot exist. We have thus shown that all recurrence conditions are met.

We therefore have $C[\mathcal{Q}, A_j, h'] = 1$. Moreover, all recurrence conditions are met, so it sets $C[\mathcal{P}, A_i, h]$ to 1.

### 6.2.5 Forget Vertex

Assume that $B_i$ is a forget vertex, and that $B_j$ is the child of $B_i$ in $T$, with $u \in B_j \backslash B_i$.

Assume that the elements of $\mathcal{S}$ are ordered as $\mathcal{S} = \{S_1, \ldots, S_h\}$ so that $S_1, \ldots, S_t$ are described by $\mathcal{P}$ and $S_{t+1}, \ldots, S_h$ are 2-clubs (this ordering is possible since $\mathcal{S}$ satisfies Definition 16). Also order $S_1, \ldots, S_t$ so that $S_1, \ldots, S_s$ have vertices in $V(T_i) \backslash (B_i \cup \{u\})$, and $S_{s+1}, \ldots, S_t \subseteq B_i \cup \{u\}$.

Consider the set of partial 2-clubs $\mathcal{S}' = \{S_1, \ldots, S_s, S_{t+1}, \ldots, S_h\}$ at $B_j$. Let $h'$ be such that $h = h' + (t - s)$, noting that $|\mathcal{S}'| = h'$. Moreover, let $\mathcal{Q} = \{Q_1, \ldots, Q_s\}$ be the set of succinct partial 2-clubs at $B_j$ that describe $S_1, \ldots, S_s$. Let $A_j = A_i \cup \{u\}$ if $s = t$ and no element of $S_1, \ldots, S_t$ contains $u$, and let $A_j = A_i \backslash (S_{s+1} \cup \ldots \cup S_t)$ otherwise. We argue that $C[\mathcal{Q}, A_j, h'] = 1$ and that the recurrence is satisfied.

We note that $\mathcal{Q}$ and $\mathcal{S}'$ satisfy Definition 16.1 since we just constructed $\mathcal{Q}$ so that they describe $S_1, \ldots, S_k$. Definition 16.2 is satisfied by $\mathcal{Q}$ and $\mathcal{S}'$ since $S_1, \ldots, S_s$ are chosen to have vertices in $V(T_i) \backslash (B_i \cup \{u\}) = V(T_j) \backslash B_j$. Definition 16.3 is satisfied since $S_{t+1}, \ldots, S_h$ are 2-clubs. Finally, Definition 16.4 is satisfied: if $A_j = A_i \cup \{u\}$, then this case occurs when $s = t$ and thus $\mathcal{S}' = \mathcal{S}$. In that situation, $\mathcal{S}$ covers $A_i \cup \{u\}$ by Definition 16.3, and thus $\mathcal{S}'$ covers $A_j$. Otherwise, $A_j = A_i \backslash (S_{s+1}, \ldots, S_t)$. Since $\mathcal{S}$ covers $A_i \cup V(T_i) \backslash B_i$ by assumption, $\mathcal{S}'$ covers $A_j$ since $\mathcal{S}' = \mathcal{S} \backslash \{S_{s+1}, \ldots, S_t\}$. Thus $\mathcal{Q}$ and $\mathcal{S}'$ satisfy Definition 16 and by induction, $C[\mathcal{Q}, A_j, h'] = 1$.

Let us show that the requirements of the recurrence are met to have

$$C[\mathcal{P}, A_i, h] = 1.$$

Consider $k$ with $1 \leq k \leq s$. Note that $S_k = S'_k$. Assume that $u \notin S_k$. Then $Q_k$ and $P_k$ describe the same partial 2-club and must be equal, as in the recurrence. Assume instead that $u \in S_k$. Then $P_k = S_k \cap B_i = (S_k \cap B_j) \backslash \{u\} = Q_k[B_j] \backslash \{u\}$ as in the recurrence. For each $v \in P_k[B_i]$, it is clear that $d_{G[S_k]}(u, v) \leq 2$ by the definition of a partial 2-club, and thus $Q_k[u, v] \leq 2$. For $v, w \in P_k[B_i]$, we must have $P_k[v, w] = Q_k[v, w]$ since they both describe $S_k$. Consider $P_k[out]$ and $Q_k[out]$.

Since $u \in B_j \setminus B_i$, it follows that if $Z \in P_k[out]$, then $Z \cup \{u\} \in Q_k[out]$ and that $N(u) \cap P_k[B_i]$ is in $P_k[out]$ and not in $Q_k[out]$.

The value of $A_j$ is set here as in the recurrence, as well as $h'$. We therefore conclude that $C[\mathcal{P}, A_i, h] = 1$.

### 6.2.6 Join Vertex

Assume that $B_i$ is a join vertex with children $B_r$ and $B_l$. Let $\mathcal{S}^l \subseteq \mathcal{S}$ be the subset of partial 2-clubs that intersect with $V(T_l)\setminus B_i$ or that are subsets of $B_i$, and let $\mathcal{S}^r \subseteq \mathcal{S}$ be the subset of partial 2-clubs that intersect with $V(T_r)\setminus B_i$ (note the difference between $\mathcal{S}^l$ and $\mathcal{S}^r$, i.e. that $\mathcal{S}^r$ does not have partial 2-clubs that are subsets of $B_i$, and that $\mathcal{S} = \mathcal{S}^l \cup \mathcal{S}^r$). Denote $h_l = |\mathcal{S}^l|$ and $h_r = |\mathcal{S}^r|$.

Let $b$ be the number of partial 2-clubs of $\mathcal{S}$ that are in both $\mathcal{S}^l$ and $\mathcal{S}^r$, and let $a$ be the number of such partial 2-clubs that are described by some $P_k \in \mathcal{P}$. Assume without loss of generality that $\mathcal{S}^l = \{S^l_1, \dots S^l_{h_l}\}$ and $\mathcal{S}^r = \{S^r_1, \dots, S^r_{h_r}\}$ are labeled so that the following holds:

(1) $S^l_k = S^r_k$ for each $1 \leq k \leq b$.
(2) $P_{k-a}$ describes $S^l_k = S^r_k$ for each $a+1 \leq k \leq b$. For later reference, note that since no entry of $\mathcal{P}$ describes $S^l_1, \dots, S^l_a$ and since $\mathcal{S}$ satisfies Definition 16.3, we know that $S^l_1, \dots, S^l_a$ are actual 2-clubs.
(3) there is an integer $s$ such that entries $S^l_{b+1}, \dots, S^l_s$ are described by some $P_k$ entry, and $S^l_{s+1}, \dots, S^l_{h_l}$ are not. Assume further that $P_{k-a}$ describes $S^l_k$ for each $b+1 \leq k \leq s$.
(4) there is an integer $q$ such that entries $S^r_{b+1}, \dots, S^r_q$ are described by some $P_k$ entry, and $S^r_{q+1}, \dots, S^r_{h_l}$ are not. Assume further that $P_{k-a+(s-b)}$ describes $S^r_k$ for each $b+1 \leq k \leq q$.

Note that since Definition 16.3 holds, $S^l_{s+1}, \dots, S^l_{h_l}, S^r_{q+1}, \dots, S^r_{h_r}$ are 2-clubs because no entry of $\mathcal{P}$ describes them. Also, summing cases (2), (3), (4), we note that $t = (b-a) + (s-b) + (q-b) = q - a + s - b$, as in the recurrence.

Also notice that for each $S^l_k \in \mathcal{S}^l$, $S^l_k \cap V(T_l)$ is a partial 2-club at $B_l$. This is because by the properties of a tree decomposition, vertices of $(S^l_k \cap V(T_l))\setminus B_l$ have distance at most 2 from each other, and distance at most 2 to vertices of $S^k_l \cap B_l$, whether the vertices of $V(T_r)\setminus B_i$ are present or not. By the same argument, for each $S^r_k \in \mathcal{S}^r$, $S^r_k \cap V(T_r)$ is a partial 2-club.

Define

$$\mathcal{S}^{l*} = \{S^l_1 \cap V(T_l), \dots, S^l_{h_l} \cap V(T_l)\} \quad \text{and}$$
$$\mathcal{S}^{r*} = \{S^r_1 \cap V(T_r), \dots, S^r_{h_r} \cap V(T_r)\}$$

which are respectively partial 2-clubs at $B_l$ and $B_r$. Our goal is to show that $C[\mathcal{L}, A_l, h_l] = C[\mathcal{R}, A_r, h_r] = 1$ for some $\mathcal{L}$ and $\mathcal{R}$, and that all requirements of the recurrence are met to have $C[\mathcal{P}, A_i, h] = 1$. Here, $A_l$ and $A_r$ are defined as

$$A_l = A_i \setminus (S^r_{b+1} \cup \dots \cup S^r_{h_r})$$

$$A_r = A_i \backslash A_l$$

We note that $A_i = A_l \cup A_r$ as in the recurrence.

Now, consider $\mathcal{L} = \{L_1, \ldots, L_s\}$ such that $L_k$ describes $S_k^l \cap V(T_l)$ for each $1 \le k \le s$, and $\mathcal{R} = \{R_1, \ldots, R_q\}$ such that $R_k$ describes $S_k^r \cap V(T_r)$ for each $1 \le k \le q$. Definition 16.1 is obviously satisfied for $\mathcal{L}$ and $\mathcal{R}$.

Let us argue that Definition 16.2 holds for $\mathcal{L}$ and $\mathcal{S}^{l*}$, and for $\mathcal{R}$ and $\mathcal{S}^{r*}$. Let $S_k^l \in \mathcal{S}^l$ with $1 \le k \le s$. We must show that $S_k^l \cap V(T_l)$ has vertices in $V(T_l) \backslash B_l$. First consider $k$ with $1 \le k \le b$. Recall that $S_k^l = S_k^r$, as described by Point (1) of the Joint Vertex proof. Also recall that $\mathcal{S}^r$ only contains partial 2-clubs that intersect with $V(T_r) \backslash B_i$, and hence $S_k^l \cap (V(T_r) \backslash B_i) \neq \emptyset$. Moreover, $\mathcal{S}^l$ only contains partial 2-clubs that either intersect with $V(T_l) \backslash B_i$, or that are subsets of $B_i$. We just argued that $S_k^l$ is not a subset of $B_i$, so it must be the case that $S_k^l$ intersects with $V(T_l) \backslash B_i$. It follows that $S_k^l \cap V(T_l)$ also intersects with $V(T_l) \backslash B_i$, as desired. Also note that $S_k^r$ intersects with $V(T_r) \backslash B_i$, by the definition of $\mathcal{S}^r$.

Now, consider $S_k^l \in \mathcal{S}^k$, with $b + 1 \le k \le s$. As described by (3) above, $S_k^l$ is described by $P_{k-a}$. Since $\mathcal{S}$ satisfies Definition 16.2, $S_k^l$ has vertices in $V(T_i) \backslash B_i$. Moreover, when $b + 1 \le k \le s$, $S_k^l$ is not in $S^r$, so it has no vertices in $V(T_r) \backslash B_r$. It follows that $S_k^l \cap V(T_l)$ has vertices in $V(T_l) \backslash B_l$. For $k$ with $b + 1 \le k \le q$, we may argue in the same manner that $S_k^r \cap V(T_r)$ has vertices in $V(T_r) \backslash B_r$. Therefore, Definition 16.2 holds for $\mathcal{L}$ and $\mathcal{S}^{l*}$, and for $\mathcal{R}$ and $\mathcal{S}^{r*}$.

We next consider Definition 16.3. We have already argued that $S_{s+1}^l, \ldots, S_{h_l}^l$ are 2-clubs, but we must argue that $S_{s+1}^l \cap V(T_l), \ldots, S_{h_l}^l \cap V(T_l)$ are also 2-clubs. This follows from the fact that only $S_1^l, \ldots, S_b^l$ have vertices in $V(T_r) \backslash B_i$, and thus that $S_k^l \cap V(T_l) = S_k^l$ for each $s + 1 \le k \le h_l$. Therefore, $\mathcal{L}$ and $\mathcal{S}^{l*}$ satisfy Definition 16.3. by the exact same reasoning, $\mathcal{R}$ and $\mathcal{S}^{r*}$ satisfy Definition 16.3.

We now turn to Definition 16.4. Since by assumption $\mathcal{S}$ covers $V(T_i) \backslash B_i$, $\mathcal{S}^{l*}$ covers $V(T_l) \backslash B_l$. Now assume that $\mathcal{S}^{l*}$ does not cover some $u \in A_l$. Since $\mathcal{S}$ covers $A_i$, $\mathcal{S}$ contains a partial 2-club $S'$ with $u \in S'$. Because $S_k^l \cap V(T_l) \cap B_i = S_k^l \cap B_i$ for each $S_k^l \in \mathcal{S}^l$, $S' \notin \mathcal{S}^l$ as otherwise $u$ would be covered. Thus, $S' \in \mathcal{S}^r \backslash \mathcal{S}^l$, which is equal to $S_{b+1}^r \cup \ldots \cup S_{h_r}^r$. But then note that $u \in A_l = A_i \backslash \{S_{b+1}^r \cup \ldots \cup S_{h_r}^r\}$, a contradiction. Hence, Definition 16.4 is satisfied by $\mathcal{L}$ and $\mathcal{S}^{l*}$. Consider now $\mathcal{R}$ and $\mathcal{S}^{r*}$. We know that $\mathcal{S}^{r*}$ covers $V(T_r) \backslash B_i$. Let $u \in A_r$. Then $u \in A_i \backslash A_l = A_i \cap (S_{b+1}^r \cup \ldots \cup S_{h_r}^r)$. Since $S_k^r \cap V(T_r) \in \mathcal{S}^{r*}$ for each $1 \le k \le h_r$, it follows that $\mathcal{S}^{r*}$ covers $u$. Therefore, Definition 16.4 is also satisfied by $\mathcal{R}$ and $\mathcal{S}^{r*}$.

We have thus shown that Definition 16 is satisfied by $\mathcal{L}$ and $\mathcal{S}^{l*}$, and by $\mathcal{R}$ and $\mathcal{S}^{r*}$. It follows that $C[\mathcal{L}, A_l, h_l] = C[\mathcal{R}, A_r, h_r] = 1$. It remains to show that all requirements of the recurrence are met to have $C[\mathcal{P}, A_i, h] = 1$.

We have already argued that $t = (s - b) + (q - b) + (b - a)$. For each $k$ with $1 \le k \le a$, $S_k^l = S_k^r = (S_k^l \cap V(T_l)) \cup (S_k^r \cap V(T_r))$ is a 2-club. In that case, $L_k[B_l] = R_k[B_l]$, as desired. Since merging the partial 2-clubs described by $L_k$ and $R_k$ forms a 2-club, it is not hard to see that the remaining elements of the recurrence must hold, so that all distances are at most 2 after merging.

For each $k$ with $a+1 \le k \le b$, $S_k^l = S_k^r = (S_k^l \cap V(T_l)) \cup (S_k^r \cap V(T_r))$ is a partial 2-club which, by construction, is described by $P_{k-a}$. Thus $P_{k-a}[B_i] = L_k[B_l] = R_k[B_r]$. Moreover, since merging the partial 2-clubs described by $L_k$ and $R_k$ forms a partial 2-club, it is not hard to see that the remaining elements of the recurrence must hold (in particular, $P_{k-a}[u, v] = \min(L_k[u, v], R_k[u, v])$ follows from the properties of tree decomposition.

For each $k$ with $b+1 \le k \le s$, $P_{k-a}$ and $L_k$ describe the same partial 2-club $S_k^l$, and for each $k$ with $b+1 \le k \le q$, $P_{k-a+(s-b)}$ and $R_k$ describe the same partial 2-club $S_k^r$, as in the recurrence.

Finally, $h = h_l + h_r - b$ since $\mathcal{S}^l$ and $\mathcal{S}^r$ have exactly $b$ partial 2-clubs in common, and $A_i = A_l \cup A_r$ was argued above.

All requirements of the recurrence are satisfied, and therefore $C[\mathcal{P}, A_i, h] = 1$. □

Even though the recurrence is shown to be correct, we have not discussed the bounds on $|\mathcal{P}|$ to be considered yet. The recurrence assumes that, for the children of a given bag $B_i$, we have access to an unbounded number of $\mathcal{P}$ entries in the children, whereas we would like to store a limited number of such entries. Specifically, for we would like to consider only the succinct partial 2-club of size at most $\delta + 1$. Consider the following algorithm.

**Algorithm 1:** Main algorithm on the treewidth decomposition.

```
1 for each bag B_i of T in a postorder traversal do
2     Initialize a map C* whose purpose is to store the recurrence entries
3     for each P of cardinality at most δ + 1, each A_i ⊆ B_i, and each h ∈ {0, ..., n} do
4         Compute and store C[P, A_i, h] using the recurrence of Lemma 18 and the C entries of
          the child or children of B_i
```

The main difference between Algorithm 1 and the recurrence of Lemma 18 is that in the algorithm, we only have access to the succinct partial 2-clubs of size at most $\delta+1$ when using the $C$ entries of the child or children of $B_i$. More specifically, denote by $C^*[\mathcal{P}, A_i, h]$ the value computed by the algorithm at bag $B_i$ on $\mathcal{P}$, $A_i$ and $h$ (we name it $C^*$ to distinguish it from the true value of $C[\mathcal{P}, A_i, h]$ as defined in Definition 16). First, notice that if $C^*[\mathcal{P}, A_i, h] = 1$, then the recurrence proof constructs an actual solution, and it follows that $C[\mathcal{P}, A_i, h] = 1$. The converse may not hold: since the algorithm has access to a limited number of entries in the children of $B_i$, it is possible that $C^*[\mathcal{P}, A_i, h] = 0$ whereas we would have found $C[\mathcal{P}, A_i, h] = 1$ if we had stored larger succinct partial 2-clubs at the children of $B_i$. Nevertheless, we show that $C[\emptyset, B_R, h] = 1$ at the root $B_R$ for the optimal value $h$. We consider this aspect in the following lemma.

**Lemma 19** *For each $\mathcal{P}, A_i, h$ triple, denote by $C^*[\mathcal{P}, A_i, h]$ be the value computed by Algorithm 1 on this triple. Then the following holds:*

- *if $C^*[\mathcal{P}, A_i, h] = 1$, then $C[\mathcal{P}, A_i, h] = 1$.*
- *Assume that $\mathcal{S}$ is an optimal 2-club cover of $G$ that contains $h$ 2-clubs. Then $C^*[\emptyset, B_R, h] = 1$.*

***Proof*** The fact that $C^*[\mathcal{P}, A_i, h] = 1$ implies $C[\mathcal{P}, A_i, h] = 1$ can be proved by inductively on $T$. If $B_i$ is a leaf, the statement is easy to verify. So assume that the statement holds for every child of $B_i$. Suppose that $C^*[\mathcal{P}, A_i, h] = 1$. Assume that $B_i$ is an introduce node with child $B_j$. Then there is some entry $C^*[\mathcal{Q}, A_j, h'] = 1$ satisfying all properties of the recurrence of Lemma 18. By induction, $C[\mathcal{Q}, A_j, h'] = 1$ as well and also satisfies the recurrence, meaning that $C[\mathcal{P}, A_i, h] = 1$. The idea is the same if $B_i$ is a forget or join node. This proves the first point.

Now, let $\mathcal{S}$ be an optimal 2-club cover of $G$. For a bag $B_i$, let $X_i$ be the set of 2-clubs of $\mathcal{S}$ that have vertices in both $V(T_i)\backslash B_i$ and in $V(G)\backslash V(T_i)$. By Lemma 15, we may assume that $|X_i| \leq \delta + 1$. Let $\mathcal{P}_i$ be the set of succinct partial 2-clubs at $B_i$ corresponding to $X_i$. Let $\mathcal{S}_i$ be the set of 2-clubs of $\mathcal{S}$ that are either in $\mathcal{P}_i$, or that have all their vertices in $V(T_i)$, and let $h_i = |\mathcal{S}_i|$. Finally, let $A_i$ be the vertices of $B_i$ that belong to some 2-club of $\mathcal{S}_i$. One can see, also by induction, that $C^*[\mathcal{P}_i, A_i, h_i] = 1$ for each $B_i$. Indeed, for a leaf $B_i = \{u\}$, we have $\mathcal{P}_i = \emptyset$ and $C^*[\emptyset, \emptyset, h] = C^*[\emptyset, \{u\}, h] = 1$ for all $h$. Consider an internal bag $B_i$. If $B_i$ is an introduce vertex with child $B_j$, then by induction $C^*[\mathcal{P}_j, A_j, h_j] = 1$. The recurrence is able to reconstruct solution $\mathcal{S}_i$ from $\mathcal{S}_j$, and thus $C^*[\mathcal{P}_j, A_j, h_j]$ can be used to obtain $C^*[\mathcal{P}_i, A_i, h_i] = 1$. The same argument holds if $B_i$ is a forget vertex with child $B_j$, and similarly, if $B_i$ is a join vertex with children $B_l, B_r$, the recurrence is able to reconstruct $\mathcal{S}_i$ from $\mathcal{S}_l, \mathcal{S}_r$, given that $C^*[\mathcal{P}_l, A_l, h_l] = C^*[\mathcal{P}_r, A_r, h_r] = 1$. □

We can conclude with the following result.

**Theorem 20** *A solution of Min 2-Club Cover on a graph $G$ having treewidth bounded by $\delta$ can be computed in time $2^{O(\delta 2^{\delta+1})}n^4$.*

***Proof*** We first argue that returning the smallest $h$ such that $C^*[\emptyset, B_R, h] = 1$, where $C^*$ is the table constructed by Algorithm 1, is correct. Suppose that $\mathcal{S}$ is an optimal 2-club cover of $G$ with $h = |\mathcal{S}|$. By Lemma 17, $C[\emptyset, B_R, h] = 1$ and, for any $h' < h$, $C[\emptyset, B_R, h'] = 0$. By the second point of Lemma 19, we have $C^*[\emptyset, B_R, h] = 1$. Moreover, Lemma 19 also implies that, for any $h' < h$, $C^*[\emptyset, B_R, h'] = 0$, as otherwise the the first point of the lemma would imply $C[\emptyset, B_R, h'] = 1$, a contradiction. This proves the correctness.

Lemma 19 implies that it is sufficient to compute $C[\mathcal{P}_i, A_i, h]$ for only entries in which $|\mathcal{P}_i| \leq \delta + 1$ for each bag $B_i$. Since there are at most $2^{4 \cdot 2^{\delta+1}}$ possible partial 2-clubs at $B_i$, which includes the empty partial 2-club, the number of ways to form $\mathcal{P}$ is bounded by $(2^{4 \cdot 2^{\delta+1}})^{\delta+1}$, which is $2^{O(\delta 2^{\delta+1})}$. Moreover, the number of possible $A_i$ subsets is at most $2^{\delta+1}$ and the number of possible $h$ values is at most $n$. Therefore, we need to compute at most $2^{O(\delta 2^{\delta+1})} \cdot 2^{\delta+1} \cdot n$ entries, which is $n \cdot 2^{O(\delta 2^{\delta+1})}$.

To compute a specific entry $C[\mathcal{P}, A_i, h]$, in the worst case $B_i$ is a join vertex and we must consider all the $(n2^{O(\delta 2^{\delta+1})})^2$ possible entries for $C[\mathcal{L}, A_l, h_l]$ and $C[\mathcal{R}, A_r, h_r]$ for the children $B_l$ and $B_r$, where $\mathcal{L}$ ($\mathcal{R}$, respectively) is a multi-set of partial 2-clubs at $B_l$ ($B_r$, respectively); the number of such entries is $n^2 2^{O(\delta 2^{\delta+1})}$. Furthermore, we need to find a matching ordering of $\mathcal{P}$, $\mathcal{L}$ and $\mathcal{R}$ (that is a correspondence between partial 2-clubs of $\mathcal{P}$, $\mathcal{L}$ and $\mathcal{R}$), which requires testing all the $((\delta + 1)!)^3$ permutations of the three sets.

Consider the time required to check each condition of the recurrence, ignoring the condition on finding the 2-clubs $R_1, \ldots, R_p$ in the introduce vertices for now. Each such condition can be verified in time $O(2^{\delta+1})$, the most time-consuming verification being to check $P[out]$ (possible neighborhoods of vertices of a succinct partial 2-clubs).

As for finding the 2-clubs $R_1, \ldots, R_p$, they must cover the uncovered elements of $A_i \subseteq B_i$. It is clear that $\delta + 1$ 2-clubs will always suffice to do so, so we can enumerate every way of obtaining at most $\delta + 1$ 2-clubs from $B_i$. There are at most $(2^{\delta+1})^{\delta+1}$ combinations of subsets to enumerate, which is $2^{O(\delta^2)}$. This is the leading term in the recurrence verification. To sum up, computing the recurrence for one specific entry takes time in

$$n^2 2^{O(\delta 2^{\delta+1})} \cdot ((\delta + 1)!)^3 \cdot 2^{O(\delta^2)}$$

which is $n^2 2^{O(\delta 2^{\delta+1})}$.

Therefore, the total spent at one particular $B_i$ is bounded by $n \cdot 2^{O(\delta 2^{\delta+1})} \cdot n^2 2^{O(\delta 2^{\delta+1})}$, which is $n^3 2^{O(\delta 2^{\delta+1})}$. As the tree decomposition has $O(n)$ vertices, the complexity result follows. □

## 7 Conclusion

We have considered the problem of covering a graph with 2-clubs, given complexity results on the problem. We have shown that the decision problem that asks whether there exists a covering of a graph with 2-clubs is W[1]-hard for parameter distance to 2-club. Moreover, for the problem that asks for a covering with minimum number of 2-clubs, on restricted graph classes, we have given negative (subcubic planar graphs, bipartite graphs) and positive (graphs of bounded treewidth) results. There are interesting open problems related to covering a graph with clubs. It would be interesting to extend some of the results for the problem of covering with $s$-clubs, with $s > 2$. For example, is it possible to extend the FPT algorithm on graphs of bounded treewidth to any $s > 2$? Moreover, the parameterized complexity of the problem has to be analyzed for other graph classes, like chordal graphs and, more generally, graphs that have a bounded distance from this class.

**Funding** Open access funding provided by Università degli studi di Bergamo within the CRUI-CARE Agreement.

## Declarations

**Conflict of interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

1. Alba, R.D.: A graph-theoretic definition of a sociometric clique. J. Math. Sociol. **3**, 113–126 (1973)
2. Asahiro, Y., Doi, Y., Miyano, E., Samizo, K., Shimizu, H.: Optimal approximation algorithms for maximum distance-bounded subgraph problems. Algorithmica **80**(6), 1834–1856 (2018)
3. Ausiello, G., Crescenzi, P., Gambosi, G., Kann, V., Marchetti-Spaccamela, A., Protasi, M.: Complexity and Approximation: Combinatorial Optimization Problems and Their Approximability Properties. Springer, Heidelberg (1999)
4. Balasundaram, B., Butenko, S., Trukhanov, S.: Novel approaches for analyzing biological networks. J. Comb. Optim. **10**(1), 23–39 (2005)
5. Bourjolly, J., Laporte, G., Pesant, G.: An exact algorithm for the maximum k-club problem in an undirected graph. Eur. J. Oper. Res. **138**(1), 21–28 (2002)
6. Cavique, L., Mendes, A.B., Santos, J.M.A.: An algorithm to discover the k-clique cover in networks. In: Lopes, L.S., Lau, N., Mariano, P., Rocha, L.M. (eds.) Progress in Artificial Intelligence, 14th Portuguese Conference on Artificial Intelligence, EPIA 2009, Aveiro, Portugal, October 12–15, 2009. Proceedings. Lecture Notes in Computer Science, vol. 5816, pp. 363–373. Springer (2009)
7. Cerioli, M.R., Faria, L., Ferreira, T.O., Martinhon, C.A.J., Protti, F., Reed, B.A.: Partition into cliques for cubic graphs: planar case, complexity and approximation. Discret. Appl. Math. **156**(12), 2270–2278 (2008)
8. Cerioli, M.R., Faria, L., Ferreira, T.O., Protti, F.: A note on maximum independent sets and minimum clique partitions in unit disk graphs and penny graphs: complexity and approximation. RAIRO Theor. Inform. Appl. **45**(3), 331–346 (2011)
9. Chang, M., Hung, L., Lin, C., Su, P.: Finding large k-clubs in undirected graphs. Computing **95**(9), 739–758 (2013)
10. Courcelle, B.: The monadic second-order logic of graphs. I. Recognizable sets of finite graphs. Inf. Comput. **85**(1), 12–75 (1990)
11. Chakraborty, D., Chandran, L.S., Padinhatteeri, S., Pillai, R.R.: Algorithms and complexity of s-club cluster vertex deletion. In: Flocchini, P., Moura, L. (eds.) Combinatorial Algorithms. IWOCA 2021. Proceedings. Lecture Notes in Computer Science, vol. 12757, pp. 152–164. Springer (2021)
12. Dondi, R., Lafond, M.: On the tractability of covering a graph with 2-clubs. In: Gasieniec, L.A., Jansson, J., Levcopoulos, C. (eds.) Fundamentals of Computation Theory—22nd International Symposium, FCT 2019, Copenhagen, Denmark, August 12–14, 2019, Proceedings. Lecture Notes in Computer Science, vol. 11651, pp. 243–257. Springer (2019)
13. Dondi, R., Mauri, G., Sikora, F., Zoppis, I.: Covering a graph with clubs. J. Graph Algorithms Appl. **23**(2), 271–292 (2019)
14. Dondi, R., Mauri, G., Zoppis, I.: On the tractability of finding disjoint clubs in a network. Theoret. Comput. Sci. **777**, 243–251 (2019)
15. Dorndorf, U., Jaehn, F., Pesch, E.: Modelling robust flight-gate scheduling as a clique partitioning problem. Transp. Sci. **42**(3), 292–301 (2008)
16. Dumitrescu, A., Pach, J.: Minimum clique partition in unit disk graphs. Graphs Comb. **27**(3), 399–411 (2011)
17. Fellows, M.R., Hermelin, D., Rosamond, F.A., Vialette, S.: On the parameterized complexity of multiple-interval graph problems. Theoret. Comput. Sci. **410**(1), 53–61 (2009)
18. Figiel, A., Himmel, A., Nichterlein, A., Niedermeier, R.: On 2-clubs in graph-based data clustering: theory and algorithm engineering. In: Calamoneri, T., Corò, F. (eds.) Algorithms and Complexity—12th International Conference, CIAC 2021, Virtual Event, May 10–12, 2021, Proceedings. Lecture Notes in Computer Science, vol. 12701, pp. 216–230. Springer (2021)
19. Garey, M.R., Johnson, D.S., Stockmeyer, L.J.: Some simplified NP-complete graph problems. Theor. Comput. Sci. **1**(3), 237–267 (1976)

20. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness (1979)
21. Golovach, P.A., Heggernes, P., Kratsch, D., Rafiey, A.: Finding clubs in graph classes. Discret. Appl. Math. **174**, 57–65 (2014)
22. Gramm, J., Guo, J., Hüffner, F., Niedermeier, R.: Data reduction and exact algorithms for clique cover. ACM J. Exp. Algorithmics **13** (2008)
23. Grbić, M., Kartelj, A., Janković, S., Matić, D., Filipović, V.: Variable neighborhood search for partitioning sparse biological networks into the maximum edge-weighted $k$k-plexes. IEEE/ACM Trans. Comput. Biol. Bioinf. **17**(5), 1822–1831 (2020)
24. Guo, J., Komusiewicz, C., Niedermeier, R., Uhlmann, J.: A more relaxed model for graph-based data clustering: s-plex cluster editing. SIAM J. Discret. Math. **24**(4), 1662–1683 (2010)
25. Hartung, S., Komusiewicz, C., Nichterlein, A.: Parameterized algorithmics and computational experiments for finding 2-clubs. J. Graph Algorithms Appl. **19**(1), 155–190 (2015)
26. Karp, R.M.: Reducibility among combinatorial problems. In: Miller, R.E., Thatcher, J.W. (eds.) 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, pp. 85–103. The IBM Research Symposia Series, Plenum Press, New York (1972)
27. Kloks, T.: Treewidth, Computations and Approximations. Lecture Notes in Computer Science, vol. 842. Springer, Berlin (1994)
28. Kochenberger, G.A., Glover, F.W., Alidaee, B., Wang, H.: Clustering of microarray data via clique partitioning. J. Comb. Optim. **10**(1), 77–92 (2005)
29. Komusiewicz, C.: Multivariate algorithmics for finding cohesive subnetworks. Algorithms **9**(1), 21 (2016)
30. Komusiewicz, C., Sorge, M.: An algorithmic framework for fixed-cardinality optimization in sparse graphs applied to dense subgraph problems. Discret. Appl. Math. **193**, 145–161 (2015)
31. Laan, S., Marx, M., Mokken, R.J.: Close communities in social networks: boroughs and 2-clubs. Soc. Netw. Anal. Min. **6**(1), 20:1-20:16 (2016)
32. Liu, H., Zhang, P., Zhu, D.: On editing graphs into 2-club clusters. In: Snoeyink, J., Lu, P., Su, K., Wang, L. (eds.) Frontiers in Algorithmics and Algorithmic Aspects in Information and Management—Joint International Conference, FAW-AAIM 2012, Beijing, China, May 14–16, 2012. Proceedings. Lecture Notes in Computer Science, vol. 7285, pp. 235–246. Springer (2012)
33. Lund, C., Yannakakis, M.: On the hardness of approximating minimization problems. J. ACM **41**(5), 960–981 (1994)
34. Mokken, R.: Cliques, clubs and clans. Qual. Quant. Int. J. Methodol. **13**(2), 161–173 (1979)
35. Mokken, R.J., Heemskerk, E.M., Laan, S.: Close communication and 2-clubs in corporate networks: Europe 2010. Soc. Netw. Anal. Min. **6**(1), 40:1-40:19 (2016)
36. Mujuni, E., Rosamond, F.A.: Parameterized complexity of the clique partition problem. In: Harland, J., Manyem, P. (eds.) Theory of Computing 2008. Proc. Fourteenth Computing: The Australasian Theory Symposium (CATS 2008), Wollongong, NSW, Australia, January 22–25, 2008. Proceedings. CRPIT, vol. 77, pp. 75–78. Australian Computer Society (2008)
37. Nelson, J.: A note on set cover inapproximability independent of universe size. Electron. Colloq. Comput. Complex. (ECCC) **14**(105) (2007)
38. Pasupuleti, S.: Detection of protein complexes in protein interaction networks using n-clubs. In: Marchiori, E., Moore, J.H. (eds.) Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics, 6th European Conference, EvoBIO 2008, Naples, Italy, March 26–28, 2008. Proceedings. Lecture Notes in Computer Science, vol. 4973, pp. 153–164. Springer (2008)
39. Paz, A., Moran, S.: Non deterministic polynomial optimization problems and their approximations. Theoret. Comput. Sci. **15**, 251–277 (1981)
40. Pirwani, I.A., Salavatipour, M.R.: A weakly robust PTAS for minimum clique partition in unit disk graphs. Algorithmica **62**(3–4), 1050–1072 (2012)
41. Schäfer, A.: Exact algorithms for s-club finding and related problems. Ph.D. thesis, Friedrich-Schiller-University Jena (2009)
42. Schäfer, A., Komusiewicz, C., Moser, H., Niedermeier, R.: Parameterized computational complexity of finding small-diameter subgraphs. Optim. Lett. **6**(5), 883–891 (2012)
43. Wu, J., Yin, M.: Local search for diversified top-k s-plex search problem (student abstract). In: Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances

in Artificial Intelligence, EAAI 2021, Virtual Event, February 2–9, 2021. pp. 15929–15930. AAAI Press (2021)
44. Zoppis, I., Dondi, R., Santoro, E., Castelnuovo, G., Sicurello, F., Mauri, G.: Optimizing social interaction—a computational approach to support patient engagement. In: Zwiggelaar, R., Gamboa, H., Fred, A.L.N., i Badia, S.B. (eds.) 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. pp. 651–657. SciTePress (2018)
45. Zou, P., Li, H., Wang, W., Xin, C., Zhu, B.: Finding disjoint dense clubs in a social network. Theoret. Comput. Sci. **734**, 15–23 (2018)
46. Zuckerman, D.: Linear degree extractors and the inapproximability of max clique and chromatic number. Theory Comput. **3**(1), 103–128 (2007)