On Structural Parameterizations for the 2-Club Problem

Sepp Hartung, Christian Komusiewicz, and André Nichterlein

Institut für Softwaretechnik und Theoretische Informatik, TU Berlin,
Berlin, Germany
{sepp.hartung, christian.komusiewicz, andre.nichterlein}@tu-berlin.de

Abstract. The NP-hard 2-Club problem is, given an undirected graph $G = (V, E)$ and a positive integer ℓ, to decide whether there is a vertex set of size at least ℓ that induces a subgraph of diameter at most two. We make progress towards a systematic classification of the complexity of 2-Club with respect to structural parameterizations of the input graph. Specifically, we show NP-hardness of 2-Club on graphs that become bipartite by deleting one vertex, on graphs that can be covered by three cliques, and on graphs with domination number two and diameter three. Moreover, we present an algorithm that solves 2-Club in $|V|^{f(k)}$ time, where k is the so-called h-index of the input graph. By showing $W[1]$-hardness for this parameter, we provide evidence that the above algorithm cannot be improved to a fixed-parameter algorithm. This also implies $W[1]$-hardness with respect to the degeneracy of the input graph. Finally, we show that 2-Club is fixed-parameter tractable with respect to "distance to co-cluster graphs" and "distance to cluster graphs".

1 Introduction

In the analysis of social and biological networks, one important task is to find cohesive subnetworks since these could represent communities or functional subnetworks within the large network. There are several graph-theoretic formulations for modeling these cohesiveness demands such as s-cliques [1], s-plexes [21], and s-clubs [15]. In this work, we study the problem of finding large s-clubs within the input network. An s-club is a vertex set that induces a subgraph of diameter at most s. Thus it is a distance-based relaxation of complete graphs, cliques, which are exactly the graphs of diameter one. For constant $s \geq 1$, the problem is defined as follows.

s-Club
Input: An undirected graph $G = (V, E)$ and an integer $\ell \geq 1$.
Question: Is there a vertex set $S \subseteq V$ of size at least ℓ such that $G[S]$ has diameter at most s?

In this work, we focus on studying the computational complexity of 2-Club. This is motivated by the following two considerations. First, 2-Club is an important special case concerning the applications: For biological networks, 2-clubs and 3-clubs have been identified as the most reasonable diameter-relaxations of cliques [18]. 2-Club also has applications in the analysis of social networks [14]. Consequently, experimental evaluations concentrate on finding 2-clubs and 3-clubs [13]. Second, 2-Club is the most basic variant of s-Club that is different from Clique. For example, being a clique is a hereditary graph property, that is, it is closed under vertex deletion. In contrast, being a 2-club is not hereditary, since deleting vertices can increase the diameter of a graph. Hence, it is interesting to spot differences in the computational complexity of the two problems.

In the spirit of multivariate algorithmics [11, 17], we aim to describe how structural properties of the input graph determine the computational complexity of 2-Club. We want to determine sharp boundaries between tractable and intractable special cases of 2-Club, and whether some graph properties, especially those motivated by the structure of social and biological networks, can be exploited algorithmically. The structural properties, called structural graph parameters, are usually described by integers; well-known examples of such parameters are the maximum degree or the treewidth of a graph. Our results use the classical framework of NP-hardness as well as the framework of parameterized complexity to show (parameterized) tractability and intractability of 2-Club with respect to the structural graph parameters under consideration.

Related Work. For all $s \geq 1$, s-Club is NP-complete on graphs of diameter $s+1$ [3]; 2-Club is NP-complete even on split graphs and, thus, also on chordal graphs [3]. For all $s \geq 1$, s-Club is NP-hard to approximate within a factor of $n^{1/2-\epsilon}$ [2]; a simple approximation algorithm obtains a factor of $n^{1/2}$ for even $s \geq 2$ and a factor $n^{2/3}$ for odd $s \geq 3$ [2]. Several heuristics [6] and integer linear programming formulations [3, 6] for s-Club have been proposed and experimentally evaluated [13]. 1-Club is equivalent to Clique and thus W[1]-hard with respect to ℓ. In contrast, for $s \geq 2$, s-Club is fixed-parameter tractable with respect to ℓ [7, 19], with respect to $n - \ell$ [19][1], and also with respect to the parameter treewidth of G [20]. Moreover, s-Club does not admit a polynomial-size kernel with respect to ℓ (unless NP $\subseteq$ coNP/poly), but admits a so-called *Turing*-kernel with at most k^2-vertices for even s and at most k^3-vertices for odd s [19]. 2-Club is solvable in polynomial time on bipartite graphs, on trees, and on interval

[1] Schäfer et al. [19] actually considered finding an s-club of size *exactly* ℓ. The claimed fixed-parameter tractability with respect to $n - \ell$ however only holds for the problem of finding an s-club of size *at least* ℓ. The other fixed-parameter tractability results hold for both variants.

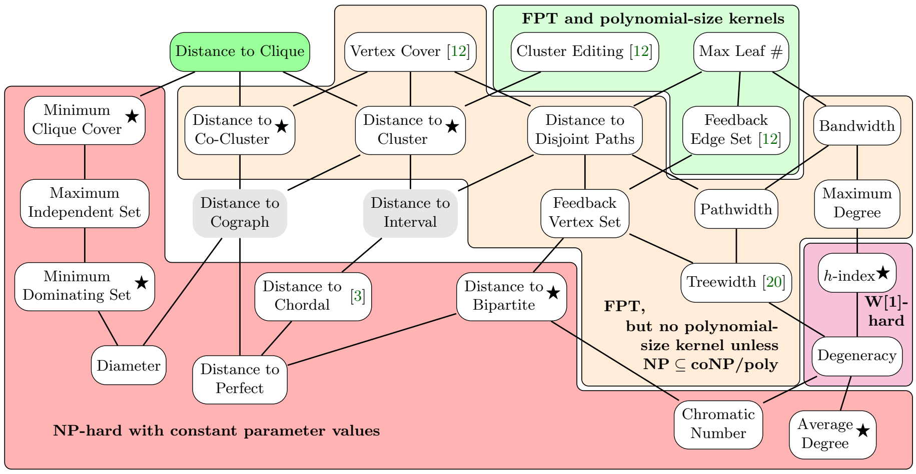

Fig. 1. Overview of the relation between structural graph parameters and of our results for 2-Club. An edge from a parameter α to a parameter β below of α means that β can be upper-bounded in a polynomial (usually linear) function in α. The boxes indicate the complexity of 2-Club with respect to the enclosed parameters. 2-Club is FPT with respect to "distance to clique", but it is open whether it admits a polynomial size kernel. The complexity with respect to "distance to interval" and "distance to cograph" is still open. Results obtained in this work are marked with ★. (For the parameters bandwidth and maximum degree, taking the disjoint union of the input graphs is a composition algorithm that proves the non-existence of polynomial-size kernels [5].)

graphs [20]. In companion work [12], we considered different structural parameters: For instance, we presented fixed-parameter algorithms for the parameters "treewidth" and "size of a vertex cover" and polynomial-size kernels for the parameters "feedback edge set" and "cluster editing number". Furthermore, we presented an efficient implementation for 2-Club based on the fixed-parameter algorithm for the dual parameter $n - \ell$.

Our Contribution. We make progress towards a systematic classification of the complexity of 2-Club with respect to structural graph parameters. Figure 1 gives an overview of our results and their implications. Therein, for a set of graphs Π (for instance the set of bipartite graphs) the parameter *distance to* Π measures the number of vertices that have to be deleted in order to obtain a graph that is isomorphic to one in Π.

In Section 2, we consider the graph parameters minimum clique cover number, minimum dominating set of G, and some related graph parameters. We show that 2-Club is NP-hard even if the minimum clique cover number of G is three, that is, the vertices of G can be covered by three cliques. In contrast, we show that if the minimum clique cover number is two, then 2-Club is polynomial-time solvable. Then, we show that 2-Club is NP-hard even if G has a dominating set of size two, that is, there are two vertices u, v in G such that every vertex in $V \setminus \{u, v\}$ is a neighbor of one of the two. This result is tight in the sense that 2-Club is trivially solvable in case G has a dominating set of size one.

In Section 3, we study the parameter distance to bipartite graphs. We show that 2-Club is NP-hard even if the input graph can be transformed into a bipartite graph by deleting only one vertex. This is somewhat surprising since 2-Club is polynomial-time solvable on bipartite graphs [20]. Then, in Section 4, we consider the graph parameter *h-index*: a graph G has *h-index* k if k is the largest number such that G has at least k vertices of degree at least k. The study of this parameter is motivated by the fact that the h-index is usually small in social networks (see Section 4 for a more detailed discussion). On the positive side, we show that 2-Club is polynomial-time solvable for constant k. On the negative side, we show that 2-Club parameterized by the h-index k of the input graph is W[1]-hard. Hence, a running time of $f(k) \cdot n^{O(1)}$ is probably not achievable. This also implies W[1]-hardness with respect to the parameter degeneracy of G.

Finally, we describe fixed-parameter algorithms for the parameters distance to cluster and co-cluster graphs. Herein, a *cluster graph* is a vertex-disjoint union of cliques, and a *co-cluster graph* is the complement graph of a cluster graph, that is, it is either an independent set or a complete p-partite graph for some $p \leq n$. Interestingly, distance to cluster/co-cluster graph are rare examples for structural graph parameters, that are unrelated to treewidth and still admit a fixed-parameter algorithm (see Figure 1).

Preliminaries. We only consider undirected and simple graphs $G = (V, E)$ where $n := |V|$ and $m := |E|$. For a vertex set $S \subseteq V$, let $G[S]$ denote the *subgraph induced by* S and $G - S := G[V \setminus S]$. We use $\operatorname{dist}_G(u, v)$ to denote the *distance between* u *and* v in G, that is, the length of a shortest path between u and v. For a vertex $v \in V$ and an integer $t \geq 1$, denote by $N_t^G(v) := \{u \in V \setminus \{v\} \mid \operatorname{dist}_G(u, v) \leq t\}$ the set of vertices within distance at most t to v. If the graph is clear from the context, we omit the superscript G. Moreover, we set $N_t[v] := N_t(v) \cup \{v\}$, $N[v] := N_1[v]$ and $N(v) := N_1(v)$. Two vertices v and w are *twins* if $N(v) \setminus \{w\} = N(w) \setminus \{v\}$ and they are *twins with respect to a vertex set* X if $N(v) \cap X = N(w) \cap X$. The twin relation is an equivalence relation; the corresponding equivalence classes are called *twin classes*. The following observation is easy to see.

Observation 1. Let S be a s-club in a graph $G = (V, E)$ and let $u, v \in V$ be twins. If $u \in S$ and $|S| > 1$, then $S \cup \{v\}$ is also an s-club in G.

For the relevant notions of parameterized complexity we refer to [9, 16]. For the parameters distance to cluster/co-cluster graph we assume that a deletion set is provided as an additional input. Note that for both of these parameters there is a polynomial-time constant factor approximation algorithm since cluster graphs and co-cluster graphs are characterized by forbidden induced subgraphs on three vertices. Due to the space restrictions, some proofs are deferred to the appendix.

2 Clique Cover Number and Domination Number

In this section, we prove that on graphs of diameter at most three, 2-Club is NP-hard even if either the clique cover number is three or the domination number is two. We first show that these bounds are tight. The size of a maximum independent set is at most the size of a minimum clique cover. Moreover, since each maximal independent set is a dominating set, a minimum dominating set is also at most the size of a minimum clique cover.

Proposition 1. 2-Club *is polynomial-time solvable on graphs where the size of a maximum independent set is at most two.*

Proof. Let $G = (V, E)$ be a graph. If a maximum independent set in G has size one or it has diameter two, then V is a 2-club. Otherwise, if the maximum independent set in G is of size two, then iterate over all possibilities to choose two vertices $v, u \in V$. Denoting by G' the graph that results from deleting $N(v) \cap N(u)$ in G, output a maximum size set $N^{G'}[v] \cup (N^{G'}(u) \cap N_2^{G'}(v))$ among all iterations.

We next prove the correctness of the above algorithm. For a maximum size 2-club $S \subset V$ in G, there are two vertices $v, u \in V$ such that $v \in S$ and $d_{G[S \cup \{u\}]}(v, u) > 2$, implying that $N(v) \cap N(u) \cap S = \emptyset$. Moreover, $N^{G'}[v]$ and $N^{G'}[u]$ are cliques: Two non-adjacent vertices in $N^{G'}(v)$ (in $N^{G'}(u)$) would form together with u (with v) an independent set.

Since $N^{G'}[v]$ is a clique and $v \in S$, $G[S \cup N^{G'}(v)]$ is a 2-club and thus $N^{G'}[v] \subseteq S$ by the maximality of S. Moreover, since $\{v, u\}$ is a maximum independent set and thus also a dominating set it remains to specify $N(u) \cap S$. However, since $N^{G'}[u]$ is a clique and each vertex in S has to be adjacent to at least one vertex in $N^{G'}(v)$, it follows that $S = N^{G'}[v] \cup (N^{G'}(u) \cap N_2^{G'}(v))$. □

The following theorem shows that the bound on the maximum independent set size in Proposition 1 is tight.

Theorem 1. 2-Club *is NP-hard on graphs with clique cover number three and diameter three.*

Proof. We describe a reduction from Clique. Let $(G = (V, E), k)$ be a Clique instance. We construct a graph $G' = (V', E')$ consisting of three disjoint vertex sets, that is, $V' = V_1 \cup V_2 \cup V_E$. Set V_i, $i \in \{1, 2\}$, to $V_i = V_i^V \cup V_i^{\text{big}}$, where V_i^V is a copy of V and V_i^{big} is a set of n^5 vertices. Let $u, v \in V$ be two adjacent vertices in G and let $u_1, v_1 \in V_1$, $u_2, v_2 \in V_2$ be the copies of u and v in G'. Then add the vertices e_{uv} and e_{vu} to V_E and add the edges $\{v_1, e_{vu}\}, \{u_2, e_{vu}\}, \{u_1, e_{uv}\}, \{v_2, e_{uv}\}$ to G'. Furthermore, add for each vertex $v \in V$ the vertices $V_E^v = e_v^1, e_v^2, \ldots, e_v^{n^3}$ to V_E and make v_1 and v_2 adjacent to all these new vertices. Finally, make the following vertex sets to cliques: V_1, V_2, V_E, and $V_1^{\text{big}} \cup V_2^{\text{big}}$. Observe that G' has diameter three and that it has a clique cover number of three.

We now prove that G has a clique of size $k \Leftrightarrow G'$ has a 2-club of size $k' = 2n^5 + kn^3 + 2k + 2\binom{k}{2}$.

"$\Rightarrow$:" Let S be a clique of size k in G. Let S_c contain all the copies of the vertices of S. Furthermore, let $S_E := \{e_{uv} \mid u_1 \in S_c \wedge v_2 \in S_c\}$ and $S_b := \{e_v^i \mid v_1 \in S_c \wedge 1 \leq i \leq n^3\}$. We now show that $S' := S_c \cup S_E \cup S_b \cup V_1^{\text{big}} \cup V_2^{\text{big}}$ is a 2-club of size k'. First, observe that $|V_1^{\text{big}} \cup V_2^{\text{big}}| = 2n^5$ and $|S_c| = 2k$. Hence, $|S_b| = kn^3$ and $|S_E| = 2\binom{k}{2}$. Thus, S' has the desired size. With a straightforward case distinction one can check that S' is indeed a 2-club.

"$\Leftarrow$:" Let S' be a 2-club of size k'. Observe that G' consists of $|V'| = 2n^5 + 2n + 2m + n^4$ vertices. Since $k' > 2n^5$ at least one vertex of V_1^{big} and of V_2^{big} is in S. Since all vertices in V_1^{big} and in V_2^{big} are twins, we can assume that all vertices of $V_1^{\text{big}} \cup V_2^{\text{big}}$ are contained in S'. Analogously, it follows that at least k sets $V_E^{v^1}, V_E^{v^2}, V_E^{v^3}, \ldots, V_E^{v^k}$ are completely contained in S'. Since S' is a 2-club, the distance from vertices in V_i^{big} to vertices in $V_E^{v^j}$ is at most two. Hence, for each set $V_E^{v^j}$ in S' the two neighbors v_1^j and v_2^j of vertices in $V_E^{v^j}$ are also contained in S'. Since the distance of v_1^i and v_2^j for $v_1^i, v_2^j \in S'$ is also at most two, the vertices $e_{v^i v^j}$ and $e_{v^j v^i}$ are part of S' as well. Consequently, v^i and v^j are adjacent in G. Therefore, the vertices $v^1, \ldots, v^k$ form a size-k clique in G. $\square$

Since a maximum independent set is also a dominating set, Theorem 1 implies that 2-Club is NP-hard on graphs with domination number three and diameter three. In contrast, for domination number one 2-Club is trivial. The following theorem shows that this cannot be extended.

Theorem 2. 2-Club *is NP-hard even on graphs with domination number two and diameter three.*

Proof. We present a reduction from CLIQUE. Let $(G = (V, E), k)$ be a CLIQUE instance and assume that G does not contain isolated vertices. We construct the graph G' as follows. First copy all vertices of V into G'. In G' the vertex set V will form an independent set. Now, for each edge $\{u, v\} \in E$ add an *edge-vertex* $e_{\{u,v\}}$ to G' and make $e_{\{u,v\}}$ adjacent to u and v. Let V_E denote the set of edge-vertices. Next, add a vertex set C of size $n + 2$ to G' and make $C \cup V_E$ a clique. Finally, add a new vertex v^* to G' and make v^* adjacent to all vertices in V. Observe that v^* plus an arbitrary vertex from $V_E \cup C$ are a dominating set of G' and that G' has diameter three. We complete the proof by showing that G has a clique of size $k \Leftrightarrow G'$ has a 2-club of size at least $|C| + |V_E| + k$.

"$\Rightarrow$:" Let K be a size-k clique in G. Then, $S := K \cup C \cup V_E$ is a size-$|C| + |V_E| + k$ 2-club in G: First, each vertex in $C \cup V_E$ has distance two to all other vertices S. Second, each pair of vertices $u, v \in K$ is adjacent in G and thus they have the common neighbor $e_{\{u,v\}}$ in V_E.

"$\Leftarrow$:" Let S be a 2-club of size $|C|+|V_E|+k$ in G'. Since $|C| > |V \cup \{v^*\}|$, it follows that there is at least one vertex $c \in S \cap C$. Since c and v^* have distance three, it follows that $v^* \notin S$. Now since S is a 2-club, each pair of vertices $u, v \in S \cap V$ has at least one common neighbor in S. Hence, V_E contains the edge-vertex $e_{\{u,v\}}$. Consequently, $S \cap V$ is a size-k clique in G. □

3 Distance to Bipartite Graphs

A 2-club in a bipartite graph is a biclique and, thus, 2-CLUB is polynomial-time solvable on bipartite graphs [20]. However, 2-CLUB is already NP-hard on graphs that become bipartite by deleting only one vertex.

Theorem 3. 2-CLUB *is NP-hard even on graphs with distance one to bipartite graphs.*

Proof. We reduce from the NP-hard MAXIMUM 2-SAT problem: Given a positive integer k and a set $\mathcal{C} := \{C_1, \ldots, C_m\}$ of clauses over a variable set $X = \{x_1, \ldots, x_n\}$ where each clause C_i contains two literals, the question is whether there is an assignment β that satisfies at least k clauses.

Given an instance of MAXIMUM 2-SAT where we assume that each clause occurs only once, we construct an undirected graph $G = (V, E)$. The vertex set V consists of the four disjoint vertex sets $V_{\mathcal{C}}$, V_F, V_X^1, V_X^2, and one additional vertex v^*. The construction of the four subsets of V is as follows.

The vertex set $V_{\mathcal{C}}$ contains one vertex c_i for each clause $C_i \in \mathcal{C}$. The vertex set V_F contains for each variable $x \in X$ exactly n^5 vertices $x^1 \ldots x^{n^5}$. The vertex set V_X^1 contains for each variable $x \in X$ two vertices: x_t

which corresponds to assigning true to x and x_f which corresponds to assigning false to x. The vertex set V_X^2 is constructed similarly, but for every variable $x \in X$ it contains $2 \cdot n^3$ vertices: the vertices $x_t^1, \ldots x_t^{n^3}$ which correspond to assigning true to x, and the vertices $x_f^1, \ldots x_f^{n^3}$ which correspond to assigning false to x.

Next, we describe the construction of the edge set E. The vertex v^* is made adjacent to all vertices in $V_{\mathcal{C}} \cup V_F \cup V_X^1$. Each vertex $c_i \in V_{\mathcal{C}}$ is made adjacent to the two vertices in V_X^1 that correspond to the two literals in C_i. Each vertex $x^i \in V_F$ is made adjacent to x_t and x_f, that is, the two vertices of V_X^1 that correspond to the two truth assignments for the variable x. Finally, each vertex $x_t^i \in V_X^2$ is made adjacent to all vertices of V_X^1 except to the vertex x_f. Similarly, each $x_f^i \in V_X^2$ is made adjacent to all vertices of V_X^1 except to x_t. This completes the construction of G which can clearly be performed in polynomial time. Observe that the removal of v^* makes G bipartite: each of the four vertex sets is an independent set and the vertices of $V_{\mathcal{C}}$, V_F, and V_X^2 are only adjacent to vertices of V_X^1.

The main idea behind the construction is as follows. The size of the 2-club forces the solution to contain the majority of the vertices in V_F and V_X^2. As a consequence, for each $x \in X$ exactly one of x_t or x_f is in the 2-club. Hence, the vertices from V_X^2 in the 2-club represent a truth assignment. In order to fulfill the bound on the 2-club size, at least k vertices from $V_{\mathcal{C}}$ are in the 2-club; these vertices can only be added if the corresponding clauses are satisfied by the represented truth assignment. It remains to show the following claim (see appendix).

Claim. $(\mathcal{C}, k)$ is a yes-instance of Maximum 2-Sat $\Leftrightarrow$ G has a 2-club of size $n^6 + n^4 + n + k + 1$. □

4 Average Degree and *h*-Index

2-Club is fixed-parameter tractable for the parameter maximum degree [19]. In social networks, the degree distribution often follows a power law, implying that there are some high-degree vertices but most vertices have low degree [4]. This suggests considering stronger parameters such as h-index, degeneracy, and average degree. Unfortunately, 2-Club is NP-hard even with constant average degree.

Proposition 2. *For any constant $\alpha > 2$, 2-Club is NP-hard on connected graphs with average degree at most α.*

Proof. Let (G, ℓ) be an instance of 2-Club where Δ is the maximum degree of G. We can assume that $\ell > \Delta + 2$ since, as shown for instance in the proof of Theorem 1, 2-Club remains NP-hard in this case. We add a path P to G and an edge from an endpoint p of P to an arbitrary vertex $v \in V$. Since $\ell > \Delta + 2$, any 2-club of size at least ℓ contains at least one vertex that is not in P. Furthermore, it cannot contain p and v since in this case it is a subset of either $N[v]$ or $N[p]$ which both have size at most $\Delta + 2$ (v has degree at most Δ in G). Hence, the instances are equivalent. Putting at least $\lceil \frac{2m}{\alpha-2} - n \rceil$ vertices in P ensures that the resulting graph has average degree at most α. $\square$

Proposition 2 suggests considering "weaker" parameters such as degeneracy or *h-index* [10] of G. Recall that having h-index k means that there are at most k vertices with degree greater than k. Since social networks have small h-index [12], fixed-parameter tractability with respect to the h-index would be desirable. Unfortunately, we show that 2-Club is W[1]-hard when parameterized by the h-index. Following this result, we show that there is "at least" an algorithm that is polynomial for constant h-index.

Theorem 4. 2-Club *parameterized by h-index is W[1]-hard.*

Since the reduction in the proof of Theorem 4 is from Multicolored Clique and in the reduction the new parameter is linearly bounded in the old one, the results of Chen et al. [8] imply the following.

Corollary 1. 2-Club *cannot be solved in $n^{o(k)}$-time on graphs with h-index k unless the exponential time hypothesis fails.*

We next prove that there is an XP-algorithm for the parameter h-index.

Theorem 5. 2-Club *can be solved in $f(k) \cdot n^{2^k} \cdot nm$ time where k is the h-index of the input graph and f solely depends on k.*

Proof. We give an algorithm that finds a maximum 2-club in $G = (V, E)$ in the claimed running time. Let $X \subseteq V$ be the set of all vertices with degree greater than k. By definition, $|X| \leq k$. In a first step, branch into the at most 2^k cases to guess a subset $X' \subseteq X$ that is contained in a maximum 2-club S for G. In case $X' = \emptyset$, one can apply the fixed-parameter algorithm for the parameter maximum degree. In each other branch, proceed as follows. First, delete all vertices from $X \setminus X'$ and while there are vertices that have distance greater than two to any vertex in X', delete them. Denote the resulting graph by $G' = (V', E')$. We next describe how to find a maximum 2-club in $G' = (V', E')$ that contains X'.

Partition all vertices in $V' \setminus X'$ into the at most 2^k twin classes $T_1, \ldots, T_p$ with respect to X'. Two twin classes T and T' are in *conflict* if

$N(T) \cap N(T') \cap X' = \emptyset$. Now, the crucial observation is that, if T and T' are in conflict, then all vertices in $(T \cup T') \cap S$ are contained in the same connected component of $G'[S \setminus X']$. Then, since all vertices in $T \cap S$ have in $G'[S \setminus X]$ distance at most two to all vertices in $T' \cap S$, it follows that all vertices in $T \cap S$ have pairwise distance at most four in $G'[S \setminus X']$.

Now, branch into the $O(n^{2^k})$ cases to choose for each twin class T a *center* c, that is, a vertex from $T \cap S$. Clearly, if $T \cap S = \emptyset$, then there is no center c and we delete all vertices in T. Consider a remaining twin class T that is in conflict to any other twin class. By the observation above, $T \cap S$ is contained in one connected component of $G'[S \setminus X']$ and in this component all vertices in $T \cap S$ have pairwise distance at most four. Moreover, the graph $G[V' \setminus X']$ has maximum degree at most k. Thus for the center c of T one can guess $N_4^S(c) := N_4(c) \cap S$ by branching into at most 2^{k^4} cases. Guess the set $N_4^S(c)$ for all centers c where the corresponding twin class is in conflict to at least one other twin class and fix them to be contained in the desired 2-club S. Delete all vertices in T guessed to be not contained in $N_4^S(c)$.

Let $\tilde{S}$ be the set of vertices guessed to be contained in S. Next, while there is a vertex $v \in V' \setminus \tilde{S}$ that has distance greater than two to any vertex in $\tilde{S}$, delete v. Afterwards, check whether all vertices in $\tilde{S}$ have pairwise distance at most two. (If this check fails, then this branch cannot lead to any solution.) We next prove that the remaining graph is a 2-club.

In the remaining graph, each pair of vertices in $V' \setminus \tilde{S}$ has distance at most two and thus the graph is a 2-club: Suppose that two remaining vertices $v, w \in V' \setminus \tilde{S}$ have distance greater than two. Let T and T' be the twin classes with the corresponding centers c, c' such that $v \in T$ and $w \in T'$. In case $T = T'$, it follows that $N(T) \cap X' = \emptyset$ (since $X' \subseteq \tilde{S}$). However, since T cannot be in conflict with any other twin class (otherwise $v, w \in N_4^S(c) \subseteq \tilde{S}$), it follows that S only contains the twin class T. This implies that v and w have distance greater than two to all vertices in X' (note that $X' \neq \emptyset$), a contradiction. In case $T \neq T'$, since $N(v) \cap N(w) = \emptyset$ it follows that T is in conflict to T', implying that $v \in N_4^S(c)$ and $w \in N_4^S(c')$, a contradiction to $v, w \in V' \setminus \tilde{S}$. $\square$

5 Distance to Cluster and Co-Cluster Graphs

We now present a simple fixed-parameter algorithm for 2-Club parameterized by distance to co-cluster graphs. The algorithm is based on the fact that each co-cluster graph is either an independent set or a 2-club.

Theorem 6. 2-Club *is solvable in* $O(2^k \cdot 2^{2^k} \cdot nm)$ *time where* k *denotes the distance to co-cluster graphs.*

Proof. Let (G, X, ℓ) be an 2-Club instance where X with $|X| = k$ and $G - X$ is a co-cluster graph. Note that the co-cluster graph $G - X$ is either a connected graph or an independent set. In the case that $G - X$ is an independent set, the set X is a vertex cover and we thus apply the algorithm we gave in companion work [12] to solve the instance in $O(2^k \cdot 2^{2^k} \cdot nm)$ time.

Hence, assume that $G - X$ is connected. Since $G - X$ is the complement of a cluster graph, this implies that $G - X$ is a 2-club. Thus, if $\ell \leq n - k$, then we can trivially answer yes. Hence, assume that $\ell > n - k$ or, equivalently, $k > n - \ell$. Schäfer et al. [19] showed that 2-club can be solved in $O(2^{n-\ell}nm)$ (simply choose a vertex pair having distance at least three and branch into the two cases of deleting one of them). Since $k > n - \ell$ it follows that 2-club can be solved in $O(2^k nm)$ time in this case. □

Next, we present a fixed-parameter algorithm for the parameter distance to cluster graphs.

Theorem 7. 2-Club *is solvable in* $O(2^k \cdot 3^{2^k} \cdot nm)$ *time where* k *denotes distance to cluster graphs.*

6 Conclusion

Although the complexity status of 2-Club is resolved for most of the parameters in the complexity landscape shown in Figure 1, some open questions remain. What is the complexity of 2-Club parameterized by "distance to interval graphs" or "distance to cographs"? The latter parameter seems particularly interesting since every induced subgraph of a cograph has diameter two. Hence, the distance to cographs measures the distance from this trivial special case. In contrast to the parameter h-index, it is open whether 2-Club parameterized by the degeneracy is in XP or NP-hard on graphs with constant degeneracy. Finally, it would be interesting to see which results carry over to 3-Club [13, 18].

Bibliography

[1] R. Alba. A graph-theoretic definition of a sociometric clique. *J. Math. Sociol.*, 3 (1):113–126, 1973.

[2] Y. Asahiro, E. Miyano, and K. Samizo. Approximating maximum diameter-bounded subgraphs. In *Proc. 9th LATIN*, volume 6034 of *LNCS*, pages 615–626. Springer, 2010.

[3] B. Balasundaram, S. Butenko, and S. Trukhanovzu. Novel approaches for analyzing biological networks. *J. Comb. Optim.*, 10(1):23–39, 2005.

[4] A. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509, 1999.

[5] H. L. Bodlaender, R. G. Downey, M. R. Fellows, and D. Hermelin. On problems without polynomial kernels. *J. Comput. System Sci.*, 75(8):423–434, 2009.

[6] J.-M. Bourjolly, G. Laporte, and G. Pesant. An exact algorithm for the maximum k-club problem in an undirected graph. *European J. Oper. Res.*, 138(1):21–28, 2002.

[7] M. S. Chang, L. J. Hung, C. R. Lin, and P. C. Su. Finding large k-clubs in undirected graphs. In *Proc. 28th Workshop on Combinatorial Mathematics and Computation Theory*, 2011.

[8] J. Chen, B. Chor, M. Fellows, X. Huang, D. W. Juedes, I. A. Kanj, and G. Xia. Tight lower bounds for certain parameterized NP-hard problems. *Inf. Comput.*, 201(2):216–231, 2005.

[9] R. G. Downey and M. R. Fellows. *Parameterized Complexity.* Springer, 1999.

[10] D. Eppstein and E. S. Spiro. The h-index of a graph and its application to dynamic subgraph statistics. In *Proc. 11th WADS*, volume 5664 of *LNCS*, pages 278–289. Springer, 2009.

[11] M. R. Fellows. Towards fully multivariate algorithmics: Some new results and directions in parameter ecology. In *Proc. 20th IWOCA*, volume 5874 of *LNCS*, pages 2–10. Springer, 2009.

[12] S. Hartung, C. Komusiewicz, and A. Nichterlein. Parameterized algorithmics and computational experiments for finding 2-clubs. In *Proc. 7th IPEC*, volume 7535 of *LNCS*, pages 231–241. Springer, 2012.

[13] F. Mahdavi and B. Balasundaram. On inclusionwise maximal and maximum cardinality k-clubs in graphs. *Discrete Optim.*, 2012. To appear.

[14] N. Memon and H. L. Larsen. Structural analysis and mathematical methods for destabilizing terrorist networks using investigative data mining. In *Proc. 2nd ADMA*, volume 4093 of *LNCS*, pages 1037–1048. Springer, 2006.

[15] R. J. Mokken. Cliques, Clubs and Clans. *Quality and Quantity*, 13:161–173, 1979.

[16] R. Niedermeier. *Invitation to Fixed-Parameter Algorithms.* Oxford University Press, 2006.

[17] R. Niedermeier. Reflections on multivariate algorithmics and problem parameterization. In *Proc. 27th STACS*, volume 5 of *LIPIcs*, pages 17–32. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2010.

[18] S. Pasupuleti. Detection of protein complexes in protein interaction networks using n-Clubs. In *Proc. 6th EvoBIO*, volume 4973 of *LNCS*, pages 153–164. Springer, 2008.

[19] A. Schäfer, C. Komusiewicz, H. Moser, and R. Niedermeier. Parameterized computational complexity of finding small-diameter subgraphs. *Optim. Lett.*, 6(5), 2012.

[20] A. Schäfer. Exact algorithms for s-club finding and related problems. Diploma thesis, Friedrich-Schiller-Universität Jena, 2009.

[21] S. B. Seidman and B. L. Foster. A graph-theoretic generalization of the clique concept. *J. Math. Sociol.*, 6:139–154, 1978.

A Proofs

A.1 Proof 1 (Claim in the proof of Theorem 3)

Proof. "$\Rightarrow$": Let β be an assignment for X that satisfies k clauses $C_1, \ldots, C_k$ of $\mathcal{C}$. Consider the vertex set S that consists of V_F, v^*, the vertex set $\{c_1, \ldots, c_k\} \subseteq V_{\mathcal{C}}$ that corresponds to the k satisfied clauses, and for each $x \in X$ of the vertex set $\{x_t, x_t^1, \ldots, x_t^{n^3}\} \subseteq V_X^1 \cup V_X^2$ if $\beta(x) = \text{true}$ and $\{x_f, x_f^1, \ldots, x_f^{n^3}\} \in V_X^1 \cup V_X^2$ if $\beta(x) = \text{false}$. Clearly, $|S| = n^6 + n^4 + n + k + 1$. In the following, we show that S is a 2-club. Herein, let $S_X^1 := V_X^1 \cap S$, $S_X^2 := V_X^2 \cap S$, and $S_{\mathcal{C}} := V_{\mathcal{C}} \cap S$.

First, v^* is adjacent to all vertices in $S_{\mathcal{C}} \cup V_F \cup S_X^1$. Hence, all vertices of $S \setminus S_X^2$ are within distance two in $G[S]$. By construction, the vertex sets S_X^1 and S_X^2 form a complete bipartite graph in G: A vertex $x_t^i \in S_X^2$ is adjacent to all vertices in V_X^1 except x_f which is not contained in S_X^1. The same argument applies to some $x_f^i \in S_X^2$. Hence, the vertices of S_X^2 are neighbors of all vertices in S_X^1. This also implies that the vertices of S_X^2 are in $G[S]$ within distance two from v^* and from every vertex in V_F since each vertex of $V_F \cup \{v^*\}$ has at least one neighbor in S_X^1. Finally, since the k vertices in $S_{\mathcal{C}}$ correspond to clauses that are satisfied by the truth assignment β, each of these vertices has at least one neighbor in S_X^1. Hence, every vertex in S_X^2 has in $G[S]$ distance at most two to every vertex in $S_{\mathcal{C}}$.

"$\Leftarrow$": Let S be a 2-club of size $n^6 + n^4 + n + k + 1$, and let $S_X^1 := V_X^1 \cap S$, $S_X^2 := V_X^2 \cap S$, $S_F := V_F \cap S$ and $S_{\mathcal{C}} := V_{\mathcal{C}} \cap S$. Clearly, neither $S_X^2 = \emptyset$ nor $S_F = \emptyset$.

Since $|V_{\mathcal{C}}| + |V_X^1| + |V_x^2| + 1 \leq n^2 + 2n + 2n^4 + 1 < n^5$ for sufficiently large n, S contains more than $n^6 - n^5$ vertices from V_F. Consequently, for each $x \in X$ there is an index $1 \leq i \leq n^5$ such that $x^i \in S_F$.

We next show that for each $x \in X$ it holds that either x_t or x_f is contained in S_X^1. Towards this, since S is a 2-club, every vertex pair $x^i \in S_F$ and $u \in S_X^2$ has at least one common neighbor in S. By construction, this common neighbor is a vertex of S_X^1 and thus either x_t or x_f. Moreover, by the observation above for each $x \in X$ at least one x^i is contained in S_F. Thus, for each $x \in X$ at least one of x_t and x_f is contained in S_X^1.

Now observe that, $G[S_X^1 \cup S_X^2]$ is a complete bipartite graph, since S_X^1 and S_X^2 are independent sets and S_X^2 has only neighbors in S_X^1. This implies that if for some $x \in X$ there exists indices $1 \leq i, j \leq n^3$ with x_t^i and x_f^j are in S_X^2, then x_t and x_f are *not* in S_X^1. This contradicts the above observation that at least one of x_t and x_f is in S_X^1. Moreover, since $|V_{\mathcal{C}}| + |V_X^1| + 1 \leq n^2 + 2n + 1 < n^3$ and $|S \setminus V_F| > n^4$, we have $|S_X^2| >$

$n^4 - n^3$. It follows that for each $x \in X$ there is an index $1 \le i \le n^3$ such that either $x_t^i \in S_X^2$ or $x_f^i \in S_X^2$. Finally, this implies that either x_t or x_f is not contained in S_X^1.

Summarizing, S has at most n^6 vertices from V_F, at most n^4 vertices belonging to S_X^2, exactly n vertices belonging to S_X^1, and thus there are $k+1$ vertices in $S_\mathcal{C} \cup \{v^*\}$. Since S is a 2-club that has nonempty S_X^2, every one of the at least k vertices from $S_\mathcal{C}$ has at least one neighbor in S_X^1. Because for each $x \in X$ either x_f or x_t is in S_X^1, the n vertices from S_X^1 correspond to an assignment β of X. By the above observation, this assignment satisfies at least k clauses of $\mathcal{C}$. $\square$

A.2 Proof 2 (Theorem 4)

Proof. We give a parameterized reduction from Multicolored Clique parameterized by the solution size k to 2-Club parameterized by the h-index of the input graph. Let $(G = (V, E), \mathrm{c}, k)$ be a Multicolored Clique instance where $\mathrm{c} : V \to \{1, \ldots, k\}$ is the vertex coloring. We construct an instance $(G' = (V', E'), \ell)$ of 2-Club with G' having an h-index of $O(k)$ as follows.

For each vertex $v \in V$ create a *vertex gadget* G_v consisting of the $3(n+1)$ vertices $\alpha_1^v, \ldots, \alpha_n^v, \beta_1^v, \ldots, \beta_{n+1}^v, \gamma_1^v, \ldots, \gamma_n^v, \omega_1^v, \omega_2^v$. These vertices form the cycle $\alpha_1^v, \beta_1^v, \gamma_1^v$, $\alpha_2^v, \beta_2^v, \gamma_2^v, \ldots, \alpha_n^v$, β_n^v, γ_n^v, $\omega_1^v, \beta_{n+1}^v, \omega_2^v, \alpha_1^v$. In the following, we call a vertex α_i^v (β_i^v, γ_i^v) an *α-vertex* (*β-vertex*, *γ-vertex*). Let $v_1, \ldots, v_n$ be an arbitrary fixed ordering of the vertices in V. Then, for each edge $\{v_i, v_j\} \in E$ add an *edge-vertex* e_{v_i,v_j} that is made adjacent to $\alpha_j^{v_i}$, $\beta_j^{v_i}$, $\beta_i^{v_j}$, and $\gamma_i^{v_j}$.

The main idea is to force the solution such that it contains exactly k vertex gadgets, each of them completely, and $\binom{k}{2}$ edge-vertices that all correspond to edges between the k vertices. Note that so far the largest 2-club in G' has size five (a star with an edge-vertex as center). We now add a *connection gadget* G_c that is contained in the solution and has the following three main purposes. First, it reduces the distance between the vertices in the vertex gadgets. Second, it enforces that for each vertex gadget either all or no vertices are in the solution. And finally, it enforces that an edge-vertex can only belong to the solution in case the two vertex gadgets in which it has neighbors are also in the solution.

The connection gadget contains three distinguished vertices u_α, u_β, u_γ that form an independent set in G'. The vertex u_α is made adjacent to ω_1^v, $v \in V$, and to each α-vertex, the vertex u_β is made adjacent to each β-vertex, and the vertex u_γ is made adjacent to ω_2^v, $v \in V$, and to

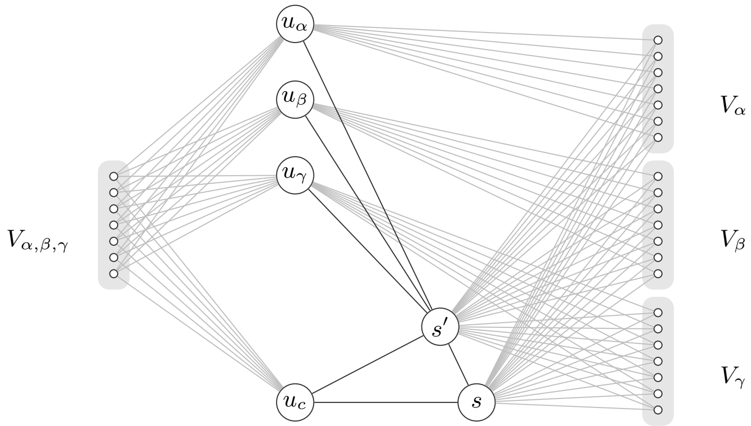

Fig. 2. The connection gadget.

each γ-vertex. We add three further vertices s, s', u_c and four large vertex sets V_α, V_β, V_γ, and $V_{\alpha,\beta,\gamma}$, each containing n^3 vertices to the gadget. Then, we add edges to the gadget such that

- the vertex s is adjacent to all vertices in V_α, V_β, and V_γ and to s' and u_c,
- each vertex in V_α (V_β, V_γ) is adjacent to u_α (u_β, u_γ, respectively) and to s',
- the vertex s' is adjacent to u_α, u_β, u_γ, and u_c, and
- each vertex in $V_{\alpha,\beta,\gamma}$ is adjacent to u_α, u_β, u_γ, and u_c.

An illustration of the connection gadget is given in Figure 2. Note that the connection gadget is a 2-club. Next, we add edges between the connection gadget and the vertex gadgets and edge-vertices such that

- ω_1^v is adjacent to u_c and u_α for each $v \in V$,
- ω_2^v is adjacent to s and u_γ for each $v \in V$,
- each α-vertex is adjacent to u_α, each β-vertex to u_β, and each γ-vertex to u_γ, and
- all edge-vertices and all α-, β-, and γ-vertices are adjacent to s and to u_c.

Next, we add the *coloring gadget*, whose purpose is to ensure that the solution does not contain vertices from two different vertex gadgets G_u and G_v when $c(u) = c(v)$. The construction of the gadget is as follows. For each color $c \in \{1, \ldots, k\}$ add four vertices $u_{1,1}^c$, $u_{1,2}^c$, $u_{2,1}^c$, and $u_{2,2}^c$, and

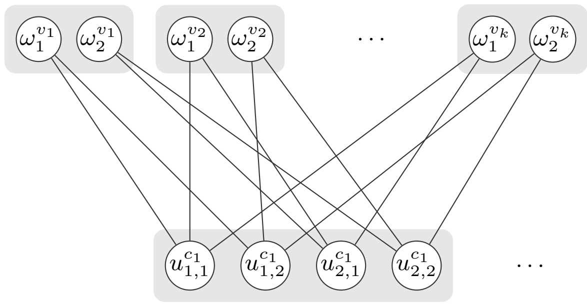

Fig. 3. A scheme of the second part of the connection gadget with $\mathrm{c}(v_i) = i$. todo..

let V_u denote the vertex set containing these vertices. Then, add edges such that

- V_u is a clique,
- each vertex of V_u is adjacent to s, s', and u_c, and
- $u_{1,1}^1$ is adjacent to all vertices in $V_\alpha \cup V_\beta \cup V_\gamma$.

Next, we add edges between the coloring gadget and the vertex gadgets and edge-vertices in G' such that for each vertex $v \in V$

- ω_1^v is adjacent to $u_{1,1}^{\mathrm{c}(v)}$ and to $u_{1,2}^{\mathrm{c}(v)}$,
- ω_2^v is adjacent to $u_{2,1}^{\mathrm{c}(v)}$ and $u_{2,2}^{\mathrm{c}(v)}$,
- ω_1^v is adjacent to $u_{1,1}^c$ and $u_{2,1}^c$ for each $c \neq c(v)$, and
- ω_2^v is adjacent to $u_{1,2}^c$ and $u_{2,2}^c$ for each $c \neq c(v)$.

See Figure 3 for an illustration.

This completes the construction of G'; the reduction is completed by setting $\ell := 4n^3 + 6 + 4k + k \cdot 3(n+1) + \binom{k}{2}$.

Note that G' is a 3-club. The pairwise distance of all vertices in G' is at most two except for the vertices ω_1^u and ω_2^v for $\mathrm{c}(u) = \mathrm{c}(v)$. Furthermore there are exactly $4k + 6$ vertices having degree more than $4k + 6$: u_α, u_β, u_γ, s, s', u_c, and the $4k$ vertices in V_u. Hence, the h-index of G' is $4k + 6 = O(k)$.

We complete the proof by showing the following claim.

(G, c, k) is a yes-instance of Multicolored Clique $\iff$ (G', ℓ) is a yes-instance of 2-Club.

"$\Rightarrow$:" Let $K \subseteq V$ be a multicolored size-k clique in G. Then, consider the vertex set $S \subseteq V'$ that contains all vertices in the connection gadget and in the coloring gadget, all vertices in the vertex gadget G_v for each $v \in K$ and all edge-vertices $e_{u,v}$ for $u, v \in K$. Observe that $|S| = \ell$ since S contains the $4n^3 + 6 + 4k$ vertices in the connection gadget and in the coloring gadget plus $3(n + 1)$ vertices in each of the k vertex gadgets plus $\binom{k}{2}$ edge-vertices.

We now show that $G[S]$ is a 2-club. By construction, the connection gadget is a 2-club. Furthermore, all vertices in $V_\alpha \cup V_\beta \cup V_\gamma \cup V_{\alpha,\beta,\gamma} \cup V_u \cup \{u_c, s, s'\}$ have distance two to all edge-vertices and all α-,β-,γ-vertices since u_c and s are adjacent to all of these vertices. This also implies that every vertex pair from the set of edge-vertices and α-,β-,γ-vertices has distance two. The vertices in the vertex gadgets also have distance at most two to the vertices in u_α,u_β, and u_γ, since only complete vertex gadgets are contained in S. It remains to show that all w-vertices have distance at most two. This is clear if they are from the same vertex gadget. If they are from different vertex gadgets, this holds because K is multicolored (see Figure 3).

"$\Leftarrow$:" Let $S \subseteq V'$ be a 2-club of size ℓ. First, we show that we can assume that $V_\alpha \cup V_\beta \cup V_\gamma \cup V_{\alpha,\beta,\gamma} \subseteq S$. Note that $|V'| = 4n^3 + 6 + 4k + n \cdot 3(n + 1) + m$. Thus, the number of vertices that are *not* in S is at most $|V'| - \ell$, which is:

$$\begin{aligned}|V'| - \ell = 4n^3 + 6 + 4k + n \cdot 3(n+1) + m& \\ - \left(4n^3 + 6 + 4k + k \cdot 3(n+1) + \binom{k}{2}\right)& \\ = 3(n+1)(n-k) + m - \binom{k}{2} < n^3.&\end{aligned}$$

Thus, less than $n^3 = |V_\alpha| = |V_\beta| = |V_\gamma| = |V_{\alpha,\beta,\gamma}|$ vertices of G' are not in S and, hence, S contains at least one vertex of each set V_α, V_β, V_γ, and $V_{\alpha,\beta,\gamma}$. Since 2-Club asks to find a 2-club of size *at least* ℓ, this implies that we can assume that *all* vertices of V_α, V_β, V_γ, and $V_{\alpha,\beta,\gamma}$ are in S. Note that the neighborhood of $V_{\alpha,\beta,\gamma}$ overlaps with the neighborhood of V_α (V_β, V_γ) in exactly one vertex: u_α (u_β, u_γ). Since S contains one vertex of each set V_α, V_β, V_γ, and $V_{\alpha,\beta,\gamma}$, this implies that S contains also u_α, u_β, and u_γ.

We prove that if for any $v \in V$ one vertex of the vertex gadget G_v is in the 2-club S, then all vertices of G_v are in S. Suppose towards a contradiction that S contains some but not all vertices of G_v. Then there is a vertex x in G_v such that $x \in S$ but a neighbor y in G_v is not in S. By construction, x and y are each adjacent to exactly one of u_α, u_β, and u_γ but not to the same vertex. Assume w.l.o.g that x is adjacent to u_α

and y is adjacent to u_β. Since y is the only common neighbor of x and u_β, x has distance three or more to u_β, a contradiction. Hence, either all or no vertices of a vertex gadget G_v are in S. Accordingly, we can interpret the vertex gadgets contained in S as a vertex set K in G.

Next, we prove that S contains at most k different vertex gadgets. Suppose towards a contradiction that S contains more than k vertex gadgets. Then, there are two vertex gadgets, G_u and G_v such that $\mathrm{c}(u) = \mathrm{c}(v)$. Then, the vertices ω_1^u and ω_2^v do not have a common neighbor in the coloring gadget nor in the connection gadget and, hence, these two vertices have distance three, a contradiction. Consequently, there are at most k vertex gadgets in S, and thus $|K| \leq k$.

Since S has size at least $4n^3 + 6 + 4k + k \cdot 3(n+1) + \binom{k}{2}$ this implies that S contains at least $\binom{k}{2}$ edge-vertices: The connection gadget G_c and the coloring gadget contain altogether $4n^3+6+4k$ vertices, and S contains at most k vertex gadgets, each of them consisting of $3(n+1)$ vertices.

We now show that the $\binom{k}{2}$ edge-vertices correspond to edges that have both endpoints in K. Each edge-vertex has exactly six neighbors: u_c, s, one α-vertex, two β-vertices, and one γ-vertex. Hence, if an edge-vertex $e_{u,v}$ is contained in S, then the adjacent α- and γ-vertex and at least one of the two β-vertices are also in S, otherwise $e_{u,v}$ has distance three to u_α, u_β, or u_γ. Consequently, if an edge-vertex $e_{u,v}$ is contained in S, then G_v and G_u are also contained in S. Hence, u and v are in K.

This means that K contains all endpoints of the edges corresponding to the at least $\binom{k}{2}$ edge-vertices, and thus that K is a size-k clique in G. As argued above, the k vertices also have pairwise distinct colors, which implies that (G, k) is a yes-instance of MULTICOLORED CLIQUE. $\square$

A.3 Proof 3 (Theorem 7)

Proof. Let (G, X, ℓ) be a 2-CLUB instance where $G - X$ is a cluster graph and $|X| = k$. First, branch into all possibilities to choose the subset $X' \subseteq X$ that is contained in the desired 2-club S. Then, remove $X \setminus X'$ and all vertices that are not within distance two to all vertices in X', and let $G' = (V', E')$ denote the resulting graph.

Let $\mathcal{T} = T_1, \ldots, T_p$ be the set of twin classes of $V' \setminus X'$ with respect to X' and let $C_1, \ldots, C_q$ denote the clusters of $G' - X'$. Two twin classes T and T' are in *conflict* if $N(T) \cap N(T') \cap X' = \emptyset$. The three main observations exploited in the algorithm are the following: First, if two conflict classes T_i and T_j are in conflict, then all vertices of T_i that are in a 2-club and all vertices from T_j that are in a 2-club must be in the same cluster of $G' - X'$. Second, every vertex from $G' - X'$ can reach all vertices in X' only via vertices of X' or via vertices in its own cluster. Third, if one 2-club-vertex $v \in S$ is in a twin class $v \in T_i$ and in a cluster $v \in C_j$, then all vertices that are in T_i and in C_j can be added to S without violating the 2-club property.

We exploit these observations in a dynamic programming algorithm. In this algorithm, we create a two-dimensional table $\mathcal{A}$ where an entry $\mathcal{A}[i, \mathcal{T}']$ stores the maximum size of a set $Y \subseteq \bigcup_{1 \leq j \leq i} C_j$ such that the twin classes of Y are *exactly* $\mathcal{T}' \subseteq \mathcal{T}$ and all vertices in Y have in $G[Y \cup X']$ distance at most two to each vertex from $Y \cup X'$.

Before filling the table $\mathcal{A}$, we calculate a value $s(i, \mathcal{T}')$ that stores the maximum number of vertices we can add from C_i that are from the twin classes in $\mathcal{T}'$ and fulfill the requirements in the previous paragraph. This value is defined as follows. Let $C_i^{\mathcal{T}'}$ denote the maximal subset of vertices from C_i whose twin classes are exactly $\mathcal{T}'$. Then, $s(i, \mathcal{T}') = |C_i^{\mathcal{T}'}|$ if $C_i^{\mathcal{T}'}$ exists and every pair of non-adjacent vertices from $C_i^{\mathcal{T}'}$ and from X' have a common neighbor. Otherwise, set $s(i, \mathcal{T}') = -\infty$. Note that as a special case we set $s(i, \emptyset) = 0$. Furthermore, for two subsets $\mathcal{T}''$ and $\tilde{\mathcal{T}}$ define the predicate $\mathrm{conf}(\mathcal{T}'', \tilde{\mathcal{T}})$ as true if there is a pair of twin classes $T_i \in \mathcal{T}''$ and $T_j \in \tilde{\mathcal{T}}$ such that T_i and T_j are in conflict, and as false, otherwise.

Using these values, we now fill $\mathcal{A}$ with the following recurrence:

$$\mathcal{A}[i, \mathcal{T}'] = \\ \max_{\mathcal{T}'' \subseteq \mathcal{T}', \tilde{\mathcal{T}} \subseteq \mathcal{T}'} \begin{cases} \mathcal{A}[i-1, \tilde{\mathcal{T}}] + s(i, \mathcal{T}'') & \text{if } \tilde{\mathcal{T}} \cup \mathcal{T}'' = \mathcal{T}' \wedge \neg \mathrm{conf}(\tilde{\mathcal{T}}, \mathcal{T}''), \\ -\infty & \text{otherwise.} \end{cases}$$

This recurrence considers all cases of combining a set Y for the clusters C_1 to C_{i-1} with a solution Y' for the cluster C_i. Herein, a positive table entry is only obtained when the twin classes of $Y \cup Y'$ is exactly $\mathcal{T}'$ and the pairwise distances between $Y \cup Y'$ and $Y \cup Y' \cup X'$ in $G[Y \cup Y' \cup X']$ are at most two. The latter property is ensured by the definition of the $s()$ values and by the fact that we consider only combinations that do not put conflicting twin classes in different clusters.

Now, the table entry $\mathcal{A}[q, \mathcal{T}']$ contains the size of a maximum vertex set Y such that in $G'[Y \cup X']$ every vertex from Y has distance two to all other vertices. It remains to ensure that the vertices from X' are within distance two from each other. This can be done by only considering a table entry $\mathcal{A}[q, \mathcal{T}']$ if each non-adjacent vertex pair $x, x' \in X'$ has either a common neighbor in X' or in one twin class contained in $\mathcal{T}'$. The maximum size of a 2-club in G' is then the maximum value of all table entries that fulfill this condition.

The running time can be bounded roughly as $O(2^k \cdot 3^{2^k} \cdot nm)$: We try all 2^k partitions of X and for each of these partitions, we fill a dynamic programming table with $2^{2^k} \cdot n$ entries. The number of overall table lookups and updates is $O(3^{2^k} \cdot n)$ since there are 3^{2^k} possibilities to partition $\mathcal{T}$ into the three sets $\mathcal{T}''$, $\tilde{\mathcal{T}}$, and $\mathcal{T} \setminus \mathcal{T}'$. Since each C_i is a clique, the entry $s(i, \mathcal{T}')$ is computable in $O(nm)$ time and the overall running time follows. $\square$