European Journal of Operational Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

# On integer programming models for the maximum 2-club problem and its robust generalizations in sparse graphs

Alexander Veremyev[a], Vladimir Boginski[a], Eduardo L. Pasiliao[b], Oleg A. Prokopyev[c,*]

[a] Department of Industrial Engineering and Management Systems, University of Central Florida, 12800 Pegasus Dr., Orlando, FL 32816, USA
[b] Munitions Directorate, Air Force Research Laboratory, 101 W. Eglin Blvd, Eglin AFB, FL 32542, USA
[c] Department of Industrial Engineering, University of Pittsburgh, 1031 Benedum Hall, Pittsburgh, PA 15261, USA

ARTICLE INFO

*Article history:*
Received 7 April 2020
Accepted 10 May 2021
Available online xxx

*Keywords:*
Networks
Graph theory
2-Clubs
Integer programming
Clique relaxations

ABSTRACT

We consider the maximum 2-club problem, which aims at finding an induced subgraph of maximum cardinality with the diameter at most two. Such subgraphs arise from a popular diameter-based clique relaxation concept, as a subgraph is a clique if and only if its diameter is one. In a 2-club every pair of non-adjacent vertices has a common neighbor; this “2-hop” property naturally arises in a variety of applications. In this paper, by exploiting a somewhat different interpretation of the problem, we provide two new mixed-integer programming (MIP) models for finding maximum 2-clubs. Our MIPs provide much tighter linear programming (LP) relaxations for sufficiently sparse graphs and have fewer constraints than the standard integer programming (IP) model at the expense of having slightly more continuous variables. We also consider feasibility versions of our MIPs that verify whether there exists a 2-club of some specified size. Then we incorporate them into a simple-to-implement “feasibility-check” algorithm that iteratively solves one of the feasibility MIPs for each possible 2-club size within some known lower and upper bounds. The upper bound is obtained from an LP relaxation of our new MIPs and is shown to be sharp. Furthermore, we show how to extend our approaches for solving some “robust” (attack- and failure-tolerant) generalizations of the maximum 2-club problem. Finally, we perform an extensive computational study with randomly generated and real-life graphs to support our theoretical results and to provide some empirical observations and insights.



## 1. Introduction

Let $G=(V,E)$ be a simple undirected graph with the sets of vertices (nodes) $V$ and edges $E$, where $|V|=n$ and $|E|=m$. Graph $G$ is called *complete* if it has all possible edges, i.e., $(i,j)\in E$ for all $i,j\in V$, $i\neq j$. A path between $i$ and $j$ in $G$ is a *shortest path* if it contains the least number of edges among all paths between $i$ and $j$ in $G$; the length (i.e., number of edges) of a shortest path between two vertices $i$ and $j$ in $G$ is also referred to as the *distance* between $i$ and $j$ in $G$ and denoted by $d_G(i,j)$. We assume that $d_G(i,j)=+\infty$ if there is no path between $i$ and $j$. The maximum distance between any two vertices in $G$ is referred to as the *diameter* of $G$, i.e., $diam(G)=\max\{d_G(i,j)\mid i,j\in V\}$. In the remainder of the paper, without loss of generality, we assume that $diam(G)<+\infty$, i.e., graph $G$ is *connected*.

For any subset of vertices $S\subseteq V$, let $G[S]=(S,E')$, where $E'=\{(i,j)\in E\mid i,j\in S\}\subseteq E$, denote the subgraph *induced* by $S$ in $G$. A *clique* $C$ is a subset of $V$ such that $G[C]$ is a complete graph; the problem of finding a clique of maximum cardinality in a given graph is referred to as the *maximum clique problem* (Pardalos & Xue, 1994). This problem is one of the classical *NP*-hard combinatorial optimization problems with numerous applications; see, e.g., surveys in Butenko and Wilhelm (2006); Garey and Johnson (2002); Pardalos and Xue (1994).

It has been observed in a number of related studies, e.g., Komusiewicz (2016) and Pattillo, Youssef, and Butenko (2013b), that the clique concept is somewhat idealized and too restrictive in many application contexts as it requires all pairwise connections in an induced subgraph. Thus, multiple clique relaxation models have been introduced in the network analysis and optimization literature to capture more realistic considerations arising in various practical settings; see, for example, a general framework outlined in Pattillo et al. (2013b). Perhaps one of the most popular clique relaxation models is the concept of a *k-club*, which is defined as a subset of vertices $S\subseteq V$ such that

* Corresponding author.
*E-mail addresses:* alexander.veremyev@ucf.edu (A. Veremyev), vladimir.boginski@ucf.edu (V. Boginski), eduardo.pasiliao@us.af.mil (E.L. Pasiliao), droleg@pitt.edu (O.A. Prokopyev).

https://doi.org/10.1016/j.ejor.2021.05.010
0377-2217/

the subgraph $G[S]$ induced by $S$ in $G$ has diameter at most $k$, i.e., $diam(G[S]) \leq k$, where $k$ is a fixed positive integer. Clearly, $k = 1$ corresponds to a clique, while having $k \geq 2$ defines a graph with somewhat less restrictive diameter requirements. In the *maximum k-club problem* one seeks to identify a $k$-club with the maximum cardinality; this problem is known to be *NP*-hard for any fixed strictly positive integer $k$; see Balasundaram, Butenko, and Trukhanov (2005) and Pajouh, Balasundaram, and Hicks (2016).

For 2-clubs every pair of non-adjacent vertices has a common neighbor. This property, often referred as the "2-hop" property, is important in many transportation and communication settings. For example, a star graph, i.e., a graph with one designated vertex as a "hub" connected to all other vertices with no additional connections between them, is a well-known example of a 2-club arising in a variety of real-life applications (Pajouh et al., 2016). This observation provides a straightforward motivation for an efficient greedy heuristic for finding large 2-clubs (and general $k$-clubs, as any 2-club is also a $k$-club, $k \geq 3$); specifically, the heuristic simply selects a maximum degree vertex and its neighborhood, see, e.g., Bourjolly, Laporte, and Pesant (2000). Such an approach is extremely effective in sparse real-life graphs as it is likely to return an optimal solution (see our further discussion in Section 4). However, it is also known that it is *NP*-hard to determine whether there exists a 2-club of a strictly larger size than the one constructed in this greedy manner (Kahruman-Anderoglu, Buchanan, Butenko, & Prokopyev, 2016).

Furthermore, 2-clubs have a variety of natural interpretations in social network contexts (Laan, Marx, & Mokken, 2016; Mokken, 1979); one can recall, for example, a well-known "a friend of a friend" concept widely used in social sciences, see, e.g., Goodreau, Kitts, and Morris (2009). Similarly, 2-clubs can be exploited for clustering in data mining as the "2-hop" property may also reflect some underlying relationships and/or similarity between objects in a given dataset (e.g., citation of the same document by two different documents in text analytics and web mining applications); see examples in Jia et al. (2018), Miao and Berleant (2004) and Terveen, Hill, and Amento (1999).

Generally speaking, 2-clubs form perhaps the "simplest" distance-based relaxation of a clique as the diameter of one is a *clique-defining property*. That is, a graph is a clique if and only if its diameter is equal to one. By "simplest" we imply that this subgraph/cluster model is, in a sense, the easiest and most intuitive to justify in many real-life contexts. In view of the above discussion, it is not surprising that from the analytical and computational perspectives the case of $k = 2$ is the most well-studied class of the maximum $k$-club problem, see, e.g., Carvalho and Almeida (2011), Hartung, Komusiewicz, and Nichterlein (2015), Hartung, Komusiewicz, Nichterlein, and Suchỳ (2015), Komusiewicz, Nichterlein, Niedermeier, and Picker (2019), Laan, Intelligentie, Marx, Mokken, and van Doornik (2012) and Pajouh et al. (2016) and the references therein.

In particular, the maximum 2-club problem admits a simple integer programming (IP) formulation, which is a straightforward generalization of the classical maximum clique IP model; see Balasundaram et al. (2005), Bourjolly, Laporte, and Pesant (2002), Veremyev and Boginski (2012a). Namely, there is a binary variable for each vertex to indicate whether a vertex is in a feasible solution (2-club); then for every pair of non-adjacent vertices (i.e., there is no edge between them) we enforce with a constraint that both vertices can be in a feasible 2-club only if they have a common neighboring vertex, which is also in the 2-club. We overview this IP model, referred to as **F2s** throughout the paper (this notation stands for "Formulation for 2-club, standard"), in Section 2.1. Polyhedral properties of this classical formulation are studied in Pajouh et al. (2016); its modeling generalizations for 3-clubs and general $k$-clubs are considered in Almeida and Carvalho (2012) and Balasundaram et al. (2005); Bourjolly et al. (2002); Buchanan and Salemi (2020); Veremyev and Boginski (2012a), respectively.

The *first contribution* of this paper is to propose two new mixed-integer programming (MIP) formulations for the maximum 2-club problem that are based on a somewhat different modeling interpretation of the problem. In particular, these two new formulations have slightly more (continuous) variables than the standard model, however, they require substantially fewer constraints for sparse graphs. We describe these models, referred to as **F2c1** and **F2c2** throughout the paper (that is, "Formulation for 2-club, compact"), in Section 2.2. We formally establish that these two new formulations have much better quality of their linear programming (LP) relaxations than the standard **F2s** model for graphs with a sufficiently small number of pairs of vertices that are within the distance of at most 2 from each other. Note that the latter is often the case in practice as real-life graphs are typically very sparse.

We perform computational experiments with an off-the-shelf commercial MIP solver (Gurobi Optimization, 2019) to explore the performance of our new models, **F2c1** and **F2c2**, against the classical model **F2s**. Our computational study provides *three* interesting observations. (*i*) In contrast to **F2s**, the new models (in particular, **F2c2**) provide much higher quality LP relaxation bounds for randomly generated and real-life graphs, which is consistent with our aforementioned theoretical results. (*ii*) When the presolve (i.e., a collection of various preprocessing routines implemented in Gurobi) is switched off, then the new MIP models substantially outperform **F2s** with respect to their running times when solving the problem to optimality. (*iii*) When the presolve is used, then the performance of **F2s** improves substantially; however, the performance of **F2c1** and **F2c2** is either not affected or deteriorates. This observation implies that the constraint structure of **F2s** is more amenable to preprocessing routines implemented in the Gurobi MIP solver. Hence, **F2s** remains a viable alternative for solving the problem in reasonably sized graphs when using solvers with advanced MIP presolve implementations. It also opens up an interesting avenue for future research to explore both computationally and, perhaps, theoretically the reasons behind the latter two observations.

Furthermore, we consider feasibility versions of these two new MIPs that verify whether there exists a 2-club of size exactly $\ell \in \mathbb{N}$ in a graph. Then we incorporate them into an easy-to-implement "feasibility-check" algorithm, see the details in Section 3, that solves one of these feasibility MIPs for each integer $\ell$, by iteratively decreasing it from some upper bound. The bound is derived by solving the LP relaxations of the new MIPs; hence, their tightness (observed both experimentally and theoretically) is critical for the computational performance of the algorithm. In our experiments this extremely simple method, which we view as the *second contribution* of the paper, outperforms **F2s** (even when the presolve is used) for large graphs with 4,000-10,000 vertices. It is important to point out that this approach does not require any sophisticated implementation.

Our *third contribution* is based on the following intuitive property of many real-life graphs. Namely, it is known from the literature, see, e.g., Hartung et al. (2015); Komusiewicz et al. (2019), and also observed in our computational experiments, that in real-life graphs any vertex with the maximum degree and its neighbors (i.e., adjacent vertices) is often a maximum 2-club. Thus, on the one hand, from the practical perspective in many real-life graphs it is rather easy to find a maximum 2-club. On the other hand, it is much more difficult to verify the global optimality of such "star-like" solutions.

Moreover, the 2-clubs identified in this manner typically do not form attack-tolerant (or fault-tolerant) subgraphs as they contain multiple vertices of degree one. Leaf vertices can be disconnected from the subgraph by removing the corresponding edges; the latter can be viewed as edge failures, e.g., due to an adversarial attack. Furthermore, such subgraphs are also susceptible to vertex failures, as they easily become disconnected (with multiple connected components), when the "hub" vertex (i.e., the vertex with the maximum degree) is unavailable/removed/failed. In this paper we consider two "robust" generalizations from Komusiewicz et al. (2019); Veremyev and Boginski (2012a,b) that address these issues, namely, the maximum 2-club problem with some specified minimum vertex degree requirement and the maximum $R$-robust 2-club problem; see Section 4. Our *(third) contribution* is that we show how our MIP models and the "feasibility-check" algorithm can be adapted to handle these two "robust" versions of the maximum 2-club problem in a fairly straightforward manner.

Finally, we summarize our paper in Section 5, where we also outline promising directions for future research.

## 2. Integer programming models

We first briefly describe the standard IP model, see Balasundaram et al. (2005), Bourjolly et al. (2002), Veremyev and Boginski (2012a), in Section 2.1. Then in Section 2.2 we introduce two new MIP models. In Section 2.3 we study the LP relaxation quality of the models and establish that for sufficiently sparse graphs our MIPs are superior to the standard IP. Finally, in Section 2.4 we perform computational experiments to support our theoretical observations using a set of randomly generated and real-life graph instances.

### 2.1. Standard formulation

Denote by $N(i)$ the neighborhood of node $i \in V$, i.e., $N(i) = \{j \in V \mid (i,j) \in E\}$. Let $x_i$, $i \in V$, be a 0–1 variable such that $x_i = 1$ if and only if vertex $i$ is in a maximum 2-club. Then the standard IP model for solving the maximum 2-club problem is given by:

$$[\mathbf{F2s}]: \quad \bar{\omega}_2 := \max \sum_{i \in V} x_i \tag{1a}$$

subject to

$$x_i + x_j - \sum_{t \in N(i) \cap N(j)} x_t \le 1 \qquad \forall (i,j) \notin E, \tag{1b}$$

$$x_i \in \{0,1\} \qquad \forall i \in V, \tag{1c}$$

where constraint (1b) ensures that a pair of non-adjacent vertices $i$ and $j$ can be simultaneously in an optimal solution only if there exists another vertex $t$ in their common neighborhood that is also in the optimal solution. The number of variables in the formulation is $|V|$; the number of constraints is $\frac{|V|(|V|-1)}{2} - |E|$, which is $\Theta(|V|^2)$ for sparse graphs. A recent detailed study of the polyhedral properties of this model can be found in Pajouh et al. (2016).

### 2.2. New formulations

The key idea behind our formulations is to completely avoid considering vertex pairs $i,j$ with $d_G(i,j) \ge 3$. We achieve this goal by introducing a new set of variables $u_{ij}$, where $u_{ij}$ can be set to 1 if and only if both vertices $i$ and $j$ are selected to be in a 2-club and the distance between them in a 2-club is at most 2. Let

$$E_2 = \{(i,j) \in \binom{V}{2} \mid d_G(i,j) \le 2\} \tag{2}$$

be the set of all vertex pairs of graph $G$ with distance at most 2 from each other. Thus, for this new set of variables it is sufficient to consider only indices in this set, i.e., $(i,j) \in E_2$. Furthermore, we treat pairs $(i,j) \in E_2$ and $(j,i) \in E_2$ as the same and hence, we assume that the corresponding variables $u_{ij}$ and $u_{ji}$ coincide.

Our first formulation is based on the observation that for any vertex $i \in V$, if it is in a 2-club, the number of vertex pairs that share 2-club membership with that vertex should be the size of the 2-club minus one. Therefore, we have the following model:

$$[\mathbf{F2c1}]: \quad \bar{\omega}_2 = \max \sum_{i \in V} x_i \tag{3a}$$

subject to

$$\sum_{j \in V:\ (i,j) \in E_2} u_{ij} \ge \sum_{j \in V} x_j - 1 - \mu(1 - x_i) \qquad \forall i \in V, \tag{3b}$$

$$u_{ij} \le x_i,\ u_{ij} \le x_j \qquad \forall (i,j) \in E_2, \tag{3c}$$

$$u_{ij} \le \sum_{t \in N(i) \cap N(j)} x_t \qquad \forall (i,j) \in E_2,\ d_G(i,j) = 2, \tag{3d}$$

$$u_{ij} \ge 0,\ \ x_i \in \{0,1\} \qquad \forall (i,j) \in E_2,\ i \in V, \tag{3e}$$

where $\mu$ is a sufficiently large constant, so that if $x_i = 0$ then constraint (3b) is inactive, e.g., $\mu$ is an upper bound on the size of the 2-club minus one. We derive a sharp upper bound on the size of the 2-club in Section 2.3. Alternatively, we can simply set $\mu = |V| - 1$.

Note that by (3c) and (3e) each $u_{ij} \in [0,1]$. Furthermore, if the right-hand sides of (3c) and (3d) do not enforce $u_{ij}$ to be equal to zero, then the left-hand side of (3b) requires the value of $u_{ij}$ to be equal to one. Therefore, we do not need to enforce binary restrictions for variables $u_{ij}$ in the resulting MIP model.

Our second model is based on the idea that in order to have a 2-club of size $\ell$, there should be exactly $\ell(\ell-1)/2$ variables $u_{ij}$ that can be set to one. This requirement can be enforced using a classical value-disjunction technique from integer programming (see, e.g., Vielma (2015)) on the 2-club size; a somewhat similar approach is used for modeling the maximum quasi-clique problem in Veremyev, Prokopyev, Butenko, and Pasiliao (2016) and the maximum clique problem in Martins (2010)).

Specifically, we define new binary variables $z_\ell \in \{0,1\}$ for all $\ell \in \{1,\dots,n\}$ such that $z_\ell = 1$ if and only if $\sum_{i \in V} x_i = \ell$. Therefore, we have the following model:

$$[\mathbf{F2c2}]: \quad \bar{\omega}_2 = \max \sum_{i \in V} x_i \tag{4a}$$

subject to

$$\sum_{(i,j) \in E_2} u_{ij} \ge \sum_{\ell = \beta_2^l}^{\beta_2^u} \frac{\ell(\ell-1)}{2} z_\ell, \tag{4b}$$

$$u_{ij} \le x_i,\ u_{ij} \le x_j \qquad \forall (i,j) \in E_2, \tag{4c}$$

$$u_{ij} \le \sum_{t \in N(i) \cap N(j)} x_t \qquad \forall (i,j) \in E_2,\ d_G(i,j) = 2, \tag{4d}$$

$$\sum_{\ell = \beta_2^l}^{\beta_2^u} \ell z_\ell = \sum_{i \in V} x_i, \quad \sum_{\ell = \beta_2^l}^{\beta_2^u} z_\ell = 1, \tag{4e}$$

$$u_{ij} \ge 0,\ \ x_i \in \{0,1\} \qquad \forall (i,j) \in E_2,\ i \in V, \tag{4f}$$

$$z_\ell \in \{0,1\} \qquad \forall \ell \in \{\beta_2^l, \ldots, \beta_2^u\}, \tag{4g}$$

where $\beta_2^l$ and $\beta_2^u$ are some lower and upper bounds on the size of a maximum 2-club in $G$, respectively. As outlined in our discussion in Section 1 we can set

$$\beta_2^l = \max_{i \in V} deg_G(i) + 1, \tag{5}$$

where $deg_G(i)$ denotes the degree of vertex $i$ in $G$, i.e., $deg_G(i) = |N(i)|$. The value of $\beta_2^u$ can be set to $|V|$; however, as stated above, another sharp upper bound on the size of the 2-club proposed in Section 2.3 can be used.

As mentioned earlier, a constraint similar to (4b) is also used in Veremyev et al. (2016) for modeling the maximum quasi-clique problem. The key difference is in defining the left-hand side of (4b). In Veremyev et al. (2016), the left-hand side of (4b) "counts" the number of edges in the subgraph, i.e., a quasi-clique, which is a density-based clique relaxation; see Pattillo et al. (2013b). However, in **F2c2** the left-hand side of (4b) represents the number of vertex pairs that are within distance 2 in a subgraph (i.e., a 2-club), which requires our definition of set $E_2$ and constraints (4c) and (4d).

The ideas behind the construction of both MIPs **F2c1** and **F2c2** are rather intuitive. We next formally state that the formulations are valid; see the detailed proof in the Appendix.

**Proposition 1.** *The largest 2-club in a graph G is of size $\ell^*$ if and only if the optimal objective values of the formulations **F2c1** and **F2c2** are equal to $\ell^*$.*

Furthermore, in **F2c2** the binary restrictions for $z_\ell$ can be relaxed, i.e., (4g) can be replaced by simply having:

$$z_\ell \geq 0 \qquad \forall \ell \in \{\beta_2^l, \ldots, \beta_2^u\}, \tag{6}$$

which is formally stated as follows.

**Proposition 2.** *There exists an optimal solution $\mathbf{x}^*, \mathbf{u}^*, \mathbf{z}^*$ of formulation **F2c2** with binary restrictions for variables $\mathbf{z}$ relaxed such that $\mathbf{z}^*$ is a binary vector.*

The proof of this result (relegated to the Appendix to streamline our discussion) is based on a rather standard application of Jensen's inequality whenever value-disjunction reformulation ideas are used; see a similar derivation in Proposition 1 in Veremyev et al. (2016).

Finally, we should point out that in both models **F2c1** and **F2c2** the total number of variables is $\Theta(|V| + |E_2|)$; the number of constraints is $O(|V| + |E_2|)$. Note that in sparse real-life graphs it is typically the case that $|E_2| \ll |V|^2$. Thus, one should expect that for many real-life graphs the numbers of constraints in **F2c1** and **F2c2** are much smaller than $\binom{|V|}{2} - |E|$, the number of constraints in **F2s**.

### 2.3. Sharp upper bound and LP relaxation analysis

In this section, we derive a sharp upper bound on the maximum 2-club size, which can be used to set the appropriate values for parameters $\mu$ and $\beta_2^u$ in **F2c1** and **F2c2**, respectively. We also explore the quality of the LP relaxations of our new formulations **F2c1** and **F2c2**, and compare them against the LP relaxation of **F2s**.

**Proposition 3** (Upper bound on a 2-club size)**.** *Let $S \subseteq V$ be a 2-club in $G = (V, E)$. Then*

$$|S| \leq \beta_2^u = \left\lfloor \frac{1 + \sqrt{1 + 8|E_2|}}{2} \right\rfloor \tag{7}$$

*and this bound is sharp.*

**Proof.** By definition, for any pair of vertices $i, j \in S$ we have $d_G(i,j) \leq d_{G[S]}(i,j) \leq 2$. Therefore,

$$\frac{|S|(|S|-1)}{2} \leq |E_2|. \tag{8}$$

Solving this quadratic inequality with respect to $|S|$ leads to condition (7). Finally, the sharpness of this bound follows immediately by considering a star graph with $\beta_2^u$ vertices. □

We should note that the derivation idea above (via a quadratic inequality) is somewhat similar in spirit to the approach used for deriving the upper bound on the size of the maximum quasi-clique in a graph; see Pattillo, Veremyev, Butenko, and Boginski (2013a). The key difference is using set $E_2$ instead of $E$; recall our discussion on **F2c2** in Section 2.2.

**Proposition 4** (LP relaxation bounds)**.** *Let $\bar{\omega}_2^s$, $\bar{\omega}_2^{c1}$ and $\bar{\omega}_2^{c2}$ be the optimal objective function values of the LP relaxations of formulations **F2s**, **F2c1** and **F2c2**, respectively, where $\mu = \beta_2^u - 1$ and $\beta_2^u = \left\lfloor \frac{1+\sqrt{1+8|E_2|}}{2} \right\rfloor$. Then the following inequalities hold,*

(i) $\bar{\omega}_2^s \geq \frac{|V|}{2}$
(ii) $\bar{\omega}_2^{c1} \leq \hat{\beta}_2^u := \frac{1+\sqrt{1+8|E_2|}}{2}$
(iii) $\bar{\omega}_2^{c2} \leq \hat{\beta}_2^u := \frac{1+\sqrt{1+8|E_2|}}{2}$,

*where $\beta_2^u = \lfloor \hat{\beta}_2^u \rfloor$ by their definitions.*

**Proof.** (*i*): The inequality follows immediately from the fact that setting $x'_i = 1/2$ for all $i \in V$ is a feasible solution of the LP relaxation of **F2s**.

(*ii*): Let $\mathbf{x}' = (x'_1, \ldots, x'_n)^T$ and $\mathbf{u}' = \{u'_{ij} \mid (i,j) \in E_2\}$ be a feasible solution of the LP relaxation of **F2c1**. Then from constraint (3b) we have:

$$\sum_{j \in V:\ (i,j) \in E_2} u'_{ij} \geq \sum_{j \in V} x'_j - 1 - (\beta_2^u - 1)(1 - x'_i),$$

$$\sum_{j \in V:\ (i,j) \in E_2} u'_{ij} \geq \sum_{j \in V} x'_j - x'_i - \beta_2^u(1 - x'_i),$$

which implies that

$$\sum_{i \in V} \sum_{j \in V:\ (i,j) \in E_2} u'_{ij} \geq \sum_{i \in V} \left( \sum_{j \in V} x'_j - x'_i - \beta_2^u(1 - x'_i) \right)$$

and hence,

$$2 \sum_{(i,j) \in E_2} u'_{ij} \geq |V| \sum_{i \in V} x'_i - \sum_{i \in V} x'_i - \beta_2^u \left( |V| - \sum_{i \in V} x'_i \right).$$

Recall that $\beta_2^u = \lfloor \hat{\beta}_2^u \rfloor$, then

$$\hat{\beta}_2^u \geq \beta_2^u,$$

and thus, from the derivation of Proposition 3, see (8), we have

$$2 \sum_{(i,j) \in E_2} u'_{ij} \leq 2|E_2| \leq \hat{\beta}_2^u(\hat{\beta}_2^u - 1).$$

Next, combining with the previous inequality we have

$$\hat{\beta}_2^u(\hat{\beta}_2^u - 1) \geq (|V| - 1) \sum_{i \in V} x'_i + \beta_2^u \left( \sum_{i \in V} x'_i \right)$$

$$- |V|\beta_2^u \geq (|V| - 1) \sum_{i \in V} x'_i + \hat{\beta}_2^u \left( \sum_{i \in V} x'_i \right) - |V|\hat{\beta}_2^u$$

**Table 1**
Problem sizes and the optimal objective function values $\bar{\omega}_2^s$, $\bar{\omega}_2^{c1}$ and $\bar{\omega}_2^{c2}$ of the LP relaxations of **F2s**, **F2c1** and **F2c2**, respectively, on path graphs with $|V|$ vertices. The maximum 2-club size is $\bar{\omega}_2 = 3$. The best results for the LP relaxation quality are in **bold**.

| $\|V\|$ | $\binom{\|V\|}{2} - \|E\|$ | $\|E_2\|$ | $\hat{\beta}_2^u$ | $\bar{\omega}_2^s$ | $\bar{\omega}_2^{c1}$ | $\bar{\omega}_2^{c2}$ |
|---|---|---|---|---|---|---|
| 5 | 6 | 7 | 4.27 | **3.0** | 3.57 | 3.75 |
| 10 | 36 | 17 | 6.35 | 5.0 | 5.09 | **4.35** |
| 15 | 91 | 27 | 7.87 | 7.5 | 6.50 | **4.55** |
| 100 | 4851 | 197 | 20.36 | 50.0 | 17.38 | **4.93** |
| 1000 | 498,501 | 1997 | 63.70 | 500.0 | 60.43 | **4.99** |
| 10000 | 49,985,001 | 19,997 | 200.48 | ≥5,000.0 | 196.17 | **4.99** |

and, equivalently,

$$(\hat{\beta}_2^u)^2 \geq \hat{\beta}_2^u \sum_{i\in V} x'_i + (|V|-1)\left(\sum_{i\in V} x'_i - \hat{\beta}_2^u\right)$$

which implies that

$$(|V|-1+\hat{\beta}_2^u)\left(\hat{\beta}_2^u - \sum_{i\in V} x'_i\right) \geq 0.$$

Observe that if $\sum_{i\in V} x'_i > \hat{\beta}_2^u$, then the above inequality is violated. Hence, $\sum_{i\in V} x'_i \leq \hat{\beta}_2^u$, which completes the proof of the proposition.

($iii$): Let $\mathbf{x}', \mathbf{u}', \mathbf{z}'$ be a feasible solution of the LP relaxation of **F2c2**. To prove the bound we apply Jensen's inequality (as in the proof of Proposition 2) to convex function $f(\ell) = \ell(\ell-1)/2$. According to Jensen's inequality:

$$\sum_{\ell=\beta_2^l}^{\beta_2^u} \frac{\ell(\ell-1)}{2} z'_\ell \geq \frac{\sum_{\ell=\beta_2^l}^{\beta_2^u} \ell z'_\ell \left(\sum_{\ell=\beta_2^l}^{\beta_2^u} \ell z'_\ell - 1\right)}{2} = \frac{\sum_{i\in V} x'_i\left(\sum_{i\in V} x'_i - 1\right)}{2}$$

Therefore, due to inequality (4b), we have:

$$|E_2| \geq \sum_{(i,j)\in E_2} u'_{ij} \geq \sum_{\ell=\beta_2^l}^{\beta_2^u} \frac{\ell(\ell-1)}{2} z'_\ell \geq \frac{\sum_{i\in V} x'_i\left(\sum_{i\in V} x'_i - 1\right)}{2}$$

Considering only the left- and right-hand sides of the above inequalities and then solving the corresponding quadratic inequality (with respect to $\sum_i x'_i$), as in the proof of Proposition 3, leads to the desired bound. □

Clearly, the developed result implies that the objective function values of the LP relaxations of **F2c1** and **F2c2** are $O(|E_2|^{1/2})$ and thus, they are tighter than the LP relaxation quality of **F2s** for sufficiently sparse graphs, where $|E_2| \ll |V|^2$. However, we would like to point out that none of the formulations dominate the others, in general, which we illustrate with the following example.

Let Path($n$) denote a path graph $G$ with $n$ vertices $V = \{1,\dots,n\}$ and edges $E = \{(i,i+1) \mid i = 1,\dots,n-1\}$. Then for $n \in \{5, 10, 15, 100, 1000, 10000\}$, in Table 1 we report the optimal objective function values $\bar{\omega}_2^s$, $\bar{\omega}_2^{c1}$ and $\bar{\omega}_2^{c2}$ of the LP relaxations of **F2s**, **F2c1** and **F2c2**, respectively. In particular, observe that, for $n \geq 10$ the LP relaxation of **F2c2** is better than the others. However, for $n = 5$ it is the worst one; furthermore, the standard model **F2s** provides the best LP relaxation bound. A real-life example, where **F2s** has the best LP relaxation bound is given in Section 2.4.3; see network **celegans** in Table 3.

As a side note it can be mentioned that to improve the performance of the solver when solving the formulation **F2c1** and its LP relaxation for large instances, the term $\sum_{j\in V} x_j$ in the right-hand side of constraint (3b) can be replaced by a new variable, say $v$, with the corresponding addition to the model of an extra constraint $v = \sum_{j\in V} x_j$. This simple modification preserves the correctness of the model, but makes the constraints matrix more sparse.

### *2.4. Computational study: Comparison of the MIP models without presolve*

#### *2.4.1. Preliminaries*

The computational experiments were conducted on an HP machine equipped with Windows 7 x64 operating system, an Intel Core i7-3520M processor (CPU 2.90 GHz, 2 Cores) and RAM 8 GB. All MIP models are solved using Gurobi Optimizer 8.1 (Gurobi Optimization, 2019) using Python 3.7 interface, and NetworkX (Hagberg, Swart, & S Chult, 2008) library to handle networks. The Gurobi parameters are kept at their default values except the time limit and presolve level, which were set to 50,000 seconds and 0 (off), respectively.

With respect to the latter, it should be pointed that the MIP presolve, i.e., a collection of various preprocessing routines (Achterberg, Bixby, Gu, Rothberg, & Weninger, 2020), implemented in MIP solvers are software specific. Hence, in this section we first explore the MIP models without presolve to have an idea about potential performance of the models with other MIP solvers (including open source ones). On the other hand, the presolve implemented in Gurobi MIP solver significantly improves the performance of the **F2s** model. Hence, we discuss this issue in Section 2.4.3 and provide additional computational experiments with the presolve set to the default value (automatic) in Section 3.1.2.

#### *2.4.2. Test instances*

We use both real-life and randomly generated network instances. We focus on a subset of various sparse real-life networks obtained from different application domains, in particular, those, where the considered problem may have some meaningful interpretation. If the original network is disconnected, then we consider (and report parameters for) its largest connected component. This set of real-life networks contains the following instances (all instances including the additional ones considered in Section 3.1.2 are also available at http://www.pitt.edu/~droleg/files/2-clubs.html):

- **bcspwr04** ($|V| = 274$, $|E| = 669$): A representation of a U.S. power network from Davis and Hu (2011).
- **bus_494** ($|V| = 494$, $|E| = 586$), **bus_662** ($|V| = 662$, $|E| = 906$), **bus_1138** ($|V| = 1138$, $|E| = 1458$) : Bus power systems (Davis & Hu, 2011).
- **cables** ($|V| = 429$, $|E| = 636$): A network adopted from the Greg's Cable Map (http://www.cablemap.info), which represents "the undersea communication infrastructure," and obtained from Nguyen, Shen, and Thai (2013); Shen, Nguyen, Xuan, and Thai (2013).
- **celegans** ($|V| = 453$, $|E| = 2025$): Metabolic network of C.elegans (Davis & Hu, 2011; DIMACS, 2011).
- **diseasome** ($|V| = 516$, $|E| = 1188$): The human disease network (Goh et al., 2007; Rossi & Ahmed, 2015)
- **erdos971** ($|V| = 429$, $|E| = 1312$): Erdős collaboration network, see Davis and Hu (2011).
- **HarvardWeb** ($|V| = 500$, $|E| = 2043$): Web connectivity matrix (Davis & Hu, 2011).
- **homer** ($|V| = 542$, $|E| = 1619$): A social network of Homer's "Iliad," see Graph Coloring and its Generalizations (2004).
- **LindenStrasse** ($|N| = 232$, $|E| = 303$): A social network of the German soap opera "Lindenstrasse" (Batagelj & Mrvar, 2006).
- **netscience** ($|V| = 379$, $|E| = 914$): A collaboration network in network science (Batagelj & Mrvar, 2006; Davis & Hu, 2011).
- **USAir97** ($|V| = 332$, $|E| = 2126$): An airline transportation network (Davis & Hu, 2011).

In addition, we consider two classes of randomly generated graph instances:

**Table 2**
Comparison of the number of constraints, the optimal objective function values of the LP relaxations and the solution times (when solved to optimality) for formulations **F2s**, **F2c1** and **F2c2** on randomly generated graphs. The average values over 10 random instances are reported. The best LP relaxation objective functions and the best running times are in **bold**.

| | | | | | | | LP rlx. objective | | | Time (seconds) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Graph | $\|V\|$ | $\|E\|$ | $\binom{\|V\|}{2} - \|E\|$ | $\|E_2\|$ | $\bar{\omega}_2$ | $\hat{\beta}_2^u$ | **F2s** | **F2c1** | **F2c2** | **F2s** | **F2c1** | **F2c2** |
| **WS** graphs | | | | | | | | | | | | |
| **WS100** | 100 | 200 | 4750 | 630.2 | 7.6 | 36.0 | 50 | 29.8 | **13.7** | **0.7** | 1.3 | 1.5 |
| **WS200** | 200 | 400 | 19,500 | 1254.5 | 7.9 | 50.6 | 100 | 43.3 | **13.6** | 11.8 | **10.0** | 11.2 |
| **WS300** | 300 | 600 | 44,250 | 1832.6 | 7.6 | 61.0 | 150 | 53.3 | **13.3** | 70.4 | **25.7** | 51.3 |
| **WS400** | 400 | 800 | 79,000 | 2521.8 | 8.1 | 71.5 | 200 | 63.2 | **13.7** | 253.3 | **54.6** | 172.6 |
| **WS500** | 500 | 1000 | 123,750 | 3148.4 | 8.2 | 79.9 | 250 | 71.3 | **13.6** | 746.2 | **121.7** | 220.4 |
| **WS600** | 600 | 1200 | 178,500 | 3800.4 | 8.3 | 87.7 | 300 | 78.7 | **13.7** | 1755.7 | **218.2** | 518.8 |
| **WS700** | 700 | 1400 | 243,250 | 4419.2 | 8.3 | 94.5 | 350 | 85.4 | **13.7** | 3768.7 | **548.1** | 726.4 |
| **WS800** | 800 | 1600 | 318,000 | 5103.5 | 8.5 | 101.5 | 400 | 92.3 | **13.8** | 7128.7 | **936.3** | 1487.0 |
| **WS900** | 900 | 1800 | 402,750 | 5716.8 | 8.6 | 107.4 | 450 | 98.2 | **13.8** | 12547.0 | **1696.6** | 2477.9 |
| **WS1000** | 1000 | 2000 | 497,500 | 6326.4 | 8.4 | 113.0 | 500 | 103.6 | **13.7** | 20083.4 | **2198.7** | 3268.0 |
| **BA** graphs | | | | | | | | | | | | |
| **BA100** | 100 | 196 | 4754 | 1249.6 | 25.3 | 50.5 | 50 | 40.9 | **28.4** | **0.5** | 2.3 | 0.7 |
| **BA200** | 200 | 396 | 19,504 | 2859.7 | 32.7 | 76.1 | 100 | 62.3 | **35.5** | 7.7 | 19.4 | **1.5** |
| **BA300** | 300 | 596 | 44,254 | 5199.1 | 47.3 | 102.4 | 150 | 84.6 | **49.0** | 53.2 | 79.1 | **2.7** |
| **BA400** | 400 | 796 | 79,004 | 6971.6 | 50.4 | 118.5 | 200 | 99.1 | **52.0** | 199.9 | 203.4 | **4.5** |
| **BA500** | 500 | 996 | 123,754 | 9127.9 | 58.3 | 135.5 | 250 | 114.3 | **59.8** | 618.8 | 449.4 | **6.0** |
| **BA600** | 600 | 1196 | 178,504 | 11517.6 | 64.2 | 152.2 | 300 | 129.1 | **65.9** | 1691.9 | 705.0 | **10.7** |
| **BA700** | 700 | 1396 | 243,254 | 13855.1 | 65.8 | 166.8 | 350 | 142.3 | **67.1** | 3518.8 | 973.3 | **11.1** |
| **BA800** | 800 | 1596 | 318,004 | 15965.6 | 67.9 | 179.1 | 400 | 153.7 | **69.7** | 7147.4 | 1247.6 | **19.4** |
| **BA900** | 900 | 1796 | 402,754 | 18455.4 | 77.3 | 192.5 | 450 | 166.1 | **79.1** | 12925.8 | 2046.7 | **22.7** |
| **BA1000** | 1000 | 1996 | 497,504 | 21317.8 | 91.0 | 206.8 | 500 | 179.2 | **92.1** | 22182.3 | 3259.5 | **73.1** |

- *Watts-Strogatz (**WS**) graphs* are constructed based on the model proposed by Watts and Strogatz (Watts & Strogatz, 1998) and generated using the corresponding function in NetworkX (Hagberg et al., 2008) library. The sampled graphs can be "highly clustered, like regular lattices, yet have small characteristic path lengths" (Watts & Strogatz, 1998). We consider all possible sizes $n$ from 100 to 1000 (with 100 vertex increment). For each $n$ we generate 10 instances and report the average results in the corresponding tables, where the instance sets are labelled as **WS**$n$ ($n$ is the graph size). The number of neighbors in the original ring topology and the edge rewiring probability are set to 4 and 0.15, respectively.
- *Barabási-Albert (**BA**) graphs* are constructed according to Barabási-Albert preferential attachment mechanism (BA model) (Albert & Barabási, 2002) and also obtained using the corresponding function in NetworkX (Hagberg et al., 2008) library. This model is widely used for generating scale-free networks. For each $n$ we also generate 10 instances and report the average results for all instances in the corresponding tables, where the instance sets are labelled as **BA**$n$ ($n$ is the graph size). To approximately match the edge density of **WS** graphs, the number of edges attached to any new vertex is set to 2.

### *2.4.3. Results and discussion*

We first discuss our experiments for the randomly constructed test instances, see Table 2. In particular, we want to point out that both new MIP models **F2c1** and **F2c2** significantly outperform the standard formulation **F2s** with respect to the quality of their LP relaxations. Note that for **F2s** the value of $\bar{\omega}_2^s$ is exactly $|V|/2$ for all randomly generated instances; on the other hand, the LP relaxations of both **F2c1** and **F2c2** are much tighter. These computational observations are consistent with the theoretical results in Proposition 4.

The results for real-life graphs are reported in Table 3. For almost all test instances (except **celegans** where the largest 2-club size is larger than $|V|/2$), both new models **F2c1** and **F2c2** provide a better LP relaxation quality than the **F2s** model, with **F2c2** being typically the best one.

With respect to the solver's running time performance, we, first, recall that in the considered set of experiments the presolve is set 0 (i.e., no preprocessing used) for the MIP solver. In this setting, both new MIP models **F2c1** and **F2c2** significantly outperform the standard formulation **F2s** with respect to the solver's running time for all instances in Table 2. For **WS** graphs model **F2c1** is the best, while **F2c2** is not far behind; for **BA** graphs **F2c2** provides the best performance. For real-life graphs in Table 3 there are only 4 instances (out of 13), namely, **celegans**, **erdos971**, **homer** and **USAir97**, where the standard MIP **F2s** outperforms the new models.

As briefly mentioned earlier, if the presolve is used, then the performance of our new models **F2c1** and **F2c2** is not affected notably, in fact, it often deteriorates. On the other hand, the performance of the **F2s** model improves significantly with respect to the solver's running time. We provide the corresponding running time results and our additional discussion on this issue in Section 3.1.2. We conjecture that the constraint structure of the **F2s** model is more amenable to various preprocessing routines implemented in the (commercial) Gurobi MIP solver than the constraint structures of the proposed MIP models. Further exploration of this issue both from the theoretical and computational perspectives (e.g., comparing different commercial and open source MIP solvers) provides an interesting avenue of further research.

Nevertheless, our empirical and theoretical (recall Proposition 4) results imply that the LP relaxations of **F2c1** and **F2c2** can be used to provide high quality (and polynomially computable) upper bounds for the maximum 2-club problem. Furthermore, as we demonstrate next, the good LP relaxation quality of our new models can be exploited within a simple iteration-based scheme, where a sequence of MIP feasibility models is solved. This new "easy-to-implement" approach turns out to be competitive with **F2s** (even when the presolve is used), and outperforms it for larger graphs. Hence, the proposed simple method allows us to consider much larger real-life graphs (up to 10,000 vertices) in our experiments than those solved by the MIPs in this section; see further details in Section 3.

**Table 3**
Comparison of the number of constraints, the optimal objective function values of the LP relaxations and the solution times (when solved to optimality) for formulations **F2s**, **F2c1** and **F2c2** on real-life graphs, see their description in Section 2.4.2. Time limit (TL) is set to 50,000 sec. The best LP relaxation objective functions and the best running times are in **bold**.

| | | | | | | | LP rlx. objective | | | Time (seconds) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Graph | $\|V\|$ | $\|E\|$ | $\binom{\|V\|}{2}-\|E\|$ | $\|E_2\|$ | $\bar{\omega}_2$ | $\hat{\beta}_2^u$ | **F2s** | **F2c1** | **F2c2** | **F2s** | **F2c1** | **F2c2** |
| **bcspwr04** | 274 | 669 | 36,732 | 2354 | 16 | 69.1 | 137.0 | 58.9 | **22.4** | 45.9 | 22.1 | **6.3** |
| **bus_494** | 494 | 586 | 121,185 | 1784 | 10 | 60.2 | 247.0 | 55.1 | **10.7** | 875.3 | 78.5 | **0.3** |
| **bus_662** | 662 | 906 | 217,885 | 2909 | 10 | 76.8 | 331.0 | 70.7 | **11.6** | 3914.8 | 197.5 | **1.4** |
| **bus_1138** | 1138 | 1458 | 645,495 | 5002 | 18 | 100.5 | 569.0 | 94.4 | **19.0** | TL | 1999.3 | **5.0** |
| **USAir97** | 332 | 2126 | 52,820 | 22,191 | 140 | 211.2 | 166.0 | 172.6 | **163.3** | **27.8** | 119.3 | TL |
| **cables** | 429 | 636 | 91,170 | 2272 | 17 | 67.91 | 214.5 | 60.9 | **17.0** | 419.8 | 55.2 | **0.2** |
| **celegans** | 453 | 2025 | 100,353 | 45,351 | 238 | 301.7 | **238.0** | 254.8 | 240.6 | **0.2** | 773.4 | 538.2 |
| **diseasome** | 516 | 1188 | 131,682 | 5814 | 51 | 108.3 | 258.0 | 94.0 | **53.4** | 959.7 | 130.8 | **5.5** |
| **LindenStrasse** | 232 | 303 | 26,493 | 1225 | 14 | 50.0 | 116.0 | 43.6 | **14.9** | 26.2 | 8.9 | **0.3** |
| **homer** | 542 | 1619 | 144,992 | 21,728 | 100 | 209.0 | 271.0 | 170.3 | **111.6** | **776.5** | 2261.3 | 1612.8 |
| **netscience** | 379 | 914 | 70,717 | 3830 | 35 | 88.0 | 189.5 | 75.6 | **37.4** | 200.4 | 51.3 | **2.4** |
| **erdos971** | 429 | 1312 | 90,494 | 9904 | 42 | 141.2 | 214.5 | 117.2 | **72.5** | **324.9** | 641.4 | TL |
| **HarvardWeb** | 500 | 2043 | 122,707 | 34,255 | 201 | 262.2 | 250.0 | 220.6 | **201.0** | 179.4 | 676.0 | **45.8** |

## 3. “Feasibility-check” algorithm

Next, in Section 3.1 we consider feasibility versions of our MIPs that verify whether there exists a 2-club of size exactly $\ell \in \mathbb{N}$; then we incorporate one of them (more promising in terms of tightness) into an easy-to-implement “feasibility-check” algorithm as outlined. Note that the 2-club property is not hereditary (i.e., a subgraph of a 2-club is not necessarily a 2-club); thus, we need to consider all possible sizes of the maximum 2-club between its lower and upper bounds. This observation implies that good quality lower and upper bounds are extremely important, as a bisection-like scheme (or, at least its naive version) cannot be exploited. The computational experiments with our new approach and its comparison against the MIP solver (with presolve) are provided in Section 3.1.2.

### 3.1. Feasibility MIPs and the algorithm

Consider the feasibility versions of formulations **F2c1** and **F2c2**. Namely, for each possible value $\ell$, the formulations further referred to as **F2c1**($\ell$) and **F2c2**($\ell$), respectively, verify whether there exists a 2-club of size exactly $\ell$. Thus, we obtain:

$$[\mathbf{F2c1}(\ell)]: \quad \sum_{j\in V:\ (i,j)\in E_2} u_{ij} \ge (\ell-1)x_i \quad \forall i\in V, \tag{9a}$$

$$u_{ij} \le x_i, u_{ij} \le x_j \forall (i,j)\in E_2, \tag{9b}$$

$$u_{ij} \le \sum_{t\in N(i)\cap N(j)} x_t \qquad \forall (i,j)\in E_2, d_G(i,j)=2, \tag{9c}$$

$$\sum_{i\in V} x_i = \ell, \tag{9d}$$

$$u_{ij}\ge 0,\ x_i\in\{0,1\} \qquad \forall (i,j)\in E_2, i\in V, \tag{9e}$$

where constraint (9d) enforces the required size of a 2-club. The other model is given by:

$$[\mathbf{F2c2}(\ell)]: \quad \sum_{(i,j)\in E_2} u_{ij} \ge \frac{\ell(\ell-1)}{2}, \tag{10a}$$

$$u_{ij} \le x_i, u_{ij} \le x_j \qquad \forall (i,j)\in E_2, \tag{10b}$$

$$u_{ij} \le \sum_{t\in N(i)\cap N(j)} x_t \qquad \forall (i,j)\in E_2, d_G(i,j)=2, \tag{10c}$$

$$\sum_{i\in V} x_i = \ell, \tag{10d}$$

$$u_{ij}\ge 0,\ x_i\in\{0,1\} \qquad \forall (i,j)\in E_2, i\in V. \tag{10e}$$

Next, we observe that if **u** and **x** form a feasible solution of the LP relaxation of **F2c1($\ell$)** for a given $\ell$, then

$$\sum_{j\in V:\ (i,j)\in E_2} u_{ij} \ge (\ell-1)x_i,$$

and

$$\sum_{i\in V}\sum_{j\in V:\ (i,j)\in E_2} u_{ij} \ge \sum_{i\in V}(\ell-1)x_i = (\ell-1)\sum_{i\in V}x_i = \ell(\ell-1)$$

or, equivalently,

$$2\sum_{(i,j)\in E_2} u_{ij} \ge \ell(\ell-1),$$

which implies that **u** and **x** also provide a feasible solution of the LP relaxation of **F2c2($\ell$)** for the same value of $\ell$. Thus, the tightness of the LP relaxation of **F2c1($\ell$)** is not worse than that of **F2c2($\ell$)**. Therefore, we use **F2c1($\ell$)** as our main MIP feasibility model in our algorithm, which we describe next. The formal pseudo-code is provided in Algorithm 1 .

Specifically, the key idea of the algorithm is to simply verify by solving MIP **F2c1($\ell$)**, whether there exists a 2-club of size *exactly* $\ell$, where $\ell \in \mathbb{N}$ is considered between some lower and upper bounds. For the lower bound we can use $\beta_2^l$ given by (5); see line 4 in Algorithm 1. For the upper bound we can use the best upper bound from those formulations considered in this paper; see line 5 of Algorithm 1. Namely, we can consider the LP relaxations of all three MIP models; recall Table 1 and our discussion at the end of Section 2.3 that, in general, none of them dominates the others.

As mentioned earlier, the 2-club property is not hereditary, i.e., a subgraph of a 2-club is not necessarily a 2-club. For example, a cycle with 5 vertices is a 2-club, but this graph does not contain a 2-club of size 4. Hence, in our algorithm in the worst case we need to solve a feasibility MIP for all possible values of $\ell$ from the upper bound to the lower bound plus one; see lines 13–16 in Algorithm 1.

As we decrease the value of $\ell$, we can stop the procedure whenever a feasible solution is found. One feasible solution is readily available from the lower bound; see lines 6–7 in Algorithm 1. Clearly, the required number of iterations (i.e., feasibility MIPs solved) depends on the difference between the upper and lower bounds used. Therefore, if this number is sufficiently small (which

**Algorithm 1:** Exact "feasibility-check" algorithm (2-club) .

1 **Input:** graph $G = (V, E)$
2 **Output:** maximum 2-club $S \subseteq V$
3 **begin**
4 $LB \leftarrow \beta_2^l := \max_{i \in V} deg(i) + 1$
5 $UB \leftarrow \min\{\lfloor \bar{\omega}_2^s \rfloor, \lfloor \bar{\omega}_2^{c1} \rfloor, \lfloor \bar{\omega}_2^{c2} \rfloor\}$
6 $i^* \leftarrow$ any vertex from $argmax_{i \in V} deg_G(i)$
7 $S \leftarrow \{N(i^*) \cup \{i^*\}\}$
8 $\ell \leftarrow UB$
9 **if** $LB = UB$ **then**
10 return $S$
11 **end**
12 Solve a feasibility MIP, denoted by **MIP**($\ell$), that verifies whether there exists a 2-club of size exactly $\ell$
13 **while MIP**($\ell$) is infeasible and $\ell \geq LB + 2$ **do**
14 $\ell \leftarrow \ell - 1$
15 Solve **MIP**($\ell$)
16 **end**
17 **if MIP**($\ell$) is feasible **then**
18 $S \leftarrow \{i : x_i^* = 1, \forall i \in V\}$, where $\mathbf{x}^*$ is feasible for **MIP**($\ell$)
19 **end**
20 return $S$
21 **end**

is the case if the bounds are tight), then the overall performance of this approach can be expected to be better than simply solving one of the original MIPs, **F2s**, **F2c1** and **F2c2**.

For comparisons with the standard model, **F2s**, in our experiments discussed next, we also consider the proposed algorithm with a feasibility version of the **F2s** model. The latter can be easily constructed by removing the objective function in (1a) and adding a cardinality constraint (9d).

### *3.2. Computational study: "feasibility-check" algorithm and MIP solver with presolve*

#### *3.2.1. Preliminaries*

The computational environment (software and hardware) is the same as in the previous set of experiments, see Section 2.4.1. The algorithm was implemented in Python 3.7. From our experiments with real-life graphs in Section 2.4, see Table 3, we observe that the LP relaxation bound provided by **F2c2** is typically the best one. Thus, in our implementation and the experiments discussed below we compute only $\lfloor \bar{\omega}_2^{c2} \rfloor$ in line 5 of Algorithm 1.

#### *3.2.2. Additional test instances*

The proposed algorithm allows us to consider an additional set of real-life networks with larger sizes, approximately up to 10,000 vertices. This additional set contains the following networks:

**Table 4**
Performance comparisons of the "feasibility-check" algorithms (multiple versions of Algorithm 1) against Gurobi with **F2s** and the presolve option turned on; see column $\mathbf{F2s}_{pre}$. The total solution times for solving the feasibility MIPs in Algorithm 1 (without the time needed for solving the LP relaxation of **F2c2**) are reported in the respective columns. In column "**F2c2**LP" we report the running time for solving the LP relaxation of **F2c2**, which is used as the upper bound (UB) in the algorithm. The total running time of Algorithm 1 is the sum of the running times of **F2c2**LP and one of the feasibility MIPs. The best approach (Algorithm 1 vs. $\mathbf{F2s}_{pre}$) is in **bold**.

| | | | | | | | Time (seconds) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | Formulation | | Algorithm 1 (lines 12–20) | | |
| Graph | $\|V\|$ | $\|E\|$ | $d_{\max}$ | $\bar{\omega}_2$ | $\bar{\omega}_2^{c2}$ | #iter | **F2c2**LP | $\mathbf{F2s}_{pre}$ | $\mathbf{F2s}(\ell)_{pre}$ | $\mathbf{F2c1}(\ell)$ | $\mathbf{F2c1}(\ell)_{pre}$ |
| **bcspwr04** | 274 | 669 | 15 | 16 | 22.44 | 6 | 0.49 | 3.88 | 2.32 | 0.23 | **0.22** |
| **bus_494** | 494 | 586 | 9 | 10 | 10.67 | 0 | **0.36** | 5.03 | - | - | - |
| **bus_662** | 662 | 906 | 9 | 10 | 11.6 | 1 | 0.94 | 7.05 | 2.15 | 0.05 | **0.04** |
| **bus_1138** | 1138 | 1458 | 17 | 18 | 19 | 1 | 1.09 | 6.93 | 5.29 | 0.1 | **0.05** |
| **USAir97** | 332 | 2126 | 139 | 140 | 163.28 | 23 | 30.24 | **1.52** | 19.72 | 4.19 | 2.31 |
| **cables** | 429 | 636 | 16 | 17 | 17 | 0 | **0.5** | 1.22 | - | - | - |
| **celegans** | 453 | 2025 | 237 | 238 | 240.59 | 2 | 118.58 | **4.38** | 4.69 | 0.84 | 0.39 |
| **diseasome** | 516 | 1188 | 50 | 51 | 53.42 | 2 | 1.91 | **1.33** | 2.6 | 0.13 | 0.1 |
| **LindenStrasse** | 232 | 303 | 13 | 14 | 14.88 | 0 | **0.13** | 0.41 | - | - | - |
| **homer** | 542 | 1619 | 99 | 100 | 111.6 | 11 | 20.75 | **3.21** | 22.5 | 1.93 | 1.26 |
| **netscience** | 379 | 914 | 34 | 35 | 37.38 | 2 | 0.86 | 1.03 | 1.26 | **0.09** | **0.09** |
| **erdos971** | 429 | 1312 | 41 | 42 | 72.51 | 30 | 5.62 | **1.68** | 35.03 | 197.15 | 277.95 |
| **HarvardWeb** | 500 | 2043 | 200 | 201 | 201.03 | 0 | 80.87 | **3.19** | - | - | - |
| **WS1000_1** | 1000 | 2000 | 7 | 8 | 13.81 | 5 | 4.28 | 55.48 | 149.87 | 212.03 | **0.64** |
| **WS1000_2** | 1000 | 2000 | 8 | 9 | 13.7 | 4 | 4.86 | 35.65 | 33.94 | 0.4 | **0.24** |
| **BA1000_1** | 1000 | 1996 | 107 | 108 | 115.75 | 7 | 33.03 | **15.78** | 61.85 | 2.41 | 0.88 |
| **BA1000_2** | 1000 | 1996 | 69 | 70 | 72.33 | 2 | 47.97 | **13.89** | 16.88 | 0.61 | 0.21 |
| **cerevisae** | 1458 | 1948 | 56 | 57 | 57 | 0 | **7.79** | 14 | - | - | - |
| **human-protein** | 1615 | 3106 | 95 | 96 | 117.42 | 21 | 78.82 | **32.35** | 554.87 | 7.79 | 5.24 |
| **yeast** | 2224 | 6609 | 64 | 65 | 118.4 | 53 | 252.38 | 665.18 | 3273.88 | 6500.71 | **141.72** |
| **bible-nouns** | 1707 | 9059 | 364 | 365 | 406.08 | 41 | 1823.08 | **81.25** | 3665.99 | 87.35 | 53.88 |
| **hamster** | 1788 | 12476 | 272 | 273 | 374.71 | 101 | 2375.67 | **80.69** | 9807.85 | 274.5 | 208.59 |
| **hamster-full** | 2000 | 16098 | 273 | 274 | 382.57 | 108 | 2843.01 | **92.1** | 13715.75 | 316.57 | 248.36 |
| **Geom** | 3621 | 9461 | 102 | 103 | 158.36 | 55 | 406.9 | 506.4 | 25946.06 | 62.65 | **47.81** |
| **GR-QC** | 4158 | 13422 | 81 | 82 | 115.28 | 33 | 234.44 | 3519.44 | 26253.4 | 25.98 | **17.41** |
| **Erdosh02** | 6927 | 11850 | 507 | 508 | 508 | 0 | **5884.51** | 10612.41 | - | - | - |
| **HighEnergy** | 8638 | 24806 | 65 | 66 | 119.8 | 53 | 2024.84 | TL | TL | 217.62 | **116.35** |
| **US_Power** | 4941 | 6594 | 19 | 20 | 21.07 | 1 | 8.71 | 12689.87 | 1392.47 | 0.37 | **0.16** |
| **PGPgiantcompo** | 10680 | 24316 | 205 | 206 | 225.07 | 19 | 1050.29 | ML | ML | 80.94 | **29.44** |

Notes: For Algorithm 1, we consider its three versions with different feasibility MIPs: **F2s**($\ell$) with presolve, and **F2c1**($\ell$) with and without presolve. If the presolve is used, then it is denoted as subscript "*pre*" in an MIP. Column "#iter" contains the number of iterations, i.e., the number of feasibility MIPs solved, given by $\lfloor \bar{\omega}_2^{c2} \rfloor - d_{\max} - 1$, where $d_{\max} = \max_{i \in V} deg_G(i)$ denotes maximum degree in a graph. Symbol "-" implies that Algorithm 1 does not need to solve any feasibility MIP as the upper bound provided by the LP relaxation of **F2c2** coincides with the available lower bound; hence, the required number of iterations is 0.

- **cerevisae** ($|V| = 1458$, $|E| = 1948$): Protein-protein interactions network in yeast Saccharomyces cerevisiae (Balasundaram et al., 2005; DIMACS, 2011).
- **human-protein** ($|V| = 1615$, $|E| = 3106$): Human protein (stelzl) network dataset (Kunegis, 2013). The network represents interacting pairs of proteins in Humans (Homo sapiens).
- **yeast** ($|V| = 2224$, $|E| = 6609$): Protein-protein interaction network in yeast (Batagelj & Mrvar, 2006).
- **bible-nouns** ($|V| = 1707$, $|E| = 9059$): The lexical network of nouns of the King James Version of the Bible; an edge indicates that two nouns appeared together in the same verse (Kunegis, 2013).
- **hamster** ($|V| = 1788$, $|E| = 12476$): The network containing friendships between users of the website hamsterster.com (Kunegis, 2013).
- **hamster-full** ($|V| = 2000$, $|E| = 16098$): The network containing friendships and family links between users of the website hamsterster.com (Kunegis, 2013)
- **Geom** ($|V| = 3621$, $|E| = 9461$): Collaboration network in computational geometry (Batagelj & Mrvar, 2006).
- **GR-QC** ($|V| = 4941$, $|E| = 13422$): Collaboration network of Arxiv General Relativity (Rossi & Ahmed, 2015).
- **Erdos02** ($|V| = 6927$, $|E| = 11850$): Erdős collaboration network from Davis and Hu (2011).
- **HighEnergy** ($|V| = 8638$, $|E| = 24806$): Collaboration network of Arxiv High Energy Physics (Rossi & Ahmed, 2015).
- **US_Power** ($|V| = 4941$, $|E| = 6594$): A network representing US power grid from Davis and Hu (2011).
- **PGPgiantcompo** ($|V| = 10680$, $|E| = 24316$): The giant component of the network of users of the Pretty-Good-Privacy algorithm for secure information interchange (Davis & Hu, 2011; Rossi & Ahmed, 2015) (compiled by Boguná, Pastor-Satorras, Díaz-Guilera, & Arenas (2004)).

*3.2.3. Results and discussion*

As pointed out in Section 2.4 if the MIP presolve (in Gurobi) is used, then the performance of the **F2s** model improves significantly. Hence, in our experiments next we compare the performance of the "feasibility-check" algorithm against the MIP solver with **F2s** model under the default MIP presolve setting. Furthermore, we consider three versions of Algorithm 1, where in each iteration we solve either a feasibility MIP **F2s**($\ell$), or **F2c1**($\ell$), and for the latter, we consider two presolve settings (default and off).

The results for the first set of our experiments with two sets of real-life graphs (see Sections 2.4.2 and 3.1.2 for their detailed descriptions) and 4 additional randomly generated graphs (the largest from those considered in Section 2.4.2) are provided in Table 4. We want to point out the following observations:

- The MIP preprocessing routines implemented in Gurobi, i.e., presolve, significantly improve the running time performance (when solving to optimality) of the **F2s** model; compare the results for **F2s** in Tables 3 and 4. On the other hand, in our experiments the running time performance of our new models **F2c1** and **F2c2** either does not change, or deteriorates for all instances in Table 3 when the presolve is used. Hence, the corresponding results for the latter MIP models (with the presolve) are omitted from Table 4 for brevity.
- Comparing the results of **F2c2** in Table 3 against **F2s** in Table 4, we observe that **F2s** becomes either competitive or outperforms **F2c2**. In fact, **F2c2** (without presolve) slightly outperforms **F2s** (with presolve) only for **bus_494**, **bus_662**, **bus_1138**, **cables**, **LindenStrasse**, while **F2s** (with presolve) is better for the remaining 8 instances.
- The upper bound provided by the LP relaxation of **F2c2** remains relatively good even for the larger networks with several thousand vertices; see column $\bar{\omega}_2^{c2}$ of Table 4. In fact, there are several real-life instances for which there is no need to solve the feasibility MIPs (as indicated by 0 in the column "#iter" of Table 4), as the upper bound provided by the LP relaxation of **F2c2** coincides with the lower bound. In particular, we refer to networks **bus_494**, **cables**, **LindenStrasse**, **HarvardWeb**, **cerevisae** and **Erdos02**, where the latter contains about 7000 vertices.
- Our simple "feasibility-check" algorithm (specifically, using a feasibility version of **F2c1** under both considered presolve options) is competitive against the MIP solver with the **F2s** model under the default MIP presolve setting for larger instances; see the bottom of Table 4. In particular, note that Algorithm 1 (with a feasibility version of **F2c1**) outperforms **F2s** for our 6 largest test instances (**Geom**, **GR-QC**, **Erdos02**, **HighEnergy**, **US_Power** and **PGPgiantcompo**). The total running time of Algorithm 1 is the sum of the running times for solving the LP relaxation of **F2c2** and the corresponding feasibility MIPs. In Table 4 we denote in **bold** the best solution approach in bold that is, either one of the versions of Algorithm 1 or simply solving model **F2s** via the MIP solver; note that for the former whenever the feasibility MIP does not need to be solved, we mark in bold the solution time of the LP relaxation.
- In Table 4 when comparing the running time needed for solving multiple feasibility MIPs (i.e., lines 12–20 in Algorithm 1) and the solution time of MIP **F2s** we observe that the performance of Algorithm 1 could be potentially improved by having better upper bounding schemes, in particular for larger graphs. That is, it could be an interesting avenue for future research to explore upper bounding schemes that can be computed faster than solving the LP relaxation of **F2c2** to optimality (as in our computations).
- The "feasibility-check" algorithm using a feasibility version of **F2s** is not competitive.

The above observations emphasize high-quality of the LP relaxation of **F2c2**. Hence, the proposed MIP models **F2c1** and **F2c2** (or, at least their LP relaxations) can be exploited for the development of more advanced solution methods for the maximum 2-club problem, and further research in this direction seems to be promising.

Next, recall from our discussion in Section 1 that a maximum degree vertex and its neighborhood form a 2-club, which often turns out to be an optimal solution. Hence, if a maximum 2-club contains a leaf vertex (i.e., a vertex with degree one), then this maximum 2-club should be a maximum degree vertex with its neighbours. Otherwise, all leaf vertices can be removed from consideration and only the 2-core (i.e., a maximum subgraph such that the degree of any vertex is at least 2) of the initial graph needs to be considered. Hence, we need to simply compare and pick as an optimal solution either a 2-club formed using a maximum degree vertex (and its neighbourhood), or a maximum 2-club found in the 2-core of the original graph. We provide additional details on this preprocessing approach in a more general setting in Section 4.1.

In Table 5, we explore the same solution methods as in Table 4, after the outlined preprocessing idea is applied. That is, we consider only the 2-cores of the same graphs as in Table 4. Furthermore, in Table 5 we do not consider graphs that do not contain leaf vertices (i.e., they are 2-cores by themselves), and graphs, for which the upper bound provided by the LP relaxation of **F2c2** is sharp (i.e., there is no need to solve feasibility MIPs). Our observations from Table 5 are fairly consistent with those made for Table 4 in our earlier discussion.

**Table 5**
Performance comparisons of the "feasibility-check" algorithms (multiple versions of Algorithm 1) against Gurobi with **F2s** and the presolve option turned (see column **F2s**$_{pre}$), after additional preprocessing is used for both methods. The size of the remaining 2-core (after preprocessing) is reported in the column "2-core". In this set of experiments we do not consider graphs that do not contain leaf vertices, and graphs for which the upper bound provided by the LP relaxation of **F2c2** is sharp.

| | | | | | | | Time (seconds) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2-core | | | | | | Formulation | | Algorithm 1 (lines 12–20) | | |
| Graph | $\|V\|$ | $\|E\|$ | $d_{\max}$ | $\bar{\omega}_2$ | $\bar{\omega}_2^{c2}$ | #iter | **F2c2**LP | **F2s**$_{pre}$ | **F2s**$(\ell)_{pre}$ | **F2c1**$(\ell)$ | **F2c1**$(\ell)_{pre}$ |
| **bcspwr04** | 217 | 612 | 15 | 16 | 22.21 | 6 | 0.53 | 2.27 | 1.58 | 0.26 | **0.29** |
| **bus_662** | 574 | 818 | 9 | 10 | 11.31 | 1 | 0.69 | 7.76 | 1.55 | 0.04 | **0.04** |
| **bus_1138** | 671 | 991 | 17 | 18 | 18.71 | 0 | **0.86** | 2.19 | - | - | - |
| **USAir97** | 277 | 2071 | 139 | 140 | 156.99 | 16 | 24.86 | **1.1** | 9.84 | 2.42 | 1.33 |
| **celegans** | 445 | 2017 | 237 | 238 | 240.59 | 2 | 128.6 | **4.08** | 4.58 | 0.79 | 0.42 |
| **diseasome** | 420 | 1092 | 50 | 51 | 52.26 | 1 | 2.12 | **1.24** | 0.88 | 0.06 | 0.04 |
| **erdos971** | 337 | 1220 | 41 | 42 | 72.35 | 30 | 6.45 | **1.31** | 22.76 | 233.27 | 341.49 |
| **homer** | 333 | 1410 | 99 | 100 | 100.98 | 0 | 11.52 | **1.43** | - | - | - |
| **netscience** | 352 | 887 | 34 | 35 | 36.68 | 1 | 0.75 | 0.83 | 0.55 | 0.04 | **0.03** |
| **human-protein** | 811 | 2302 | 95 | 96 | 114.32 | 18 | 42.85 | **6.54** | 117.29 | 4.35 | 2.96 |
| **yeast** | 1488 | 5873 | 64 | 65 | 118.36 | 53 | 198.76 | **73.19** | 1560.09 | 6767.32 | 102.39 |
| **bible-nouns** | 1707 | 9059 | 364 | 365 | 406.08 | 41 | 1971.95 | **81.75** | 3753.94 | 89.81 | 54.97 |
| **hamster** | 1535 | 12223 | 272 | 273 | 373.69 | 100 | 2279.85 | **59.68** | 7626.9 | 257.63 | 197.34 |
| **hamster-full** | 1872 | 15970 | 273 | 274 | 382.09 | 108 | 2890.21 | **93.89** | 11424.42 | 307.26 | 239.17 |
| **Geom** | 2811 | 8651 | 102 | 103 | 157.2 | 54 | 247.26 | **187.6** | 7965.09 | 52.56 | 40.18 |
| **GR-QC** | 3413 | 12677 | 81 | 82 | 115.1 | 33 | 184.07 | 577.81 | 10638.77 | 23.31 | **15.17** |
| **HighEnergy** | 7059 | 23227 | 65 | 66 | 119.8 | 53 | 1880.02 | TL | TL | 216.95 | **117.95** |
| **US_Power** | 3353 | 5006 | 19 | 20 | 21.04 | 1 | 5.57 | 1793.25 | 74.27 | 0.15 | **0.09** |
| **PGPgiantcompo** | 5434 | 19070 | 205 | 206 | 215.31 | 9 | 416 | 482.48 | 17891.21 | 25.03 | **8.93** |

Notes: See our discussion on preprocessing in Section 3.1.5 and also the caption of Table 4 for additional details on the notation used in the table.

## 4. "Robust" generalizations

As briefly outlined in Sections 1 and 3.1.2, for many real-life graphs an induced subgraph that contains the maximum degree vertex with all its neighbors, i.e., adjacent vertices, very often turns out to be either an optimal solution or a solution that is nearly optimal; see, e.g., a recent study in Komusiewicz et al. (2019). Another supporting evidence for these earlier experimental results in the literature is also provided in our computational study in Section 3.1.2 if one compares the values computed in the columns denoted by "$d_{\max}$" and "$\bar{\omega}_2$" in Table 4. These empirical observations lead to the following two important viewpoints.

First, in real-life graphs the greedy heuristic provides a very good feasible solution, which is often also optimal. Thus, branch-and-bound and other enumerative approaches usually need to focus most of their efforts on proving optimality of such solutions. It implies that good quality upper bounds, e.g., those provided by the LP relaxations of our MIPs, are critical for improving computational performance of the exact methods. These arguments further highlight the importance of our theoretical (Proposition 3) and numerical (Tables 2, 3, 4, and 5) results on the LP relaxation tightness of our MIP models.

More importantly, the empirical observations in this paper and in the related literature imply that in many, if not most, sufficiently sparse real-life graphs maximum 2-clubs (in particular, those obtained by the aforementioned greedy procedure) can be viewed structurally as very close to star graphs. Therefore, as also outlined in Section 1, such 2-clubs contain multiple leaf vertices, which are typically connected to the vertex with the maximum degree; the latter is often referred to as the "hub" vertex. Hence, such 2-clubs are also susceptible to both edge and vertex failures, as they easily become disconnected with two or more connected components, if either the edge connecting the "hub" and one of the leaf vertices fails, or the "hub" vertex fails itself, e.g., due to a natural failure or an adversarial attack. These considerations resulted in a number of studies that focus on possible "robust" generalizations of the problem (Almeida & Brás, 2019; Carvalho & Almeida, 2017; Komusiewicz et al., 2019; Veremyev & Boginski, 2012a; 2012b; Veremyev, Prokopyev, Boginski, & Pasiliao, 2014; Yezerska, Pajouh, & Butenko, 2017).

Next, we discuss two intuitive robust versions of the maximum 2-club problem that directly address the outlined concerns on edge and vertex failures. Formally, let $S \subseteq V$ be a 2-club. Then:

($i$) The first generalization simply requires that the degree of any vertex in $G[S]$ should be at least $d_{\min}$, where $d_{\min} \in \mathbb{N}$ is some predefined constant parameter. Clearly, if $d_{\min} \geq 2$, then $G[S]$ does not contain leaf vertices; also, in order to disconnect such 2-club there should be at least $d_{\min}$ edge failures (Veremyev & Boginski, 2012b). We refer to the problem of finding such $S$ with maximum cardinality as *the problem of finding a maximum 2-club with the minimum degree requirement*, see (Veremyev & Boginski, 2012b). In Section 4.2 we describe how to extend the MIP-based approaches from Sections 2 and 3 in order to model such 2-clubs.

($ii$) The other generalization is known as *the maximum R-robust 2-club problem* (Komusiewicz et al., 2019; Veremyev & Boginski, 2012a). In this problem we seek a 2-club of maximum cardinality that also has at least $R$ vertex-disjoint paths of length at most 2 between any pair of vertices, where $R \in \mathbb{N}$. (The paths with the same endpoints are vertex-disjoint if they do not have any other vertex in common.) Hence, such 2-clubs are "protected" (i.e., keep the 2-club property and remain vertex pairwise connected with short paths of length at most 2) against up to $R-1$ vertex (and/or edge) failures. We discuss the MIP-based methods for this generalization in Section 4.3.

In Table 6 we compare the maximum cardinality of such "robust" 2-clubs for the same set of real-life graphs described in Section 2.4.2. From the results in this table we can make the following observations. First, the minimum degree requirement is much less restrictive than the other one based on the availability of $R$ "short" vertex-disjoint paths. For some graphs (see, e.g., **bcspwr04** and **celegans**) the minimum degree requirement does not substantially influence the sizes of the maximum 2-club. However, in general, the extra "robustness" condition (in particular, the availability of $R$ "short" vertex-disjoint paths) considerably reduces

**Table 6**
Size comparison of the maximum 2-clubs with extra “robustness” requirements: either the minimum degree for each vertex $d_{\min}$ or the minimum number $R$ of “short” vertex-disjoint paths between any pair of vertices. Symbol “-” indicates that such 2-clubs do not exist.

| | | | | 2-club with $d_{\min}$ | | | $R$-robust 2-club | |
|---|---|---|---|---|---|---|---|---|
| Graph | $\|V\|$ | $\|E\|$ | $d_{\max}$ | $d_{\min}=1$ | $d_{\min}=2$ | $d_{\min}=3$ | $R=2$ | $R=3$ |
| **bcspwr04** | 274 | 669 | 15 | 16 | 16 | 16 | 12 | 10 |
| **bus_494** | 494 | 586 | 9 | 10 | 7 | - | 3 | - |
| **bus_662** | 662 | 906 | 9 | 10 | 8 | 8 | 8 | 4 |
| **bus_1138** | 1138 | 1458 | 17 | 18 | 9 | 7 | 7 | 5 |
| **USAir97** | 332 | 2126 | 139 | 140 | 137 | 133 | 84 | 69 |
| **cables** | 429 | 636 | 16 | 17 | 10 | 4 | 5 | 4 |
| **celegans** | 453 | 2025 | 237 | 238 | 238 | 228 | 104 | 54 |
| **diseasome** | 516 | 1188 | 50 | 51 | 49 | 46 | 20 | 14 |
| **LindenStrasse** | 232 | 303 | 13 | 14 | 7 | - | 4 | - |
| **homer** | 542 | 1619 | 99 | 100 | 80 | 65 | 42 | 33 |
| **netscience** | 379 | 914 | 34 | 35 | 33 | 31 | 22 | 15 |
| **erdos971** | 429 | 1312 | 41 | 42 | 41 | 38 | 26 | 20 |
| **HarvardWeb** | 500 | 2043 | 200 | 201 | 162 | 131 | 43 | 40 |

the sizes of the maximum 2-clubs. Moreover, in some cases such 2-clubs do not even exist under some rather modest extra “robustness” requirements (see, e.g., **bus_494** and **LindenStrasse**). Note that this non-existence scenario occurs for very sparse real-life graphs, which is very intuitive.

We should note that there exist other “robust” generalizations of the maximum 2-club problem; see, e.g., Komusiewicz et al. (2019); Pattillo et al. (2013b); Yezerska et al. (2017) and the references therein. In particular, the study by Komusiewicz et al. (2019) considers the maximum $R$-robust 2-club problem and its generalizations ($t$-Robust/$t$-Hereditary/$t$-Connected 2-clubs), and develops specialized exact combinatorial algorithms. The latter exploits various efficient data reduction and preprocessing techniques. As a benchmark, Komusiewicz et al. (2019) uses a version of the **F2s** model that is extended to capture the considered generalizations. We leave it as a possible direction of future research to explore extensions of our models to the “robust” generalizations considered in Komusiewicz et al. (2019). Furthermore, it could be of interest to study the advanced data reduction and preprocessing techniques proposed in Komusiewicz et al. (2019) to enhance the performance of our approaches.

Finally, for the details on the computational setting used in our experiments discussed below, we refer the reader to Section 2.4.1.

### 4.1. Preprocessing

For both of the considered generalizations we observe that any vertex such that its degree is smaller than $r$, where either $r=d_{\min}$, or $r=R$, cannot belong to an optimal solution. Thus, all such vertices can be removed from the graph, which, in turn, may reduce the degrees of the remaining vertices. Consequently, this procedure, often referred to in the related literature (see, e.g., Pastukhov, Veremyev, Boginski, & Prokopyev, 2018; Verma, Buchanan, & Butenko, 2015) as “peeling” (or “vertex peeling”), can be performed in an iterative manner until the remaining subgraph contains only vertices with degrees at least $r$. Such subgraph is known as an *r-core*, see, e.g., Pattillo et al. (2013b). We apply this efficient preprocessing procedure (its running time is $O(|E|)$) for both generalizations and report the sizes of the remaining subgraphs (further referred to as either $d_{\min}$-core, or $R$-core, respectively) in our computational results; see Tables 7 and 8. By comparing the graph sizes given in Table 6 with those reported in Tables 7 and 8 we conclude that the preprocessing procedure is very effective for all of our test instances except **celegans** when $d_{\min}=2$ and $R=2$.

### 4.2. 2-clubs with the minimum degree requirement

Next, we assume that $G=(V,E)$ contains only vertices with their degrees at least $d_{\min}$. That is, the preprocessing procedure from Section 4.1 is applied, and $G$ is a $d_{\min}$-core itself in the remainder of this section.

To solve the problem of finding a maximum 2-club with the minimum degree requirement, it is sufficient to add for each MIP from Sections 2 an extra set of linear constraints in the form:

$$\sum_{j\in N(i)} x_j \geq d_{min} x_i \qquad \forall i \in V, \tag{11}$$

which ensures that the resulting 2-club contains only vertices with degrees at least $d_{\min}$.

We refer to the resulting MIPs as **FD2s**, **FD2c1** and **FD2c2**, which are obtained by adding (11) into **F2s**, **F2c1** and **F2c2**, respectively. Also, the optimal objective function value, i.e., the size of the optimal 2-club, is denoted by $\bar{\omega}_{d2}$.

Furthermore, we note that the theoretical results on the LP relaxation quality from Section 2.3, namely, Propositions 3 and 4, also hold for the considered “robust” version of the problem. The corresponding computational results, see the columns denoted by “LP rlx. objective” in Table 7, are consistent with the corresponding results in Table 3. That is, both LP relaxations of **FD2c1** and **FD2c2** are better than that of **FD2s** in most cases (except **celegans** and **USAir97**), with **FD2c2** being the best one. However, we observe that the presence of (11) decreases the quality of the LP relaxation based bounds (if one compares the latter values against the maximum size of the “robust” 2-club given by $\bar{\omega}_{d2}$). Consequently, the running times of the solver with MIPs **FD2c1** and **FD2c2** significantly deteriorate, see Table 7.

However, the “feasibility-check” algorithm (Algorithm 1) from Section 3 can be adapted with some modifications to handle both considered “robust” versions of the problem, see its pseudo-code in Algorithm 2 . The modified algorithm provides consistent results and outperforms the considered MIPs, see Table 7, with respect to the running time for most of the instances.

To conclude the discussion, we briefly describe Algorithm 2 next. The feasibility MIP **FD2c1**($\ell$) can be created by adding (11) into **F2c1**($\ell$). As in the previous approach, **FD2c1**($\ell$) is solved iteratively for different values of $\ell$, see line 14 in Algorithm 2. There are two differences between Algorithms 1 and 2. First, the lower bound given by (5) cannot be applied. Instead, it is replaced by a trivial lower bound $\ell \geq d_{\min}+1$, see line 13 in Algorithm 2. Second, the problem is not guaranteed to have a feasible solution, see **LindenStrasse** in Table 7; recall also our

**Table 7**
Comparison of the number of constraints, the optimal objective function values of the LP relaxations and the solution times for formulations **FD2s**, **FD2c1** and **FD2c2** as well as Algorithm 2 on real-life graphs; see their description and other settings in Section 2.4. Time limit (TL) is set to 50,000 sec. The best LP relaxation objective functions and the best running times are in **bold**. For each graph we report its size after preprocessing, i.e., the size of the corresponding $d_{\min}$-core.

| | $d_{\min}$-core | | | | | | LP rlx. objective | | | Time (seconds) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Graph | $\|V\|$ | $\|E\|$ | $\binom{\|V\|}{2}$ | $\|E_2\|$ | $\bar{\omega}_{d2}$ | $\hat{\beta}^u_2$ | **FD2s** | **FD2c1** | **FD2c2** | **FD2s** | **FD2c1** | **FD2c2** | Alg. 2 |
| $d_{\min}$=2 | | | | | | | | | | | | | |
| **bcspwr04** | 217 | 612 | 22824 | 1999 | 16 | 63.7 | 108.5 | 53.1 | **22.2** | 10.6 | 11.9 | 5.7 | **0.4** |
| **bus_494** | 277 | 369 | 37857 | 1044 | 7 | 46.2 | 138.5 | 40.5 | **10.2** | 4.4 | 4.1 | 0.5 | **0.2** |
| **bus_662** | 574 | 818 | 163633 | 2521 | 8 | 71.5 | 287.0 | 65.0 | **11.2** | 156.5 | 46.8 | 2.2 | **0.6** |
| **bus_1138** | 671 | 991 | 223794 | 3229 | 9 | 80.9 | 335.5 | 73.3 | **17.7** | 1165.9 | 86.6 | 2.5 | **1.1** |
| **USAir97** | 277 | 2071 | 36155 | 18331 | 137 | 191.9 | **138.5** | 159.1 | 156.7 | **11.1** | 61.7 | TL | 11.5 |
| **cables** | 337 | 544 | 56072 | 1935 | 10 | 62.7 | 168.5 | 54.8 | **16.0** | 33.7 | 14.7 | 2.1 | **0.8** |
| **celegans** | 445 | 2017 | 96773 | 45256 | 238 | 301.4 | **238.0** | 253.9 | 240.6 | **0.2** | 954.4 | 832.4 | 42.2 |
| **diseasome** | 420 | 1092 | 86898 | 4929 | 49 | 99.8 | 210.0 | 85.3 | **52.3** | 251.2 | 57.3 | 3.6 | **1.0** |
| **LindenStrasse** | 116 | 187 | 6483 | 680 | 7 | 37.4 | 58.0 | 30.5 | **13.1** | 0.7 | 1.2 | **0.6** | 0.8 |
| **homer** | 333 | 1410 | 53868 | 13568 | 80 | 165.2 | 166.5 | 132.3 | **101.0** | 45.5 | 166.0 | TL | **6.4** |
| **netscience** | 352 | 887 | 60889 | 3551 | 33 | 84.8 | 176.0 | 72.1 | **36.3** | 122.4 | 26.3 | 3.4 | **0.6** |
| **erdos971** | 337 | 1220 | 55396 | 8974 | 41 | 134.5 | 168.5 | 108.9 | **72.4** | **71.6** | 417.8 | TL | 215.1 |
| **HarvardWeb** | 421 | 1964 | 86446 | 30130 | 162 | 245.9 | 210.5 | 208.2 | **198.9** | 64.6 | 455.2 | TL | **29.3** |
| $d_{\min}$=3 | | | | | | | | | | | | | |
| **bcspwr04** | 169 | 528 | 13668 | 1606 | 16 | 57.2 | 84.5 | 46.7 | **22.1** | 6.3 | 9.2 | 7.7 | **0.3** |
| **bus_494** | 0 | 0 | - | - | - | - | - | - | - | - | - | - | - |
| **bus_662** | 17 | 37 | 99 | 65 | 8 | 11.9 | 8.5 | **10.1** | **10.1** | **0.1** | **0.1** | 0.4 | **0.1** |
| **bus_1138** | 16 | 29 | 91 | 37 | 7 | 9.1 | 8.0 | 7.63 | **7.0** | < **0.1** | 0.1 | 0.2 | < **0.1** |
| **USAir97** | 227 | 1982 | 23669 | 15308 | 133 | 175.5 | **133.7** | 148.3 | 151.1 | **0.6** | 272.9 | TL | 9.5 |
| **cables** | 43 | 74 | 829 | 169 | 4 | 18.9 | 21.5 | 15.3 | **11.0** | 0.2 | **0.1** | 0.5 | 0.2 |
| **celegans** | 429 | 1988 | 89818 | 43325 | 228 | 294.9 | **230.2** | 247.8 | 234.3 | **1.4** | 2434.8 | 45092.1 | 36.2 |
| **diseasome** | 290 | 873 | 41032 | 3343 | 46 | 82.3 | 145.0 | 69.3 | **49.6** | 155.3 | 123.4 | 35.2 | **0.6** |
| **LindenStrasse** | 0 | 0 | - | - | - | - | - | - | - | - | - | - | - |
| **homer** | 226 | 1223 | 24202 | 9296 | 65 | 136.9 | 113.0 | 109.6 | **93.4** | 32.4 | 626.5 | TL | **4.4** |
| **netscience** | 265 | 736 | 34244 | 2464 | 31 | 70.7 | 132.5 | 59.6 | **33.7** | 77.7 | 81.3 | 4.6 | **0.3** |
| **erdos971** | 257 | 1073 | 31823 | 7636 | 38 | 124.1 | 128.5 | 99.3 | **71.7** | **33.5** | 433.5 | TL | 391.9 |
| **HarvardWeb** | 315 | 1755 | 47700 | 18599 | 131 | 193.4 | **157.5** | 166.1 | 160.0 | 72.5 | 1197.5 | TL | **13.3** |

**Table 8**
Comparison of the number of constraints, the optimal objective function values of the LP relaxations and the solution times for formulations **FR2s**, **FR2c1** and **FR2c2** as well as Algorithm 2 on real-life graphs; see their description and other settings in Section 2.4. Time limit (TL) is set to 50,000 sec. The best LP relaxation objective functions and the best running times are in **bold**. For each graph we report its size after preprocessing, i.e., the size of the corresponding $R$-core.

| | $R$-core | | | | | | LP rlx. objective | | | Time (seconds) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Graph | $\|V\|$ | $\|E\|$ | $\binom{\|V\|}{2}$ | $\|E_{R2}\|$ | $\bar{\omega}_{R2}$ | $\hat{\beta}^u_{R2}$ | **FR2s** | **FR2c1** | **FR2c2** | **FR2s** | **FR2c1** | **FR2c2** | Alg. 2 |
| $R$=2 | | | | | | | | | | | | | |
| **bcspwr04** | 217 | 612 | 23,436 | 990 | 12 | 45.0 | 108.5 | 38.8 | **16.3** | 26.8 | 8.8 | 0.5 | **0.3** |
| **bus_494** | 277 | 369 | 38,226 | 101 | 3 | 14.7 | 138.5 | 13.4 | **3.7** | 90.2 | 3.58 | < **0.1** | < **0.1** |
| **bus_662** | 574 | 818 | 164,451 | 229 | 8 | 21.9 | 287.0 | 20.3 | **9.0** | 3104.2 | 47.9 | 0.1 | 0.1 |
| **bus_1138** | 671 | 991 | 224,785 | 499 | 7 | 32.1 | 335.5 | 30.7 | **7.5** | 6441.3 | 71.2 | < **0.1** | 0.1 |
| **USAir97** | 277 | 2071 | 38,226 | 11,937 | 84 | 155.0 | 138.5 | 127.3 | **120.2** | **41.3** | 185.7 | TL | 200.9 |
| **cables** | 337 | 544 | 56,616 | 334 | 5 | 26.35 | 168.5 | 24.4 | **6.3** | 188.6 | 7.0 | **0.1** | **0.1** |
| **celegans** | 445 | 2017 | 102,378 | 17,872 | 104 | 189.6 | 227.7 | 154.3 | **116.9** | 317.9 | 924.5 | TL | **16.8** |
| **diseasome** | 420 | 1092 | 87,990 | 1664 | 20 | 58.2 | 210.0 | 52.0 | **24.9** | 432.8 | 36.8 | 1.2 | **0.4** |
| **LindenStrasse** | 116 | 187 | 6670 | 71 | 4 | 12.4 | 58.0 | 11.1 | **4.6** | 2.7 | 0.4 | 0.1 | < **0.1** |
| **homer** | 333 | 1410 | 55,278 | 5219 | 42 | 102.7 | 166.5 | 85.4 | **59.4** | 186.9 | 202.0 | TL | **8.5** |
| **netscience** | 352 | 887 | 61,776 | 1457 | 22 | 54.5 | 176.0 | 48.0 | **22.8** | 206.7 | 12.3 | 0.3 | **0.2** |
| **erdos971** | 337 | 1220 | 56,616 | 3465 | 26 | 83.8 | 168.5 | 70.9 | **43.4** | 159.1 | 42.8 | 38483.2 | **15.5** |
| **HarvardWeb** | 421 | 1964 | 88,410 | 6291 | 43 | 112.7 | 210.5 | 94.3 | **52.1** | 378.2 | 119.9 | 373.3 | **3.2** |
| $R$=3 | | | | | | | | | | | | | |
| **bcspwr04** | 169 | 528 | 14,196 | 560 | 10 | 34.0 | 84.5 | 28.7 | **12.9** | 10.5 | 3.2 | 0.3 | **0.1** |
| **bus_494** | 0 | 0 | - | - | - | - | - | - | - | - | - | - | - |
| **bus_662** | 17 | 37 | 136 | 38 | 4 | 9.2 | 8.5 | 7.6 | **7.3** | < **0.1** | 0.1 | < **0.1** | 0.1 |
| **bus_1138** | 16 | 29 | 120 | 26 | 5 | 7.7 | 8.0 | 6.0 | **5.0** | < **0.1** | < **0.1** | < **0.1** | < **0.1** |
| **USAir97** | 227 | 1982 | 54,946 | 8076 | 69 | 127.6 | 118.2 | 104.6 | **99.5** | **20.5** | 110.7 | TL | 126.6 |
| **cables** | 43 | 74 | 903 | 23 | 4 | 7.3 | 21.5 | 6.3 | **4.0** | 0.1 | 0.1 | < **0.1** | < **0.1** |
| **celegans** | 429 | 1988 | 91,806 | 7814 | 54 | 125.5 | 217.8 | 104.9 | **69.2** | 514.2 | 269.7 | 48688.5 | **6.1** |
| **diseasome** | 290 | 873 | 41,905 | 950 | 14 | 44.1 | 145.0 | 39.1 | **17.3** | 96.9 | 15.3 | 0.6 | **0.2** |
| **LindenStrasse** | 0 | 0 | - | - | - | - | - | - | - | - | - | - | - |
| **homer** | 226 | 1223 | 25,425 | 2873 | 33 | 76.3 | 113.0 | 63.2 | **47.0** | 32.0 | 18.5 | 3748.0 | **1.4** |
| **netscience** | 265 | 736 | 34,980 | 847 | 15 | 41.7 | 132.5 | 36.4 | **16.1** | 56.3 | 2.6 | **0.1** | 0.2 |
| **erdos971** | 257 | 1073 | 32,896 | 1674 | 20 | 58.4 | 128.5 | 50.1 | **32.2** | 75.8 | 10.1 | 236.4 | **4.6** |
| **HarvardWeb** | 315 | 1755 | 49,455 | 2781 | 40 | 75.1 | 158.3 | 64.1 | **41.6** | 116.1 | 20.3 | 5.0 | **0.7** |

**Algorithm 2:** Exact "feasibility-check" algorithm ("robust" 2-club).

1 **Input:** graph $G=(V,E)$, either "robustness" parameter $d_{\min}$ or $R \in \mathbb{N}$
2 **Output:** maximum "robust" 2-club $S \subseteq V$ (either with the minimum degree $d_{\min}$ requirement or $R$-robust 2-club)
3 **begin**
4 $G \leftarrow$ either $d_{\min}$-core or $R$-core of the input graph $G$, depending on the "robust" version considered, after the preprocessing procedure from Section 4.1 is applied
5 **FIPs** $\leftarrow$ either **FD2s**, or **FR2s**~depending on the "robust" version considered
6 **FMIP1** $\leftarrow$ either **FD2c1**, or **FR2c1**, as above
7 **FMIP2** $\leftarrow$ either **FD2c2**, or **FR2c2**, as above
8 **FMIP1**$(\ell)$ $\leftarrow$ either **FD2c1**$(\ell)$, or **FR2c1**$(\ell)$, as above
9 $UB \leftarrow$ minimum value of the optimal objective function values after solving the LP relaxations of **FIPs**, **FMIP1** and **FMIP2**
10 $\ell \leftarrow \lfloor UB \rfloor$
11 $\ell_{\min} \leftarrow$ either $d_{\min}$ or $R$ depending on the "robust" version considered
12 $S \leftarrow \emptyset$
13 **while** $S$ is empty and $\ell \geq \ell_{\min}+1$ **do**
14 Solve **FMIP1**$(\ell)$
15 **if FMIP1**$(\ell)$ is feasible **then**
16 $S \leftarrow \{i:\ x_i^*=1, \forall i \in V\}$, where $\mathbf{x}^*$ is feasible for **FMIP1**$(\ell)$
17 **end**
18 $\ell \leftarrow \ell - 1$
19 **end**
20 return $S$
21 **end**

earlier discussion of Table 6, where the entry "-" corresponds to the values of $d_{\min}$ for which a feasible solution does not exists. Hence, in contrast to Algorithm 1, in its modified version given by Algorithm 2 we have an empty initial feasible solution (see lines 12 and 13 of Algorithm 2). Furthermore, we need to consider $\ell = d_{\min}+1$ (see line 13 of Algorithm 2), while in Algorithm 1 a feasible solution is readily available for the corresponding lower bound (recall lines 6–8 in Algorithm 1).

Finally, it should be pointed out that similar to Algorithm 1, in our experiments with real-life graphs we compute only the LP relaxation of **FD2c2** in line 9 of Algorithm 2, as **FD2c2** almost always provides the best LP relaxation bound.

### 4.3. R-robust 2-club

The standard formulation for the maximum $R$-robust 2-club problem (Veremyev & Boginski, 2012a) is given by:

$$[\mathbf{FR2s}]: \quad \bar{\omega}_{R2} := \max \sum_{i \in V} x_i \tag{12a}$$

subject to

$$\mathbb{1}_{(i,j)\in E} + \sum_{t \in N(i)\cap N(j)} x_t \geq R(x_i + x_j - 1) \quad \forall (i,j) \in \binom{V}{2}, \tag{12b}$$

$$x_i \in \{0,1\} \qquad \forall i \in V, \tag{12c}$$

where (12b) is a generalization of (1b), and $\mathbb{1}_{(i,j)\in E}$ indicates whether $(i,j)\in E$, i.e., $\mathbb{1}_{(i,j)\in E}=1$ if $(i,j)\in E$, and $\mathbb{1}_{(i,j)\in E}=0$ if $(i,j)\notin E$. Specifically, (12b) ensures that for any pair of vertices $(i,j)\in\binom{V}{2}$ selected to be in an $R$-robust 2-club (i.e., $x_i=x_j=1$), there must be at least $R-1$ or $R$ common neighbors of vertices $i,j$ in that $R$-robust 2-club depending on whether the vertices are adjacent or non-adjacent, respectively. The formulation requires $\binom{|V|}{2}$ constraints.

To apply our new formulation technique in a more efficient way we modify the definition of $E_2$ in (2) as follows:

$$E_{R2} = \{(i,j) \in \binom{V}{2} \mid |N(i)\cap N(j)| + \mathbb{1}_{(i,j)\in E} \geq R\},$$

i.e., $E_{R2}$ is a set of all vertex pairs in $G$ such that there exist at least $R$ vertex-disjoint paths of length at most 2 between them. Similar in spirit to Section 2, one would expect that $|E_{R2}| \ll |V|^2$ for sufficiently sparse graphs, see the appropriate columns in Table 8 for comparison.

Given the above notation model **F2c1** is generalized as follows:

$$[\mathbf{FR2c1}]: \quad \bar{\omega}_{R2} = \max \sum_{i \in V} x_i \tag{13a}$$

subject to

$$\sum_{j \in V:\ (i,j)\in E_{R2}} u_{ij} \geq \sum_{j\in V} x_j - 1 - \mu(1-x_i) \qquad \forall i \in V, \tag{13b}$$

$$u_{ij} \leq x_i, u_{ij} \leq x_j \qquad \forall (i,j)\in E_{R2}, \tag{13c}$$

$$u_{ij} \leq \frac{1}{R}\left(\mathbb{1}_{(i,j)\in E} + \sum_{t\in N(i)\cap N(j)} x_t\right) \qquad \forall (i,j)\in E_{R2}, \tag{13d}$$

$$u_{ij} \geq 0,\ \ x_i \in \{0,1\} \qquad \forall (i,j)\in E_{R2}, i \in V, \tag{13e}$$

where (13d) is a generalization of (3d), and $E_2$ is replaced by $E_{R2}$ in the appropriate terms. Similarly, **F2c2** becomes:

$$[\mathbf{FR2c2}]: \quad \bar{\omega}_{R2} = \max \sum_{i \in V} x_i \tag{14a}$$

subject to

$$\sum_{(i,j)\in E_{R2}} u_{ij} \geq \sum_{\ell=\beta_{R2}^l}^{\beta_{R2}^u} \frac{\ell(\ell-1)}{2} z_\ell, \tag{14b}$$

$$u_{ij} \leq x_i, u_{ij} \leq x_j \qquad \forall (i,j)\in E_{R2}, \tag{14c}$$

$$u_{ij} \leq \frac{1}{R}\left(\mathbb{1}_{(i,j)\in E} + \sum_{t\in N(i)\cap N(j)} x_t\right) \qquad \forall (i,j)\in E_{R2}, \tag{14d}$$

$$\sum_{\ell=\beta_{R2}^l}^{\beta_{R2}^u} \ell z_\ell = \sum_{i\in V} x_i, \quad \sum_{\ell=\beta_{R2}^l}^{\beta_{R2}^u} z_\ell = 1, \tag{14e}$$

$$u_{ij} \geq 0,\ \ x_i \in \{0,1\} \qquad \forall (i,j)\in E_{R2}, i \in V, \tag{14f}$$

$$z_\ell \in \{0,1\} \qquad \forall \ell \in \{\beta_{R2}^l, \ldots, \beta_{R2}^u\}, \tag{14g}$$

where (14d) corresponds to modified (4d) from **F2c2**. Also, $\beta_{R2}^l$ and $\beta_{R2}^u$ denote some lower and upper bounds on the size of a maximum $R$-robust 2-club. We use $\beta_{R2}^l = 0$ as such 2-club does not necessarily exist; recall Table 8. For the upper bound, Proposition 3 can easily be extended to:

**Corollary 1** (Upper bound on an $R$-robust 2-club size)**.** *Let $S \subseteq V$ be an $R$-robust 2-club in $G = (V, E)$. Then*

$$|S| \le \beta^u_{R2} = \left\lfloor \frac{1 + \sqrt{1 + 8|E_{R2}|}}{2} \right\rfloor \tag{15}$$

*and this bound is sharp.*

Furthermore, Proposition 2 also holds for **FR2c2** and thus, (14g) can be replaced by $z_\ell \ge 0$ for all $\ell \in \{\beta^l_{R2}, \ldots, \beta^u_{R2}\}$. With respect to the LP relaxation quality, next we assume that $G = (V, E)$ contains only vertices with their degrees at least $R$ after the pre-processing procedure from Section 4.1 is applied. Then:

**Corollary 2** (LP relaxation bounds)**.** *Let $\bar\omega^s_{R2}$, $\bar\omega^{c1}_{R2}$ and $\bar\omega^{c2}_{R2}$ be the optimal objective function values of the LP relaxations of formulations **FR2s**, **FR2c1** and **FR2c2**, respectively, where $\mu = \beta^u_{R2} - 1$ and $\beta^u_{R2} = \left\lfloor \frac{1+\sqrt{1+8|E_{R2}|}}{2} \right\rfloor$. Then the following inequalities hold,*

(i) $\bar\omega^s_{R2} \ge \frac{|V|}{2}$
(ii) $\bar\omega^{c1}_{R2} \le \hat\beta^u_{R2} := \frac{1+\sqrt{1+8|E_{R2}|}}{2}$
(iii) $\bar\omega^{c2}_{R2} \le \hat\beta^u_{R2} := \frac{1+\sqrt{1+8|E_{R2}|}}{2}$

To conclude our theoretical development, we point out that the feasibility MIPs can be extended in a similar manner, and the "feasibility-check" algorithm is outlined in Algorithm 2. Note that in the pseudo-code, see Algorithm 2, we use notation for the "robust" feasibility MIPs similar to those used in Section 3.1.

The computational results are provided in Table 8, which are consistent with those in our previous results. Namely, **FR2c2** provides the best LP relaxation quality (hence, only this model is used in line 9 of Algorithm 2); furthermore the "feasibility-check" algorithm is typically the best approach with respect to the overall running time.

## 5. Concluding remarks

In this paper we consider new MIP models for the maximum 2-club problem and compare them against a classical IP model from the literature. The new models are based on exploiting slightly different interpretations of the original "small diameter" requirement of this popular clique relaxation model. We demonstrate both theoretically and numerically that our MIP models have much better LP relaxation quality than the standard IP model for sufficiently sparse graphs. Hence, these LP relaxations can be used to provide high quality (and polynomially computable) upper bounds for the sizes of maximum 2-clubs in real-life graphs.

We perform a computational study with real-life and randomly generated graphs to explore the running time performance (when solving to optimality) of our models using an off-the-shelf commercial MIP solver Gurobi (Gurobi Optimization, 2019). If the presolve (i.e., a collection of various preprocessing routines implemented in Gurobi (Achterberg et al., 2020)) is switched off, then the new models substantially outperform the standard model (due to a better quality of the LP relaxation). If the presolve is used, then the performance of the standard model improves substantially; however, the performance of our models is either not affected or deteriorates. This observation implies that the new models are structurally less amenable to preprocessing routines implemented in the Gurobi MIP solver. Hence, the classical IP model remains a viable alternative for solving the problem in reasonably sized graphs when using solvers with advanced presolve implementations. It also opens up an interesting avenue for future research to explore both computationally (e.g., by comparing commercial and open source MIP solvers) and, perhaps, theoretically the reasons behind these observations.

Furthermore, we consider a "feasibility-check" algorithm that iteratively solves feasibility versions of the considered MIPs for each possible 2-club size within some known lower and upper bounds. The latter bound is computed using the LP relaxations of new models. Their high quality allows this algorithm to outperform the standard IP model (even when the presolve used in the solver) for sufficiently large graphs with up to 10,000 vertices. As emphasized earlier, the key advantage of this feasibility based approach is that it is extremely simple to implement and does not require significant implementation efforts.

Our numerical experiments also support earlier observations from the related literature that in sparse real-life networks maximum 2-clubs are typically not "robust" with respect to edge and/or vertex failures and are easily disconnected into multiple connected components. Hence, we show how to extend our approaches to solving two "robust" (attack- and failure-tolerant) generalizations of the maximum 2-club problem.

With respect to future research directions our results provide numerous avenues for further studies. For example, the developed formulations can be directly extended to the 2-club partition or coverage problems (Dondi & Lafond, 2019; Dondi, Mauri, Sikora, & Zoppis, 2018; Dondi, Mauri, & Zoppis, 2019; Gschwind, Irnich, Furini, & Calvo, 2020; Yezerska, Pajouh, Veremyev, & Butenko, 2019). In addition, it may be interesting to extend our modeling approach to other $k$-club problems with $k \ge 3$.

Finally, more results on the polyhedral properties of our MIPs are also of interest; see, e.g., (Buchanan & Salemi, 2020; Pajouh et al., 2016). For example, the study in Carvalho and Almeida (2011) builds upon the classical IP model (and its polyhedral properties) to develop specialized algorithms and heuristics for solving the maximum 2-club problem. A similar direction could be pursued in order to explore whether the new models could be strengthened, perhaps, in some some combinations with the classical approach. Such results may lead to the development of substantially more advanced algorithms for solving the considered class of combinatorial optimization problems.

## Acknowledgments

The authors would like to thank the Associate Editor and three anonymous referees for their extremely helpful and constructive comments that allowed us to substantially improve this paper. The research of A. Veremyev and V. Boginski was partially supported by the Air Force Research Laboratory (AFRL) award FA8651-16-2-0009. The research of O.A. Prokopyev was partially supported by the US National Science Foundation Grant CBET 1803527.

## Appendix A

**Proof of Proposition 1.** We demonstrate that any optimal solution of formulations **F2c1** or **F2c2** corresponds to a certain 2-club with the size equal to their respective objective function values, and for any 2-club $S' \subseteq V$ of graph $G$ we can construct feasible solutions of formulations **F2c1** and **F2c2** with the objective function value $|S'|$. Without loss of generality, we assume that $\beta^l_2 = 1$ and $\beta^u_2 = |V| = n$ in **F2c2**.

First, we consider **F2c1**. Let $\mathbf{x}^* = (x^*_1, \ldots, x^*_n)^T$ and $\mathbf{u}^* = \{u^*_{ij} \mid (i, j) \in E_2\}$ be an optimal solution of **F2c1**, and let $S^* = \{i \in V \mid x^*_i = 1\}$. We need to show that $S^*$ is a 2-club.

Note that constraints (3c) imply that $u^*_{ij} \in [0, 1]$. Furthermore, if $u^*_{ij} > 0$ in an optimal solution, then setting $u^*_{ij} = 1$ does not violate constraint (3b) as it only increases its left-hand side; also, it does not violate constraints (3c) and (3d) as variable $x_i$ is binary for all $i \in V$. More importantly, this modification does not change the objective function value of **F2c1**. Thus, without loss of generality we can assume that $u^*_{ij} \in \{0,1\}$ for all $(i,j) \in E_2$.

For any vertex $i \in S^*$, we have

$$\sum_{j\in V:(i,j)\in E_2} u^*_{ij} \geq |S^*| - 1$$

due to constraint (3b). Moreover, since $u^*_{ij} = 0$ for any $j \notin S^*$ and $(i,j) \in E_2$ due to constraints (3c) we have:

$$\sum_{j\in V:\ (i,j)\in E_2} u^*_{ij} = \sum_{j\in S^*\setminus\{i\}:\ (i,j)\in E_2} u^*_{ij} \leq |S^*| - 1.$$

Therefore,

$$\sum_{j\in V:\ (i,j)\in E_2} u^*_{ij} = \sum_{j\in S^*\setminus\{i\}:\ (i,j)\in E_2} u^*_{ij} = |S^*| - 1,$$

which implies that for any $i,j \in S^*$, we have $(i,j) \in E_2$ and $u^*_{ij} = 1$. Hence, it is either $(i,j) \in E$, or $\sum_{t\in N(i)\cap N(j)} x^*_t \geq 1$ due to constraint (3d). The latter implies that if vertices $i \in S^*$ and $j \in S^*$ are not directly connected by an edge, then they have at least one common neighbor $t \in S^*$. Thus, $S^*$ is a 2-club.

Next, we consider **F2c2**. Let $\mathbf{x}^* = (x^*_1,\dots,x^*_n)^T$, $\mathbf{z}^* = (z^*_1,\dots,z^*_n)^T$, $\mathbf{u}^* = \{u^*_{ij} \mid (i,j) \in E_2\}$ be an optimal solution of **F2c2**, and let $S^* = \{i \in V \mid x^*_i = 1\}$. Similarly, to the discussion above we can assume that $u^*_{ij} \in \{0,1\}$ for all $(i,j) \in E_2$.

Note that $z^*_{|S^*|} = 1$ and $z^*_\ell = 0$ for $\ell \in \{1,\dots,n\}, \ell \neq |S^*|$, due to constraint (4e). Therefore, constraint (4b) becomes

$$\sum_{(i,j)\in E_2} u^*_{ij} \geq \sum_{\ell=1}^{n} \frac{\ell(\ell-1)}{2} z^*_\ell = \frac{|S^*|(|S^*|-1)}{2}.$$

Moreover, due to constraints (4c) we have $u^*_{ij} = 0$ if either $i \in V \setminus S^*$ or $j \in V \setminus S^*$ and $(i,j) \in E_2$. Hence,

$$\sum_{(i,j)\in E_2} u^*_{ij} = \sum_{(i,j)\in E_2:\ i,j\in S^*} u^*_{ij} \geq \frac{|S^*|(|S^*|-1)}{2}$$

and since $|\{(i,j) \in E_2 \mid i,j \in S^*\}| \leq |\{i,j \in S^* \mid i<j\}| = \frac{|S^*|(|S^*|-1)}{2}$, it follows that

$$\sum_{(i,j)\in E_2:\ i,j\in S^*} u^*_{ij} = \frac{|S^*|(|S^*|-1)}{2}.$$

The latter implies that for any $i,j \in S^*$, we have $(i,j) \in E_2$ and $u^*_{ij} = 1$. Using the same arguments as in the case of **F2c1** above, we conclude that $S^*$ is a 2-club. Namely, it is either $(i,j) \in E$, or $\sum_{t\in N(i)\cap N(j)} x^*_t \geq 1$, which implies that if vertices $i \in S^*$ and $j \in S^*$ are not directly connected by an edge, then they have at least one common neighbor $t \in S^*$.

Next, assume that $S'$ is a 2-club. For any $i,j \in V$ such that $(i,j) \in E_2$, let

$$x'_i = \begin{cases} 1, & \text{if } i \in S', \\ 0, & \text{if } i \notin S', \end{cases} \quad \text{and } u'_{ij} = \begin{cases} 1, & \text{if } i \in S', \text{ and } j \in S', \\ 0, & \text{if } i \notin S', \text{ or } j \notin S'. \end{cases}$$

Clearly, if $i,j \in S'$, then $(i,j) \in E_2$. Also, it can be verified that $\mathbf{x}', \mathbf{u}'$, where $\mathbf{x}' = (x'_1,\dots,x'_n)^T$ and $\mathbf{u}' = \{u'_{ij} \mid (i,j) \in E_2\}$, is a feasible solution of **F2c1**, i.e., it satisfies all modeling constraints.

Moreover, let $z'_{|S'|} = 1$ and $z'_\ell = 0$ for $\ell \in \{1,\dots,n\}, \ell \neq |S'|$, and $\mathbf{z}' = (z'_1,\dots,z'_n)^T$. It is also easy to verify that $\mathbf{x}', \mathbf{u}', \mathbf{z}'$ is a feasible solution of formulation **F2c2**. Both objective functions are equal to $|S'|$, which completes the proof. □

**Proof of Proposition 2.** Without loss of generality, we assume that $\beta^l_2 = 1$ and $\beta^u_2 = n$. Let $\mathbf{x}^* = (x^*_1,\dots,x^*_n)^T$, $\mathbf{z}' = (z'_1,\dots,z'_n)^T$, $\mathbf{u}^* = \{u^*_{ij} \mid (i,j) \in E_2\}$ be an optimal solution of formulation **F2c2** with relaxed variables $\mathbf{z}$, i.e., (4g) is replaced by (6). Next, suppose that $\mathbf{z}'$ is not a binary vector. Let $\mathbf{z}^* = (z^*_1,\dots,z^*_n)^T$ be defined as follows:

$$z^*_\ell = \begin{cases} 1, & \text{if } \ell = \sum_{i\in V} x^*_i, \\ 0, & \text{if } \ell \neq \sum_{i\in V} x^*_i. \end{cases} \tag{16}$$

The definition above and (4e) imply that

$$\sum_{\ell=1}^{n} \ell z'_\ell = \sum_{i\in V} x^*_i = \sum_{\ell=1}^{n} \ell z^*_\ell \tag{17}$$

To prove that $\mathbf{x}^*, \mathbf{u}^*, \mathbf{z}^*$ is also a feasible solution of **F2c2** we apply Jensen's inequality to convex function $f(\ell) = \ell(\ell-1)/2$. Specifically, using Jensen's inequality and (17) we have:

$$\frac{\sum_{\ell=1}^{n} \frac{\ell(\ell-1)}{2} z'_\ell}{\sum_{\ell=1}^{n} z'_\ell} \geq \frac{\sum_{\ell=1}^{n} \ell z'_\ell \left(\sum_{\ell=1}^{n} \ell z'_\ell - 1\right)}{2} = \frac{\sum_{\ell=1}^{n} \ell z^*_\ell \left(\sum_{\ell=1}^{n} \ell z^*_\ell - 1\right)}{2}$$
$$= \sum_{\ell=1}^{n} \frac{\ell(\ell-1)}{2} z^*_\ell,$$

which implies that

$$\sum_{(i,j)\in E_2} u^*_{ij} \geq \sum_{\ell=1}^{n} \frac{\ell(\ell-1)}{2} z'_\ell \geq \sum_{\ell=1}^{n} \frac{\ell(\ell-1)}{2} z^*_\ell$$

and inequality (4b) is also valid for $\mathbf{z}^*$. Note that the other constraints involving $\mathbf{z}$, i.e., (4e) are valid for $\mathbf{z}^*$ due to (16) and (17). This observation completes the proof. □

## References


Achterberg, T., Bixby, R. E., Gu, Z., Rothberg, E., & Weninger, D. (2020). Presolve reductions in mixed integer programming. *INFORMS Journal on Computing, 32*(2), 473–506.

Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics, 74*(1), 47.

Almeida, M. T., & Brás, R. (2019). The maximum $\ell$-triangle $k$-club problem: Complexity, properties, and algorithms. *Computers & Operations Research, 111*, 258–270.

Almeida, M. T., & Carvalho, F. D. (2012). Integer models and upper bounds for the 3-club problem. *Networks, 60*(3), 155–166.

Balasundaram, B., Butenko, S., & Trukhanov, S. (2005). Novel approaches for analyzing biological networks. *Journal of Combinatorial Optimization, 10*(1), 23–39.

Batagelj, V., & Mrvar, A. (2006). Pajek datasets. http://vlado.fmf.uni-lj.si/pub/networks/data/. Last accessed March, 2020.

Boguná, M., Pastor-Satorras, R., Díaz-Guilera, A., & Arenas, A. (2004). Models of social networks based on social distance attachment. *Physical Review E, 70*(5), 056122.

Bourjolly, J.-M., Laporte, G., & Pesant, G. (2000). Heuristics for finding $k$-clubs in an undirected graph. *Computers & Operations Research, 27*(6), 559–569.

Bourjolly, J.-M., Laporte, G., & Pesant, G. (2002). An exact algorithm for the maximum k-club problem in an undirected graph. *European Journal of Operational Research, 138*(1), 21–28.

Buchanan, A., & Salemi, H. (2020). Parsimonious formulations for low-diameter clusters. *Mathematical Programming Computation*. To appear

Butenko, S., & Wilhelm, W. (2006). Clique-detection models in computational biochemistry and genomics. *European Journal of Operational Research, 173*(1), 1–17. https://doi.org/10.1016/j.ejor.2005.05.026.

Carvalho, F. D., & Almeida, M. T. (2011). Upper bounds and heuristics for the 2-club problem. *European Journal of Operational Research, 210*(3), 489–494.

Carvalho, F. D., & Almeida, M. T. (2017). The $\ell$-triangle $k$-club problem. *Journal of Combinatorial Optimization, 33*(3), 814–846.

Davis, T. A., & Hu, Y. (2011). The University of Florida sparse matrix collection. *ACM Transactions on Mathematical Software (TOMS), 38*(1), 1–25.

DIMACS (2011). 10th DIMACS Implementation Challenge. Available at http://www.cc.gatech.edu/dimacs10/index.shtml, last accessed accessed March, 2020.

Dondi, R., & Lafond, M. (2019). On the tractability of covering a graph with 2-clubs. In *International symposium on fundamentals of computation theory* (pp. 243–257). Springer.

Dondi, R., Mauri, G., Sikora, F., & Zoppis, I. (2018). Covering with clubs: Complexity and approximability. In C. Iliopoulos, H. W. Leong, & W.-K. Sung (Eds.), *Combinatorial algorithms* (pp. 153–164). Cham: Springer International Publishing.

Dondi, R., Mauri, G., & Zoppis, I. (2019). On the tractability of finding disjoint clubs in a network. *Theoretical Computer Science, 777*, 243–251.

Garey, M. R., & Johnson, D. S. (2002). *Computers and intractability*: 29. W.H. Freeman.

Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., & Barabási, A.-L. (2007). The human disease network. *Proceedings of the National Academy of Sciences, 104*(21), 8685–8690.

Goodreau, S. M., Kitts, J. A., & Morris, M. (2009). Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks. *Demography, 46*(1), 103–125.

Graph Coloring and its Generalizations (2004). http://mat.gsia.cmu.edu/COLOR03/. Last accessed accessed March, 2020.

Gschwind, T., Irnich, S., Furini, F., & Calvo, R. W. (2020). A branch-and-price framework for decomposing graphs into relaxed cliques. *INFORMS Journal on Computing*.

Gurobi Optimization (2019). Gurobi optimizer reference manual. http://www.gurobi.com, last accessed accessed March, 2020.

Hagberg, A., Swart, P., & S Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX. *Technical Report*. Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

Hartung, S., Komusiewicz, C., & Nichterlein, A. (2015). Parameterized algorithmics and computational experiments for finding 2-clubs. *Journal of Graph Algorithms and Applications, 19*(1), 155–190. https://doi.org/10.7155/jgaa.00352.

Hartung, S., Komusiewicz, C., Nichterlein, A., & Suchỳ, O. (2015). On structural parameterizations for the 2-club problem. *Discrete Applied Mathematics, 185*, 79–92.

Jia, S., Gao, L., Gao, Y., Nastos, J., Wen, X., Huang, X., & Wang, H. (2018). Viewing the meso-scale structures in protein-protein interaction networks using 2-clubs. *IEEE Access, 6*, 36780–36797.

Kahruman-Anderoglu, S., Buchanan, A., Butenko, S., & Prokopyev, O. A. (2016). On provably best construction heuristics for hard combinatorial optimization problems. *Networks, 67*(3), 238–245.

Komusiewicz, C. (2016). Multivariate algorithmics for finding cohesive subnetworks. *Algorithms, 9*(1), 21.

Komusiewicz, C., Nichterlein, A., Niedermeier, R., & Picker, M. (2019). Exact algorithms for finding well-connected 2-clubs in sparse real-world graphs: Theory and experiments. *European Journal of Operational Research, 275*(3), 846–864.

Kunegis, J. (2013). Konect: The koblenz network collection. In *Proceedings of the 22nd international conference on world wide web* (pp. 1343–1350).

Laan, S., Intelligentie, B. O. K., Marx, M., Mokken, R., & van Doornik, J. (2012). Fast finding of 2-clubs. *University of Amsterdam, Informatics Institute, Amsterdam*.

Laan, S., Marx, M., & Mokken, R. J. (2016). Close communities in social networks: Boroughs and 2-clubs. *Social Network Analysis and Mining, 6*(1), 20.

Martins, P. (2010). Extended and discretized formulations for the maximum clique problem. *Computers & Operations Research, 37*(7), 1348–1358.

Miao, J., & Berleant, D. (2004). From paragraph networks to document networks.. In *Proceedings of the international conference on information technology (icc 2004)* (pp. 295–302). Citeseer.

Mokken, R. J. (1979). Cliques, clubs and clans. *Quality & Quantity, 13*(2), 161–173.

Nguyen, D., Shen, Y., & Thai, M. (2013). Detecting critical nodes in interdependent power networks for vulnerability assessment. *IEEE Transactions on Smart Grid, 4*(1), 151–159.

Pajouh, F. M., Balasundaram, B., & Hicks, I. V. (2016). On the 2-club polytope of graphs. *Operations Research, 64*(6), 1466–1481.

Pardalos, P. M., & Xue, J. (1994). The maximum clique problem. *Journal of Global Optimization, 4*(3), 301–328. https://doi.org/10.1007/BF01098364.

Pastukhov, G., Veremyev, A., Boginski, V., & Prokopyev, O. A. (2018). On maximum degree-based-quasi-clique problem: Complexity and exact approaches. *Networks, 71*(2), 136–152.

Pattillo, J., Veremyev, A., Butenko, S., & Boginski, V. (2013a). On the maximum quasi-clique problem. *Discrete Applied Mathematics, 161*(1-2), 244–257.

Pattillo, J., Youssef, N., & Butenko, S. (2013b). On clique relaxation models in network analysis. *European Journal of Operational Research, 226*(1), 9–18.

Rossi, R. A., & Ahmed, N. K. (2015). The network data repository with interactive graph analytics and visualization. In *Proceedings of the twenty-ninth aaai conference on artificial intelligence*.

Shen, Y., Nguyen, N., Xuan, Y., & Thai, M. (2013). On the discovery of critical links and nodes for assessing network vulnerability. *IEEE/ACM Transactions on Networking, 21*(3), 963–973.

Terveen, L., Hill, W., & Amento, B. (1999). Constructing, organizing, and visualizing collections of topically related web resources. *ACM Transactions on Computer-Human Interaction (TOCHI), 6*(1), 67–94.

Veremyev, A., & Boginski, V. (2012a). Identifying large robust network clusters via new compact formulations of maximum *k*-club problems. *European Journal of Operational Research, 218*(2), 316–326.

Veremyev, A., & Boginski, V. (2012b). Robustness and strong attack tolerance of low-diameter networks. In A. Sorokin, R. Murphey, M. T. Thai, & P. M. Pardalos (Eds.), *Dynamics of information systems: Mathematical foundations* (pp. 137–156). Springer.

Veremyev, A., Prokopyev, O. A., Boginski, V., & Pasiliao, E. L. (2014). Finding maximum subgraphs with relatively large vertex connectivity. *European Journal of Operational Research, 239*(2), 349–362.

Veremyev, A., Prokopyev, O. A., Butenko, S., & Pasiliao, E. L. (2016). Exact mip-based approaches for finding maximum quasi-cliques and dense subgraphs. *Computational Optimization and Applications, 64*(1), 177–214.

Verma, A., Buchanan, A., & Butenko, S. (2015). Solving the maximum clique and vertex coloring problems on very large sparse networks. *INFORMS Journal on Computing, 27*(1), 164–177.

Vielma, J. P. (2015). Mixed integer linear programming formulation techniques. *Siam Review, 57*(1), 3–57.

Watts, D., & Strogatz, S. (1998). Collective dynamics of “small-world” networks. *Nature, 393*(6684), 440–442.

Yezerska, O., Pajouh, F. M., & Butenko, S. (2017). On biconnected and fragile subgraphs of low diameter. *European Journal of Operational Research, 263*(2), 390–400.

Yezerska, O., Pajouh, F. M., Veremyev, A., & Butenko, S. (2019). Exact algorithms for the minimum s-club partitioning problem. *Annals of Operations Research, 276*(1–2), 267–291.