Discrete Optimization

An analytical comparison of the LP relaxations of integer models for the k-club problem

Maria Teresa Almeida *, Filipa D. Carvalho

Instituto Superior de Economia e Gestão, Universidade Técnica de Lisboa, Rua do Quelhas 6, 1200-781 Lisboa, Portugal
CIO, FC, Universidade de Lisboa, Bloco C6, Piso 4, 1749-016 Lisboa, Portugal

ABSTRACT

Given an undirected graph $G = (V, E)$, a k-club is a subset of nodes that induces a subgraph with diameter at most k. The k-club problem is to find a maximum cardinality k-club. In this study, we use a linear programming relaxation standpoint to compare integer formulations for the k-club problem. The comparisons involve formulations known from the literature and new formulations, built in different variable spaces. For the case $k = 3$, we propose two enhanced compact formulations. From the LP relaxation standpoint these formulations dominate all other compact formulations in the literature and are equivalent to a formulation with a non-polynomial number of constraints. Also for $k = 3$, we compare the relative strength of LP relaxations for all formulations examined in the study (new and known from the literature). Based on insights obtained from the comparative study, we devise a strengthened version of a recursive compact formulation in the literature for the k-club problem ($k > 1$) and show how to modify one of the new formulations for the case $k = 3$ in order to accommodate additional constraints recently proposed in the literature.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Given an undirected graph $G = (V, E)$ and an integer k, a k-club S is a cluster of nodes such that any two of its elements can reach each other through at most $k - 1$ intermediate cluster members. If $k = 1$, then every cluster member is directly linked to all the others and S is a clique (Bomze, Budinich, Pardalos, & Pelillo, 1999; Sorensen, 2004; Alidaee, Glover, Kochenberger, & Wang, 2007; Martins, 2010). If $k > 1$, a k-club can be interpreted as a distance-based relaxation of a clique.

The need for graph models to represent clusters was first noted during the 1970s in the literature of social networks analysis (Alba, 1973; Mokken, 1979). It was also noted then that the clique model is too restrictive for many real-world applications for not all social relations require face-to-face interaction. In fact, social relations are frequently established through intermediaries. This explains the importance of clique relaxation models for social sciences.

Recently, the OR community has produced a significant number of studies on clique-related models with applications in both the social sciences as well as other fields such as computational biology (Błażewicz, Formanowicz, & Kasprak, 2005; Balasundaram,

Butenko, & Trukhanov, 2005; Butenko & Wilhelm, 2006) and data mining (Boginski, Butenko, & Pardalos, 2006).

For any $k > 1$, the k-club problem is to find a maximum cardinality k-club of a graph. The k-club problem is NP-hard (Bourjolly, Laporte, & Pesant, 2002). The computational performance of all the approaches reported in the literature depends on the value of k and the edge density of the graph. To be able to devise better exact and approximate approaches, it is important to gain a deeper understanding of the problem from a theoretical point of view. A detailed study of the non-hereditary nature of k-clubs for $k > 1$ is presented in Mahdavi Pajouh and Balasundaram (2012). In this study, we investigate the relative strength of the LP relaxations of integer programming formulations for the k-club problem built in different variable spaces.

The rest of the paper is organized as follows. In Section 2 we provide definitions, terminology, and notation. Section 3 contains the motivation and an overview of the study. Section 4 contains a review of the literature on k-club models. Sections 5 and 6 are devoted to the comparison of models and the development of enhanced versions, from the linear programming standpoint. In Section 7 we derive a formulation for a problem that has recently been proposed in the literature: a variant of the 3-club problem with additional robustness constraints. Conclusions are presented in Section 8.

2. Definitions, terminology, and notation

Given an undirected graph $G = (V, E)$ and a node $v \in V$, we represent by N_v the set of all nodes adjacent to v, and call it the

---

* Corresponding author. Address: Instituto Superior de Economia e Gestão, Departamento de Matemática, Rua do Quelhas 6, 1200-781 Lisboa, Portugal. Tel.: +351 213 925 800; fax: +351 213 922 781.
E-mail addresses: talmeida@iseg.utl.pt (M.T. Almeida), filipadc@iseg.utl.pt (F.D. Carvalho).

neighbourhood of v. The distance $dist_G(u, v)$ between two nodes u and v is the minimum number of edges needed to link them, and the diameter $diam(G)$ is the maximum distance between any u and v. If $S \subseteq V$, the subgraph induced by S is $G_{[S]} = (S, E(S))$, where $E(S)$ is the set of edges with both end nodes in S. For any $k \geqslant 1$, if $diam(G_{[S]}) \leqslant k$, S is a k-club. A k-club $S \subseteq V$ can be represented by its incidence vector $X = (x_1, \ldots, x_{|V|})$. If C is a subset of nodes and s, $t \in V \backslash C$, we say that C is an $s-t$ node cut set if C intersects all chains that link s and t.

Formulations with variables to represent only nodes are called natural whereas formulations with additional variables to represent edges, chains or paths are called extended. A formulation is compact if the number of its variables and constraints is bounded from above by a polynomial function of the number of nodes. If an integer linear programming formulation is represented by [F], then $[F]_{LP}$ represents its linear programming (LP) relaxation, their optimal values are represented by $Z[F]$ and $Z[F]_{LP}$, and their feasible sets are represented by $S[F]$ and $S[F]_{LP}$, respectively. Feasible solutions to LP relaxations of natural (resp. extended) integer formulations are represented by $\overline{X}$, (resp. $(\overline{X}, \overline{Y})$, $(\overline{X}, \overline{V})$, and $(\overline{X}, \overline{Z})$). Integer formulations are compared by projecting their LP relaxation feasible sets onto the space of node variables. Given a polyhedron $P = \{(x, y) \in \mathfrak{R}^{n \times q} : Ax + By \leqslant b\}$, where A, B, an b have m rows, the projection of P onto the x-space is defined as $\text{Proj}_x(P) = \{x \in \mathfrak{R}^n : \exists y \in \mathfrak{R}^q (x, y) \in P\}$. For more on polyhedral theory and projection the reader is referred to Nemhauser and Wolsey (1988) and Balas (2005). Given two integer formulations [F1] and [F2], whose objective function is the same and is to be maximized, if $\text{Proj}_x(S[F1]_{LP})$ is a subset of $\text{Proj}_x(S[F2]_{LP})$, then $Z[F1]_{LP} \leqslant Z[F2]_{LP}$, and we say that $[F1]_{LP}$ is stronger than $[F2]_{LP}$ and formulation [F1] dominates [F2]. If $\text{Proj}_x(S[F1]_{LP}) = \text{Proj}_x(S[F2]_{LP})$, then [F1] and [F2] are LP-equivalent.

3. Motivation and overview

Like many other combinatorial optimization problems, the k-club-problem has alternative integer linear programming formulations that are stated with different sets of variables. The first formulation in the literature, proposed in Bourjolly et al. (2002), is a compact extended formulation with node and chain variables. Following Almeida and Carvalho (2012), we shall call it the chain formulation and denote it by [F_C]. Since two-edge chains $t = (i, r, j)$ that link nodes i and j can be identified by the central node r, there is no need for chain variables to represent them. Based on this observation, Bourjolly et al. (2002) also proposed a simplified natural formulation for the case $k = 2$. Very recently, four new formulations have been proposed for k-clubs. Veremyev and Boginski (2012) present a compact extended recursive formulation (henceforth referred to as [F_R]) derived from a nonlinear model by a linearization procedure that takes into account the structure of k-clubs. For $k = 3$, Almeida and Carvalho (2012) propose a node cut set natural formulation [F_S] with a non-polynomial number of constraints and a compact extended formulation [F_N] with node and edge variables. For the case $k = 2$, Carvalho and Almeida (2011) devise a natural formulation whose constraints define facets of the 2-club polytope P_{2c}. This formulation, based on Bourjolly et al.'s model, is obtained by lifting redundant constraints.

Many authors have carried out comparative studies of alternative formulations for important combinatorial optimization problems – *e.g.* tree problems (Magnanti & Wolsey, 1995) and travelling salesman-related problems (Gouveia & Voss, 1995; Oncan, Altinel, & Laporte, 2009). To the best of our knowledge, no similar study can be found in the literature for the k-club problem. Veremyev and Boginski (2012, pp. 320–321) comment on the tightness of binary formulations based on the results

obtained with a set of test instances. They guess that their recursive formulation "*is generally at least as good (or, at the very least, not substantially worse) than the one by* Bourjolly et al. (2002)" and that their formulation "*is rather tight for larger values of k*". Almeida and Carvalho (2012) have proved that formulation [F_N] dominates chain formulation [F_C] for $k = 3$, and they present a weakened version of [F_N] which is LP-equivalent to [F_C]. The authors also prove that for $k = 3$ node cut set formulation [F_S] dominates [F_N], but do not show how to strengthen the latter to obtain a formulation LP-equivalent to the former.

With this study we aim to investigate the relative strength of the LP relaxations of different formulations for the maximum k-club problem. Since the formulations are stated with different sets of variables, the comparisons are made by projecting the LP feasible sets onto the space of node variables. The insight gained into the k-club structure by means of these comparisons may be of help to explore possible ways to formulate related problems, as illustrated in Section 7 with the R-robust k-club problem recently proposed in the literature by Veremyev and Boginski (2012).

First, for the case $k = 3$, we show how to modify [F_C] in order to obtain an enhanced version [F_EC], which is LP-equivalent to [F_S]. Then, we address the tightness of the recursive formulation [F_R] in comparison with the tightness of [F_C] and show that the conclusion depends on the value of k: for $k = 3$ [F_R] dominates [F_C] but the dominance does not hold for $k = 4$. For $k = 3$, we also show that [F_R] is dominated by [F_EC].

Based on the insight gained from the comparisons of [F_R], [F_C], and [F_EC] for $k = 3$, we derive a strengthened version of [F_R] that is valid for any $k > 1$, denoted by [F_ER]. We then show that for $k = 3$ [F_ER] and [F_EC] are LP-equivalent.

We conclude the paper with a formulation for the R-robust 3-club problem, introduced in the literature by Veremyev and Boginski (2012). To the best of our knowledge, there is in the literature no other formulation for the robustness condition when $k > 2$.

4. Review of integer programming formulations for the k-club problem

To set the grounds for the comparisons presented in the following sections, we shall first review integer programming formulations in the literature.

To simplify the notation, we will denote by $\mathscr{P}^k$ the set of all pairs of nodes that cannot belong simultaneously to a k-club, because their distance in $G = (V, E)$ is greater than k, and by $\mathscr{N}^k$ the set of all pairs of nonadjacent nodes whose distance in $G = (V, E)$ does not exceed k.

4.1. Chain formulation

The formulation proposed in Bourjolly et al. (2002), with node and chain variables, can be presented as follows. For any two non-adjacent nodes $i, j \in V$, let C_{ij}^k be the set of all chains, of length at most k, that link i and j, and let $C = \cup_{i,j \in V} C_{ij}^k$. Let y_t, $t \in C$, be a binary variable associated with chain t and V_t be the set of all nodes in the chain.

A maximum k-club is an optimal solution for

$$[\text{F\_C}] \quad \max \quad Z = \sum_{i \in V} x_i$$

$$\text{s.t.} \quad x_i + x_j \leqslant 1 \quad \{i, j\} \in \mathscr{P}^k \tag{1}$$

$$x_i + x_j \leqslant 1 + \sum_{t \in C_{ij}^k} y_t \quad \{i, j\} \in \mathscr{N}^k \tag{2}$$

$$y_t \leqslant x_r \quad t \in C, \quad r \in V_t \tag{3}$$

$$x_i \in \{0, 1\} \quad i \in V \tag{4}$$

$$y_t \in \{0, 1\} \quad t \in C \tag{5}$$

Constraints (1) guarantee that if $dist_G(i,j) > k$, then nodes i and j are not both selected for the k-club. Constraints (1) will be referred to as packing constraints. Constraints (2) impose that a pair of non-adjacent nodes in a k-club must be linked by at least one selected chain with at most k edges. Constraints (3) impose the inclusion in the k-club of all nodes in selected chains. Conditions (4) and (5) define the variables as binary.

Note that for every pair $\{i,j\} \in \mathcal{N}^k$ and every chain $t \in C_{ij}^k$, $r \in V_t \setminus \{i,j\}$ can be substituted for $r \in V_t$, in constraints (3), since in any solution such that $\bar{y}_t > \min\{\bar{x}_i, \bar{x}_j\}$ the value of variable y_t can be decreased to $\min\{\bar{x}_i, \bar{x}_j\}$, without altering the solution in the x-space.

The number of chain variables y_t and constraints in [F_C] is $O(|V|^{k+1})$.

4.2. Recursive formulation

Veremyev and Boginski (2012) use a different approach to devise a formulation for the k-club problem. First, they develop a nonlinear model. Then, they transform it into a linear model with binary variables x_i $(i = 1,\ldots,|V|)$ and integer variables $w_{ij}^{(l)}(i,j = 1,\ldots,|V|; l = 2,\ldots,k)$ to represent the number of paths of length l that link nonadjacent nodes i and j. Finally, taking into account the k-club structure, they devise a recursive binary formulation with node variables $x_i(i = 1,\ldots,|V|)$ and binary variables $v_{ij}^{(l)}(i,j = 1,\ldots,|V|; l = 2,\ldots,k)$ such that $v_{ij}^{(l)} = 1$ if and only if there is at least one path of length l from node i to node j in the subgraph induced by the set $S \subseteq V$ whose incidence vector is $X = (x_1,\ldots,x_{|V|})$.

The recursive formulation can be presented as follows:

[F_R] $\max \quad Z = \sum_{i \in V} x_i$

s.t. $x_i + x_j \leqslant 1 + \sum_{l=2}^{k} v_{ij}^{(l)} \quad i,j \in V, (i,j) \notin E$ (6)

for $i,j \in V, j > i$,

$v_{ij}^{(2)} \leqslant x_i \quad v_{ij}^{(2)} \leqslant x_j \quad v_{ij}^{(2)} \leqslant \sum_{k \in (N_i \cap N_j)} x_k$ (7)

$v_{ij}^{(2)} \geqslant \frac{1}{|V|}\left(\sum_{k \in (N_i \cap N_j)} x_k\right) + (x_i + x_j - 2)$ (8)

for $l = 3,\ldots,k; i,j \in V, j > i$,

$v_{ij}^{(l)} \leqslant x_i \quad v_{ij}^{(l)} \leqslant \sum_{k \in N_i} v_{kj}^{(l-1)}$ (9)

$v_{ij}^{(l)} \geqslant \frac{1}{|V|}\left(\sum_{k \in N_i} v_{kj}^{(l-1)}\right) + (x_i - 1)$ (10)

$x_i \in \{0,1\} \quad i \in V$ (4)

$v_{ij}^{(l)} \in \{0,1\} \quad i,j \in V; l = 2,\ldots,k$ (11)

Constraints (6) impose that each pair of selected nodes must be linked by at least a path with at most k edges. Constraints (7) and (8) define paths with two edges. Constraints (9) and (10) define the recursion for paths with three or more edges. Constraints (11) define the recursion variables as binary.

Recursive formulation [F_R] has size $O(k|V|^2)$.

4.3. Neighbourhood and node cut set formulations ($k = 3$)

For the case $k = 3$, Almeida and Carvalho (2012) propose two alternative formulations – neighbourhood formulation [F_N] and

node cut set formulation [F_S]. The rationale for [F_N] and [F_S] is summarized next.

Consider a pair of nonadjacent nodes i and j of $G = (V,E)$. They may be included in a 3-club S only if at least one of the following conditions holds:

(I) There is a node r in S that is a neighbour of i and j, i.e., $r \in (N_i \cap N_j)$;
(II) There are two adjacent nodes in S, p and q, such that p is a neighbour of i, and q is a neighbour of j, i.e., $p \in N_i$, $q \in N_j$, and $(p,q) \in E$.

Condition (I) holds if and only if $dist_{G[S]}(i,j) = 2$.

If condition (II) holds and $p \in (N_i \cap N_j)$ or $q \in (N_i \cap N_j)$, then $dist_{G[S]}(i,j) = 2$ and condition (I) holds as well. So, in (II), we can consider only nodes $p \in (N_i \setminus N_j)$ and $q \in (N_j \setminus N_i)$. Let E_{ij} denote the set of edges that link those nodes:

$$E_{ij} = \{(p,q) \in E : p \in (N_i \setminus N_j), q \in (N_j \setminus N_i)\}.$$

Let us now associate a variable z_{ij} with each edge $(i,j) \in E$.

A maximum 3-club is an optimal solution for

[F_N] $\max \quad Z = \sum_{i \in V} x_i$

s.t. $x_i + x_j \leqslant 1 \quad \{i,j\} \in \mathcal{P}^3$ (12)

$x_i + x_j \leqslant 1 + \sum_{r \in (N_i \cap N_j)} x_r + \sum_{(p,q) \in E_{ij}} z_{pq} \quad \{i,j\} \in \mathcal{N}^3$ (13)

$z_{ij} \leqslant x_i, \; z_{ij} \leqslant x_j, \; z_{ij} \geqslant x_i + x_j - 1 \quad (i,j) \in E$ (14)

$x_i \in \{0,1\} \quad i \in V$ (4)

$z_{ij} \geqslant 0 \quad (i,j) \in E$ (15)

Conditions (12) are packing constraints. Neighbourhood constraints (13) impose that two nonadjacent nodes i and j cannot be both in a 3-club unless a common neighbour is in the 3-club or a pair of neighbours, p and q, of i and j respectively, linked by an edge, are in the 3-club. Constraints (14) guarantee that an edge (i,j) is used if and only if both its end nodes belong to the 3-club. Constraints (15) define edge variables as nonnegative.

Conditions $z_{ij} \geqslant x_i + x_j - 1$ can be ignored because in any solution with $\bar{z}_{ij} < \bar{x}_i + \bar{x}_j - 1$ the value of variable z_{ij} can be increased to $\min\{\bar{x}_i, \bar{x}_j\}$, without altering the solution in the x-space.

Note that, if in the last term of conditions (13) $E(N_i, N_j)$ is substituted for E_{ij}, the resulting formulation is LP-equivalent to [F_C] (Almeida & Carvalho, 2012); if in [F_C] for each pair of nodes $\{i,j\} \in \mathcal{N}^3$ all variables y_t associated with three-edge chains $t = (i,p,q,j)$ with $p \in (N_i \cap N_j)$ or $q \in (N_i \cap N_j)$ are removed, the resulting formulation is LP-equivalent to [F_N] (Almeida & Carvalho, 2012).

Neighbourhood formulation [F_N] has $|V| + |E|$ variables and $\frac{|V|^2 - |V|}{2} + 2|E|$ constraints.

[F_S] is a natural formulation for the 3-club problem based on node cut sets of auxiliary graphs built for pairs $\{i,j\} \in \mathcal{N}^3$. This formulation is described next.

Consider again a pair of nonadjacent nodes i and j of $G = (V,E)$ and $E_{ij} = \{(p,q) \in E : p \in (N_i \setminus N_j), q \in (N_j \setminus N_i)\}$. E_{ij} is the set of inner edges of three-edge chains that link nodes i and j, and whose inner nodes are not in $N_i \cap N_j$. Let V_{ij} represent the set of their end nodes. Each node $v \in V_{ij}$ is either adjacent to node i or to node j. Therefore V_{ij} can be partitioned into subsets $A_{ij} = \{v \in V_{ij} : v \in (N_i \setminus N_j)\}$ and $B_{ij} = \{v \in V_{ij} : v \in (N_j \setminus N_i)\}$.

Let us now associate with i and j a subgraph $G_{ij} = (N_{ij}, H_{ij})$, where $N_{ij} = V_{ij} \cup \{i,j\}$ and $H_{ij} = E_{ij} \cup \{(i,v) \in E : v \in A_{ij}\} \cup \{(v,j) \in E : v \in B_{ij}\}$.

To include nodes i and j in a 3-club S, it is also necessary to include a node $r \in (N_i \cap N_j)$ or a node of each set $S_{ij} \subseteq V_{ij}$ such that $E(V_{ij} \setminus S_{ij}) = \varnothing$, since otherwise $dist_{G[S]}(i,j) > 3$. These sets S_{ij} are $i - j$ node cut sets. The set of all minimal S_{ij} will be denoted by $\mathcal{S}^{ij}$.

A maximum 3-club is an optimal solution for

$$[\text{F\_S}] \quad \max \quad Z = \sum_{i \in V} x_i$$

$$\text{s.t.} \quad x_i + x_j \leqslant 1 \quad \{i,j\} \in \mathscr{P}^3 \tag{12}$$

$$x_i + x_j \leqslant 1 + \sum_{r \in (N_i \cap N_j)} x_r + \sum_{s \in S_{ij}} x_s \quad \{i,j\} \in \mathscr{N}^3, S_{ij} \in \mathcal{S}^{ij} \tag{16}$$

$$x_i \in \{0,1\} \quad i \in V \tag{4}$$

Conditions (12) are packing constraints. Node cut set constraints (16) impose that, if two nonadjacent nodes i and j are both selected for the 3-club, then a common neighbour of i and j or one node from each $S_{ij} \in \mathcal{S}^{ij}$ is also selected for the 3-club.

Node cut set formulation [F_S] dominates neighbourhood formulation [F_N] (Almeida & Carvalho, 2012).

5. Enhanced chain formulation for the case k = 3

For $k = 3$, Almeida and Carvalho (2012) show that chain formulation [F_C] is dominated by formulation [F_N], which is in turn dominated by natural formulation [F_S]. They also show how to modify [F_C] to obtain a chain formulation, LP-equivalent to [F_N]. All these formulations are compact, except for [F_S], which has a non-polynomial number of node cut set constraints. The comparisons in Almeida and Carvalho (2012) suggest that if we aim to obtain a chain formulation LP-equivalent to [F_S], we have to devise more stringent constraints to link node and chain variables. The intuition for the new linking constraints is provided by the example in Almeida and Carvalho (2012, p. 157): the dominance of [F_S]$_{LP}$ over [F_N]$_{LP}$ may be strict only when there is at least one pair $\{i,j\} \in \mathscr{N}^3$ and a node $v \in V_{ij}$ such that $\deg(v)_{G_{ij|V_{ij}}} > 1$. In this case, node v is shared by two or more three-edge chains whose central edges are represented by variables z_{pq}, with $(p,q) \in E_{ij}$. In a non-integer feasible solution $(\overline{X}, \overline{Z})$ for [F_N]$_{LP}$, if $v \in A_{ij}$, it may happen that $\sum_{q:(v,q)\in E_{ij}} \overline{z}_{vq} > \overline{x}_v$ or, if $v \in B_{ij}$, it may happen that $\sum_{p:(p,v)\in E_{ij}} \overline{z}_{pv} > \overline{x}_v$. In either case, the solution can be cut off by imposing the condition that the sum of the values of all variables that represent the chains that share node v cannot be greater than $\overline{x}_v$. To formulate these new linking constraints, we shall rewrite chain variables y_t, $t = (i,p,q,j)$, as y_{pq}^{ij}, in order to make the identification of inner nodes immediate.

The enhanced version of chain formulation [F_C] for $k = 3$ we propose in this section combines the new linking constraints sketched above with the rationale used to remove y_t variables associated with three-edge chains $t = (i,p,q,j)$ with at least one inner node in $N_i \cap N_j$ (see Section 4.3).

Let us consider chain formulation [F_C] for $k = 3$.

For each pair of nodes $\{i,j\} \in \mathscr{N}^3$, let $C_{ij}^{=2}$ (resp. $C_{ij}^{=3}$) represent all chains with two (resp. three) edges that link i and j.

Each chain $t = (i,r,j) \in C_{ij}^{=2}$ can be represented by its central node r. Therefore, all variables y_t that represent chains of $C_{ij}^{=2}$ can be discarded and constraints (2) rewritten as

$$x_i + x_j \leqslant 1 + \sum_{r \in (N_i \cap N_j)} x_r + \sum_{t \in C_{ij}^{=3}} y_t \quad \{i,j\} \in \mathscr{N}^3 \tag{17}$$

As discussed in Section 4.3, for each $\{i,j\} \in \mathscr{N}^3$, variables y_t associated with chains $t = (i,p,q,j) \in C_{ij}^{=3}$ with $p \in (N_i \cap N_j)$ or $q \in (N_i \cap N_j)$ can be eliminated. To simplify the notation, for all

other chains $t = (i,p,q,j) \in C_{ij}^{=3}$ let us write y_{pq}^{ij}, instead of y_t, with constraints (3) rewritten as

$$y_{pq}^{ij} \leqslant x_i \quad y_{pq}^{ij} \leqslant x_j \quad \{i,j\} \in \mathscr{N}^3 \ (p,q) \in E_{ij} \tag{18}$$

$$y_{pq}^{ij} \leqslant x_p \quad y_{pq}^{ij} \leqslant x_q \quad \{i,j\} \in \mathscr{N}^3 \ (p,q) \in E_{ij} \tag{19}$$

and constraints (5) rewritten as

$$y_{pq}^{ij} \in \{0,1\} \quad \{i,j\} \in \mathscr{N}^3 \ (p,q) \in E_{ij} \tag{20}$$

Constraints (2) in [F_C] can now be lifted to

$$x_i + x_j \leqslant 1 + \sum_{r \in (N_i \cap N_j)} x_r + \sum_{(p,q) \in E_{ij}} y_{pq}^{ij} \quad \{i,j\} \in \mathscr{N}^3 \tag{21}$$

A chain that includes edge (i,p) (resp. (q,j)) can only be used if node p (resp. q) is in the 3-club, and to include a pair of nonadjacent nodes i and j in the 3-club, there is no need to select more than one chain with three edges. Therefore, constraints (19) can be lifted to

$$\sum_{q:(p,q) \in E_{ij}} y_{pq}^{ij} \leqslant x_p \quad \{i,j\} \in \mathscr{N}^3 \quad p \in A_{ij} \tag{22}$$

$$\sum_{p:(p,q) \in E_{ij}} y_{pq}^{ij} \leqslant x_q \quad \{i,j\} \in \mathscr{N}^3 \quad q \in B_{ij} \tag{23}$$

An enhanced chain formulation for the 3-club problem is

$$[\text{F\_EC}] \quad \max \left\{ \sum_{i \in V} x_i : (12), (21), (18), (22), (23), (4), (20) \right\}.$$

By construction, [F_EC] dominates [F_C].

Proposition 1. *For $k = 3$, $Z [\text{F\_EC}]_{LP} \leqslant Z [\text{F\_C}]_{LP}$.*

Note that constraints (18) can be ignored because in any solution with $\overline{y}_{pq}^{ij} > \min \{\overline{x}_i, \overline{x}_j\}$ the value of variable y_{pq}^{ij} can be decreased to $\min \{\overline{x}_i, \overline{x}_j\}$, without altering the solution in the x-space. Constraints (18) can also be written as $\sum_{(p,q) \in E_{ij}} y_{pq}^{ij} \leqslant x_i$ and $\sum_{(p,q) \in E_{ij}} y_{pq}^{ij} \leqslant x_j$ because the upper bounding constraints $x_v \leqslant 1$, for all v, dominate (21) if $\sum_{(p,q) \in E_{ij}} \overline{y}_{pq}^{ij} > \min \{\overline{x}_i, \overline{x}_j\}$. This version of constraints (18) imposes the extra condition that at most one chain with three edges can be selected to link nodes i and j in the induced subgraph. The condition is not restrictive for the 3-club problem, but it is incompatible with multiple-chain conditions for robustness that have been recently proposed in the
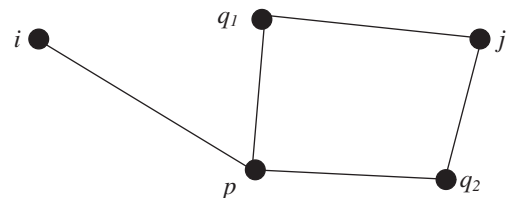
Fig. 1. (a) Consider a feasible solution for [F_C]$_{LP}$ in which $\overline{x}_i = \frac{2}{3}$, $\overline{x}_j = 1$, $\overline{x}_p = \overline{x}_{q_1} = \overline{x}_{q_2} = \frac{1}{3}$, and $\overline{y}_{t_1} = \overline{y}_{t_2} = \frac{1}{3}$, where $t_1 = (i,p,q_1,j)$ and $t_2 = (i,p,q_2,j)$. In the corresponding solution for [F_EC]$_{LP}$, to comply with constraint $y_{p,q_1}^{ij} + y_{p,q_2}^{ij} \leqslant x_p$, constraint (21) for the pair $\{i,j\}$ is violated. (b) Consider again a feasible solution for [F_C]$_{LP}$ in which $\overline{x}_i = \frac{2}{3}$, $\overline{x}_j = 1$, $\overline{x}_p = \overline{x}_{q_1} = \overline{x}_{q_2} = \frac{1}{3}$, and $\overline{y}_{t_1} = \overline{y}_{t_2} = \frac{1}{3}$, where $t_1 = (i,p,q_1,j)$ and $t_2 = (i,p,q_2,j)$. In the corresponding solution for [F_R]$_{LP}$, constraint (6) for the pair $\{i,j\}$ is violated because $\overline{v}_{pj}^{(2)} \leqslant \min \{\overline{x}_p, \overline{x}_j, \overline{x}_{q_1} + \overline{x}_{q_2}\} = \frac{1}{3}$, $\overline{v}_{ij}^{(3)} \leqslant \min \{\overline{x}_i, \overline{v}_{pj}^{(2)}\} \leqslant \frac{1}{3}$, and $\overline{v}_{ij}^{(2)} = 0$. (c) Consider a solution in the x-space such that $\overline{x}_i = \frac{2}{3}$, $\overline{x}_j = 1$, $\overline{x}_p = \overline{x}_{q_1} = \overline{x}_{q_2} = \frac{1}{3}$. In the corresponding solution for [F_R]$_{LP}$, constraint (6) for the pair $\{i,j\}$ is violated (see Section 4.2). By contrast, $\overline{z}_{p,q_1} = \overline{z}_{p,q_2} = \frac{1}{3}$ yields a feasible solution for [F_N]$_{LP}$.

literature (see Section 7). By contrast, the lifting of constraints (19) to constraints (22) and (23) strengthens the LP relaxation of the integer model and paves the way to the introduction of robustness conditions.

The dominance of [F_EC] over [F_C] may be strict. An illustration is provided by the graph depicted in Fig. 1(a).

For $k = 3$, the number of variables in [F_EC] and [F_C] is $O(|V|^4)$. In practice, the number of variables in [F_EC] is significantly smaller than the number of variables in [F_C] because all variables that represent chains with two edges as well as many variables that represent chains with three edges are eliminated.

6. Comparison of formulations for k = 3

In this section, we compare the LP relaxations of the integer models described in Sections 4 and 5, for the case $k = 3$. In these relaxations, constraints (4), (5), (11), and (20) are replaced by

$$0 \leqslant x_i \leqslant 1 \quad i \in V \tag{4'}$$

$$y_t \geqslant 0 \quad t \in C \tag{5'}$$

$$v_{ij}^{(l)} \geqslant 0 \quad i,j \in V; l = 2, \ldots, k \tag{11'}$$

$$y_{pq}^{ij} \geqslant 0 \quad \{i,j\} \in \mathcal{N}^3 \quad (p,q) \in E_{ij} \tag{20'}$$

In Section 6.1, we establish the LP-equivalence of [F_EC] and [F_S]. In Section 6.2, we show that formulation [F_R] is dominated by enhanced chain formulation [F_EC] and dominates chain formulation [F_C]. Based on the comparative study made for the case $k = 3$, we propose an enhanced version of [F_R] which is valid for the k-club problem, for any $k > 1$. We then establish the LP-equivalence between [F_EC] and this enhanced version of [F_R] for $k = 3$. In Section 6.3, we show that there is no dominance relation between formulations [F_N] and [F_R].

6.1. Comparison of the LP relaxations of [F_EC] and [F_S]

The rationale used to derive formulation [F_EC] suggests that with constraints (22) and (23) we may obtain a chain formulation as strong as [F_S] from the LP relaxation standpoint. In this section we present a formal proof that [F_EC] and [F_S] are, in fact, LP-equivalent.

To prove that [F_EC] and [F_S] are LP-equivalent, we shall show that the feasible set of [F_S]$_{LP}$ is the projection of the feasible set of [F_EC]$_{LP}$ onto the x-space.

Let $(\overline{X}, \overline{Y})$ be any feasible solution for [F_EC]$_{LP}$.

For any $\{i,j\} \in \mathcal{N}^3$ and $S_{ij} \in \mathcal{S}^{ij}$,

$$\sum_{(p,q)\in E_{ij}} \bar{y}_{pq}^{ij} = \sum_{\substack{(p,q)\in E_{ij} \\ p\in(A_{ij}\cap S_{ij}) \\ q\in B_{ij}}} \bar{y}_{pq}^{ij} + \sum_{\substack{(p,q)\in E_{ij} \\ p\in(A_{ij}\setminus S_{ij}) \\ q\in(B_{ij}\cap S_{ij})}} \bar{y}_{pq}^{ij} \leqslant \sum_{p\in(A_{ij}\cap S_{ij})} \bar{x}_p + \sum_{q\in(B_{ij}\cap S_{ij})} \bar{x}_q = \sum_{s\in S_{ij}} \bar{x}_s.$$

Therefore, $\overline{X}$ satisfies all node cut set constraints (16), for the pair $\{i,j\} \in \mathcal{N}^3$. Since $\{i,j\} \in \mathcal{N}^3$ was chosen arbitrarily, $\overline{X}$ is feasible for [F_S]$_{LP}$.

Conversely, let $\overline{X}$ be any feasible solution for [F_S]$_{LP}$, and let us build a feasible solution $(\overline{X}, \overline{Y})$ for [F_EC]$_{LP}$.

For every $\{i,j\} \in \mathcal{N}^3$, let us consider subgraph $G_{ij} = (N_{ij}, H_{ij})$ (see Section 4.3), and a layered digraph $\mathscr{D}_{ij}$ built from G_{ij} as follows.

The layers of $\mathscr{D}_{ij}$ are: $L_1 = \{i\}$, $L_2 = A_{ij}$, $L_3 = B_{ij}$, and $L_4 = \{j\}$. For each edge in H_{ij}, we create a directed arc in $\mathscr{D}_{ij}$ and define the arc capacities as:

$$\text{cap}(u,v) = \begin{cases} \bar{x}_v & \text{if } u = i \quad \text{and} \quad v \in A_{ij} \\ \infty & \text{if } u \in A_{ij} \quad \text{and} \quad v \in B_{ij} \\ \bar{x}_u & \text{if } u \in B_{ij} \quad \text{and} \quad v = j \end{cases}$$

Let f be a maximum $i - j$ flow in $\mathscr{D}_{ij}$. By construction, value$(f) = \sum_{(p,q)\in E_{ij}} f_{pq}$. Let us set $\bar{y}_{pq}^{ij} = f_{pq}$ for all $(p,q) \in E_{ij}$. By the max-flow min-capacity cut theorem, value(f) is also the capacity of an $i - j$ cut of minimum capacity in $\mathscr{D}_{ij}$. Hence, value$(f) = \sum_{s\in C_{ij}} \bar{x}_s$, where C_{ij} is the set of nodes of V_{ij} that are incident to the arcs in a minimum capacity $i - j$ cut in $\mathscr{D}_{ij}$. Therefore, value$(f) = \min\left\{\sum_{s\in S_{ij}} \bar{x}_s : S_{ij} \in \mathcal{S}^{ij}\right\}$, and constraints (16) for $\{i,j\}$ are satisfied. Since the pair $\{i,j\}$ was chosen arbitrarily, we can conclude that $(\overline{X}, \overline{Y})$ is feasible for [F_EC]$_{LP}$.

Proposition 2.

Z[F_EC]$_{LP}$ = Z[F_S]$_{LP}$.

6.2. Comparison of the LP relaxations of [F_R], [F_C], and [F_EC]

As pointed out by Veremyev and Boginski (2012), a theoretical comparison of the tightness of the LP relaxation of their formulation [F_R] with that of the LP relaxation of [F_C] is a very challenging task, due to the recursive nature of [F_R]. As we shall show next, the result of the comparison is not independent of the value of k: [F_R] dominates [F_C] for $k = 3$, but the dominance does not hold for $k = 4$.

We shall conclude this section by showing that [F_EC] dominates [F_R] for $k = 3$.

We shall first show that [F_R] dominates [F_C] for the case $k = 3$. Recalling the rationale presented in the proof of Proposition 2, it is quite intuitive that constraints (7) play an important role in any comparison of [F_R]$_{LP}$ and [F_C]$_{LP}$ for $k = 3$. Given a pair $\{i,j\} \in \mathcal{N}^3$ and a node $p \in A_{ij}$, constraint $v_{pj}^{(2)} \leqslant x_p$ may cut off feasible solutions of [F_C]$_{LP}$ in the x-space, in a way that bears a clear resemblance with the effect of constraints (22) in [F_EC].

Proposition 3. For $k = 3$, $Z[F\_R]_{LP} \leqslant Z[F\_C]_{LP}$.

Proof. Let $(\overline{X}, \overline{V})$ be feasible for [F_R]$_{LP}$. We shall assume, without loss of generality, that $\bar{v}_{ab}^{(2)} = \min\left\{\bar{x}_a, \bar{x}_b, \sum_{r\in(N_a\cap N_b)} \bar{x}_r\right\}$ and $\bar{v}_{ab}^{(3)} = \min\left\{\bar{x}_a, \sum_{k\in N_a} \bar{v}_{kb}^{(2)}\right\}$, for all $a, b \in V$.

Let us now build a feasible solution $(\overline{X}, \overline{Y})$ for [F_C]$_{LP}$.

All packing constraints are satisfied by $\overline{X}$ because $\bar{v}_{ij}^{(2)} = \bar{v}_{ij}^{(3)} = 0$ for all pairs $\{i,j\} \in \mathscr{P}^3$.

Let $\{i,j\} \in \mathcal{N}^3$ be chosen arbitrarily. If $N_i \cap N_j = \varnothing$, then $\bar{v}_{ij}^{(2)} = 0$ and $\sum_{r\in(N_i\cap N_j)} \bar{x}_r = 0$. Otherwise, let $\mathscr{D}_{ij}$ be a complete layered digraph with five layers:

$L_1 = \{i\}$, $L_2 = \{i'\}$, $L_3 = N_i \cap N_j$, $L_4 = \{j'\}$, and $L_5 = \{j\}$

(where i' and j' are copies of i and j, respectively) and arc capacities defined by

$$\text{cap}(u,v) = \begin{cases} \bar{x}_i & \text{if } u = i \quad \text{and} \quad v = i' \\ \bar{x}_r & \text{if } u = i' \quad \text{and} \quad v = r \in (N_i \cap N_j) \\ \infty & \text{if } u \in (N_i \cap N_j) \quad \text{and} \quad v = j' \\ \bar{x}_j & \text{if } u = j' \quad \text{and} \quad v = j \end{cases}$$

Suppose that f is a maximum $i - j$ flow in $\mathscr{D}_{ij}$. By construction, value$(f) = \sum_{r\in(N_i\cap N_j)} f_{i'r}$. If we set $\bar{y}_t = f_{i'r}$, for every chain $t = (i,r,j) \in C_{ij}^{-2}$, it follows that $\sum_{t\in C_{ij}^{-2}} \bar{y}_t = \bar{v}_{ij}^{(2)}$. We are now left with variables $\bar{v}_{ij}^{(3)}$ and chains $t \in C_{ij}^{-3}$, for all $\{i,j\} \in \mathcal{N}^3$. For every pair $\{i,j\} \in \mathcal{N}^3$, let us build a layered digraph $\mathscr{D}_{ij}$, with layers

$L_1 = \{i\}$, $L_2 = \{i'\}$, $L_3 = \{p : p \in N_i \text{ and } N_p \cap N_j \neq \varnothing\}$, $L_4 = \left\{ t : t \in C_{ij}^{=3} \right\}$, and $L_5 = \{j\}$ (where i' is a copy of i) and whose set of arcs is

$$\mathscr{A}_{ij} = \{(i, i')\} \cup \{(i', p) : p \in N_i \text{ and } N_p \cap N_j \neq \varnothing\} \cup$$
$$\cup \left\{ (p, t) : t \in C_{ij}^{=3} \text{ and edge } (i, p) \text{ belongs to chain } t \right\} \cup \left\{ (t, j) : t \in C_{ij}^{=3} \right\}$$

with arc capacities defined by

$$cap(u, v)$$
$$= \begin{cases} \bar{x}_i & \text{if } u = i & \text{and} & v = i' \\ \bar{v}_{pj}^{(2)} & \text{if } u = i' & \text{and} & v = p \in L_3 \\ \infty & \text{if } u \in L_3 & \text{and} & v = t \in L_4 \\ \min\{\bar{x}_w : w \text{ is a node of } t\} & \text{if } u = t \in L_4 & \text{and} & v = j \end{cases}.$$

Suppose that f is a maximum $i - j$ flow in $\mathscr{D}_{ij}$. By construction, $value(f) = \sum_{t \in C_{ij}^{=3}} f_{tj} = \bar{v}_{ij}^{(3)}$. By setting $\bar{y}_t = f_{tj}$, for all $t \in C_{ij}^{=3}$, it follows that $\sum_{t \in C_{ij}^{=3}} \bar{y}_t = value(f)$. We can now conclude that $\sum_{t \in C_{ij}^3} \bar{y}_t = \bar{v}_{ij}^{(2)} + \bar{v}_{ij}^{(3)}$ and consequently, since $\{i, j\} \in \mathscr{N}^3$ was chosen arbitrarily, that $(\overline{X}, \overline{Y})$ is feasible for $[F\_C]_{LP}$. $\quad\square$

The dominance of $[F\_R]$ over $[F\_C]$ for $k = 3$ may be strict. An illustration is provided by the graph depicted in Fig. 1(b).

The dominance does not hold for $k = 4$, as illustrated by the example presented in Appendix A.

Let us now consider $[F\_EC]$ and $[F\_R]$ for $k = 3$.

Let $\{i, j\} \in \mathscr{N}^3$ be chosen arbitrarily. In $[F\_EC]$ constraints (23) impose that $\sum_{p:(p,q) \in E_{ij}} y_{pq}^{ij} \leqslant x_q$ for every $q \in B_{ij}$. In $[F\_R]$ constraints (9) impose only that $v_{ij}^{(3)} \leqslant x_i$ and $v_{ij}^{(3)} \leqslant \sum_{k \in N_i} v_{kj}^{(2)}$. Due to constraints (23), feasible solutions for $[F\_R]_{LP}$ on the x-space, may be infeasible for $[F\_EC]_{LP}$. An illustration is provided by the graph depicted in Fig. 2(a).
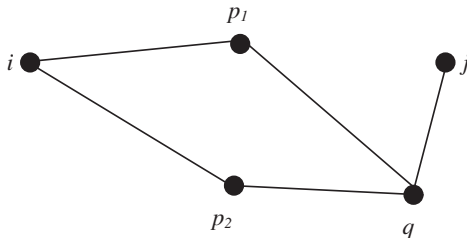
We shall next show that for $k = 3$ $[F\_EC]$ dominates $[F\_R]$.

Proposition 4. *For $k = 3$, $Z [F\_EC]_{LP} \leqslant Z [F\_R]_{LP}$.*

Proof. We shall show that the projection onto the x-space of the feasible set of $[F\_EC]_{LP}$ is included in the projection onto the x-space of the feasible set of $[F\_R]_{LP}$ by showing that if $(\overline{X}, \overline{Y})$ is feasible for $[F\_EC]_{LP}$, then $(\overline{X}, \overline{V})$ is feasible for $[F\_R]_{LP}$, where

$$\bar{v}_{ab}^{(2)} = \min \left\{ \bar{x}_a, \bar{x}_b, \sum_{r \in (N_a \cap N_b)} \bar{x}_r \right\} \text{ and } \bar{v}_{ab}^{(3)} = \min \left\{ \bar{x}_a, \sum_{k \in N_a} \bar{v}_{kb}^{(2)} \right\},$$

for all $a, b \in V$.

Fig. 2. (a) Consider a feasible solution for $[F\_R]_{LP}$ in which $\bar{x}_i = \frac{2}{3}$, $\bar{x}_j = 1$, $\bar{x}_{p_1} = \bar{x}_{p_2} = \bar{x}_q = \frac{1}{3}$, $\bar{v}_{ij}^{(2)} = 0$, $\bar{v}_{p_1,j}^{(2)} = \bar{v}_{p_2,j}^{(2)} = \frac{1}{3}$, and $\bar{v}_{ij}^{(3)} = \frac{2}{3}$. In the corresponding solution for $[F\_EC]_{LP}$, constraint (21) for the pair $\{i, j\}$ is violated because $y_{p_1,q}^{ij} + y_{p_2,q}^{ij} \leqslant x_q$ must hold. (b) Consider again $k = 3$ and a feasible solution for $[F\_R]_{LP}$ in which $\bar{x}_i = \frac{2}{3}$, $\bar{x}_j = 1$, $\bar{x}_{p_1} = \bar{x}_{p_2} = \bar{x}_q = \frac{1}{3}$, $\bar{v}_{ij}^{(2)} = 0$, $\bar{v}_{p_1 j}^{(2)} = \bar{v}_{p_2 j}^{(2)} = \frac{1}{3}$, and $\bar{v}_{ij}^{(3)} = \frac{2}{3}$. To comply with (9''), constraint (6) is violated.

We shall assume, without loss of generality, that for all pairs $\{i, j\} \in \mathscr{N}^3$, $\sum_{(p,q) \in E_{ij}} \bar{y}_{pq}^{ij} \leqslant \bar{x}_i$ and $\sum_{(p,q) \in E_{ij}} \bar{y}_{pq}^{ij} \leqslant \bar{x}_j$. Note that, to satisfy constraints (7) and (9), $v_{ij}^{(2)}$ and $v_{ij}^{(3)}$ cannot assume values greater than the ones they are being assigned, which makes constraints (8) and (10) irrelevant.

If $\{i, j\} \in \mathscr{P}^3$, then $\bar{v}_{ij}^{(2)} = \bar{v}_{ij}^{(3)} = 0$, and consequently, $\bar{x}_i + \bar{x}_j \leqslant 1 + \bar{v}_{ij}^{(2)} + \bar{v}_{ij}^{(3)}$ if and only if $\bar{x}_i + \bar{x}_j \leqslant 1$.

Let $\{i, j\} \in \mathscr{N}^3$ be chosen arbitrarily.

If $\bar{v}_{ij}^{(2)} = \bar{x}_i$ or $\bar{v}_{ij}^{(2)} = \bar{x}_j$, constraint (6) is satisfied, regardless of the value of $v_{ij}^{(3)}$. Otherwise, $\bar{v}_{ij}^{(2)} = \sum_{r \in (N_i \cap N_j)} \bar{x}_r$.

If $\bar{v}_{ij}^{(3)} = \bar{x}_i$, constraint (6) is satisfied, regardless of the value of $v_{ij}^{(2)}$. Otherwise, $\sum_{k \in N_i} \bar{v}_{kj}^{(2)} < \bar{x}_i$ and $\bar{v}_{ij}^{(3)} = \sum_{k \in N_i} \bar{v}_{kj}^{(2)}$.

Let us assume that $\bar{v}_{ij}^{(3)} = \sum_{k \in N_i} \bar{v}_{kj}^{(2)}$.

By construction, $\sum_{(p,q) \in E_{ij}} \bar{y}_{pq}^{ij} = \sum_{p \in A_{ij}} \left(\sum_{q:(p,q) \in E_{ij}} \bar{y}_{pq}^{ij} \right)$.

Let p^* be chosen arbitrarily in A_{ij} and remember we are assuming, without loss of generality, that $\sum_{(p,q) \in E_{ij}} \bar{y}_{pq}^{ij} \leqslant \bar{x}_j$.

Since $\sum_{q:(p^*,q) \in E_{ij}} \bar{y}_{p^*q}^{ij} \leqslant \sum_{(p,q) \in E_{ij}} \bar{y}_{pq}^{ij}$, we can conclude that $\sum_{q:(p^*,q) \in E_{ij}} \bar{y}_{p^*q}^{ij} \leqslant \bar{x}_j$.

By (22), $\sum_{q:(p^*,q) \in E_{ij}} \bar{y}_{p^*q}^{ij} \leqslant \bar{x}_{p^*}$, and by (19), $\sum_{q:(p^*,q) \in E_{ij}} \bar{y}_{p^*q}^{ij} \leqslant \sum_{q:(p^*,q) \in E_{ij}} \bar{x}_q = \sum_{q \in N_{p^*} \atop q \in N_j \backslash N_i} \bar{x}_q \leqslant \sum_{q \in (N_{p^*} \cap N_j)} \bar{x}_q$.

It follows that $\sum_{q:(p^*,q) \in E_{ij}} \bar{y}_{p^*q}^{ij} \leqslant \min \left\{ \bar{x}_{p^*}, \ \bar{x}_j, \ \sum_{r \in (N_{p^*} \cap N_j)} \bar{x}_r \right\} = \bar{v}_{p^*j}^{(2)}$ and, therefore $\sum_{(p,q) \in E_{ij}} \bar{y}_{pq}^{ij} \leqslant \min \left\{ \bar{x}_i, \ \sum_{k \in N_i} \bar{v}_{kj}^{(2)} \right\} = \bar{v}_{ij}^{(3)}$.

Since $\{i, j\} \in \mathscr{N}^3$ was chosen arbitrarily, we can conclude that if $(\overline{X}, \overline{V})$ is feasible for $[F\_EC]_{LP}$, then $(\overline{X}, \overline{Y})$ is feasible for $[F\_R]_{LP}$. $\quad\square$

6.2.1. Enhanced version of [F_R]

The study of the relations between $[F\_EC]_{LP}$ and $[F\_R]_{LP}$ for $k = 3$ gives hints on how to strengthen the recursive formulation from the LP relaxation standpoint for any integer $k \geqslant 3$. Note that for $k = 2$, $[F\_C]$ and $[F\_R]$ are essentially the same (Veremyev & Boginski, 2012).

For $k = 3$, the dominance of $[F\_EC]_{LP}$ over $[F\_R]_{LP}$ stems from two reasons: the elimination of all variables that represent chains $t = (i, p, q, j) \in C_{ij}^{=3}$ with $p \in (N_i \cap N_j)$ or $q \in (N_i \cap N_j)$ in $[F\_EC]$ and the lifting effect provided by constraints (23). For any $k \geqslant 3$, a similar elimination is obtained by replacing (9) with

$$v_{ij}^{(l)} \leqslant x_i \quad v_{ij}^{(l)} \leqslant \sum_{k \in (N_i \backslash N_j)} v_{kj}^{(l-1)} \tag{9'}$$

Due to their recursive nature, variables $v_{ij}^{(l)}$ represent paths directed from node i to node j. Given any undirected graph, there is at least one path from i to j if and only if there is at least one path from j to i. Therefore, for any $k \geqslant 3$, we can add to $[F\_R]$ the following conditions

for $l = 3, \ldots, k$; $i, j \in V, j > i$,

$$v_{ij}^{(l)} \leqslant x_j \quad v_{ij}^{(l)} \leqslant \sum_{k \in (N_j \backslash N_i)} v_{ki}^{(l-1)} \tag{9''}$$

$$v_{ij}^{(l)} \geqslant \frac{1}{|V|} \left(\sum_{k \in N_j} v_{ki}^{(l-1)} \right) + (x_j - 1) \tag{10'}$$

The resulting enhanced recursive formulation for the k-club problem is

$$[F\_ER] \quad \max \left\{ \sum_{i \in V} x_i : (6) - (8), (9'), (9''), (10), (10'), (4), (11) \right\}.$$

By construction, [F_ER] dominates [F_R].

Proposition 5. $Z[F\_ER]_{LP} \leqslant Z[F\_R]_{LP}$.

The dominance of [F_ER] over [F_R] may be strict. An illustration is provided by the graph depicted in Fig. 2(b).

It is quite intuitive that if $k = 3$, [F_EC] and [F_ER] are LP-equivalent.

The proof of Proposition 6 follows the rationale used in the proof of Proposition 4. A detailed proof is presented in Appendix B.

Proposition 6. *For $k = 3$, $Z[F\_EC]_{LP} = Z[F\_ER]_{LP}$.*

6.3. Comparison of the LP relaxations of [F_N] and [F_R]

For $k = 3$, formulation [F_C] is dominated both by [F_N] and [F_R], but the dominances stem from different reasons. The dominance of [F_N] over [F_C] is due to the substitution of E_{ij} for $E(N_i, N_j)$ in the last term of inequalities (13) (Almeida & Carvalho, 2012). The dominance of [F_R] over [F_C] is, roughly speaking, a consequence of (7). It is then natural to suspect that in the x-space the feasible sets of [F_N]_{LP} and [F_R]_{LP} are different and no inclusion relation holds for them.

The graph used to illustrate the dominance of [F_R] over [F_C] (see Fig. 1(c)) can also be used to show that in the x-space a feasible solution for [F_N]_{LP} may not be feasible for [F_R]_{LP}.

On the other hand, in the x-space a feasible solution for [F_R]_{LP} may not be feasible for [F_N]_{LP}. An illustration is provided by the graph depicted in Fig. 3.

Since there is no inclusion relation in the x-space between the feasible sets of [F_R]_{LP} and [F_N]_{LP}, there is no domination relation between [F_R] and [F_N].

6.4. Neighbourhood formulation revisited

For the case $k = 3$, neighbourhood formulation [F_N] is more compact than any other formulation in the literature. Since, also for $k = 3$, [F_EC] and [F_S] are LP-equivalent, and [F_S] dominates [F_N], we can conclude that [F_N] is also dominated by [F_EC]. Note that there is no way of imposing constraints on the set of three-edge chains that link each pair of nodes $\{i, j\} \in \mathcal{N}^3$ using only node and edge variables. However, it is possible to derive [F_EC] from [F_N] by introducing chain variables and adding valid inequalities for the 3-club polytope.

Let us consider again variables y_{pq}^{ij} (see Section 5), and let us add constraints (18), (20), (22), and (23) to [F_N]. After adding these constraints, constraints (13) can be replaced with

$$x_i + x_j \leqslant 1 + \sum_{r \in (N_i \cap N_j)} x_r + \sum_{(p,q) \in E_{ij}} y_{pq}^{ij} \quad \{i,j\} \in \mathcal{N}^3 \qquad (21)$$

Now, edge variables z_{ij} play no role in the formulation and can be dropped, together with constraints (14). To summarize, (13) and (14) in [F_N] can be replaced with 18 and (20)–(23). Thus, by introducing chain variables y_{pq}^{ij} and lifting variable linking
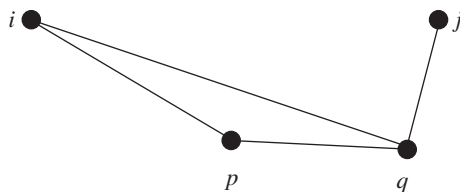
Fig. 3. Consider a solution in the x-space such that $\bar{x}_i = \frac{2}{3}$, $\bar{x}_j = 1$, $\bar{x}_p = \bar{x}_q = \frac{1}{3}$. By setting $\bar{v}_{ij}^{(2)} = \bar{v}_{pj}^{(2)} = \bar{v}_{ij}^{(3)} = \frac{1}{3}$ we get a feasible solution for [F_R]_{LP}. Since $\sum_{r \in (N_i \cap N_j)} \bar{x}_r = \frac{1}{3}$ and $E_{ij} = \varnothing$, constraint (13) in [F_N]_{LP} for the pair $\{i,j\}$ is violated.

constraints, [F_N] is transformed into [F_EC]. This procedure provides an alternative proof of the dominance of [F_S] over [F_N] presented in Almeida and Carvalho (2012).

The large number of variables added to [F_N] to transform it into [F_EC] raises the question of the magnitude of the difference $Z[F\_N]_{LP} - Z[F\_EC]_{LP}$. An upper bound on this difference is easy to derive by considering a variant of formulation [F_S], denoted by [F_S$_\gamma$], which results from replacing constraints (16) with

$$x_i + x_j \leqslant 1 + \sum_{r \in (N_i \cap N_j)} x_r + \sum_{s \in S_{ij}} (\gamma_s - 1) x_s \quad \{i,j\} \in \mathcal{N}^3, S_{ij} \in \mathcal{S}^{ij} \qquad (24)$$

where γ_s is the degree of node s in subgraph G_{ij}.

Since $\sum_{s \in S_{ij}} (\gamma_s - 1) x_s \geqslant 1$ if and only if $\sum_{s \in S_{ij}} x_s \geqslant 1$, formulation [F_S$_\gamma$] is valid for the 3-club problem. Since $\gamma_s \geqslant 2$ for all $s \in S_{ij}$ and $S_{ij} \in \mathcal{S}^{ij}$, $Z[F\_S]_{LP} \leqslant Z[F\_S_\gamma]_{LP}$.

We shall next show that [F_N] dominates [F_S$_\gamma$].

Proposition 7. $Z[F\_N]_{LP} \leqslant Z[F\_S_\gamma]_{LP}$.

Proof. Let $(\bar{X}, \bar{Z})$ be feasible for [F_N]_{LP}. For all $\{i,j\} \in \mathcal{N}^3$ and $S_{ij} \in \mathcal{S}^{ij}$,

$$\sum_{\substack{(p,q) \in E_{ij}}} \bar{z}_{pq} = \sum_{\substack{(p,q) \in E_{ij} \\ p \in (A_{ij} \cap S_{ij}) \\ q \in B_{ij}}} \bar{z}_{pq} + \sum_{\substack{(p,q) \in E_{ij} \\ p \in (A_{ij} \backslash S_{ij}) \\ q \in (B_{ij} \cap S_{ij})}} \bar{z}_{pq}$$

$$\leqslant \sum_{s \in (A_{ij} \cap S_{ij})} (\gamma_s - 1) \bar{x}_s + \sum_{s \in (B_{ij} \cap S_{ij})} (\gamma_s - 1) \bar{x}_s = \sum_{s \in S_{ij}} (\gamma_s - 1) \bar{x}_s$$

and $\bar{X}$ is feasible for [F_S$_\gamma$]_{LP}. ☐

The dominance of [F_N] over [F_S$_\gamma$] may be strict. An illustration is provided by the graph depicted in Fig. 4.

The graphs depicted in Figs. 3 and 4 provide an illustration that there is no dominance relation between [F_S$_\gamma$] and [F_C].

7. *k*-Clubs with additional constraints

A k-club is by definition a subset of nodes of a graph which induces a subgraph with diameter at most equal to k. If k is small, a k-club represents a cluster with good connectivity between each pair of its members– every member needs at most $k - 1$ intermediaries to reach any other member. If the underlying graph represents a communication network, the diameter has been considered a natural way of describing the reliability of the network (Boesch, Harary, & Kabell, 1981). However, a k-club is a fragile structure in that if one of its elements breaks down by accident, the communications among members may be severely affected. A k-club is also quite vulnerable to external attacks: the destruction of a selected single element (node or link) may render the communication among members totally impossible. To identify network clusters that combine good connectivity among members with better
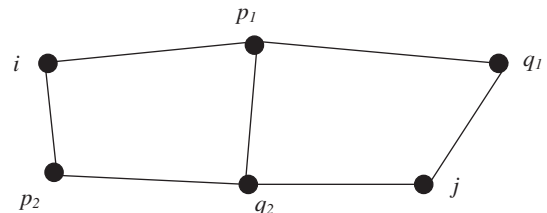
Fig. 4. Consider a solution in the x-space in which $\bar{x}_i = \frac{5}{8}$, $\bar{x}_j = 1$, $\bar{x}_{p_1} = \bar{x}_{q_2} = \frac{1}{4}$, and $\bar{x}_{p_2} = \bar{x}_{q_1} = \frac{1}{8}$. In [F_S$_\gamma$]_{LP}, $S^{ij} = \{A_{ij}, \{p_1, q_2\}, B_{ij}\}$, thus node cut set constraints (24) are satisfied for all $S_{ij} \in \mathcal{S}^{ij}$, but in [F_N]_{LP}, to comply with (14), neighbourhood constraint (13) for $\{i,j\}$ is violated.

reliability properties, Veremyev and Boginski (2012) introduced the new concept of *R*-robust *k*-club, which extends the original *k*-club definition by imposing the additional condition that there must be at least *R* internally node-disjoint paths between every pair of cluster members. A key feature of this new concept is that the elimination of up to $R - 1$ elements will not destroy the *k*-club structure, which means that the *R*-robust *k*-clubs have much better error and attack tolerance characteristics than *k*-clubs (Veremyev & Boginski, 2012). As pointed out by the authors, developing mathematical programming approaches for finding large network clusters with good error and attack tolerance characteristics is not an easy task.

For the case *k* = 2, all distinct paths that link a given pair of nodes *i* and *j* are internally node-disjoint, since each path is either edge (*i*,*j*) or a two-edge path of the form (*i*,*r*,*j*) with node *r* in $N_i \cap N_j$. The *R*-robust 2-club problem is formulated in Veremyev and Boginski (2012) as follows:

$$\max \quad Z = \sum_{i \in V} x_i$$

$$\text{s.t.} \quad a_{ij} + \sum_{k \in (N_i \cap N_j)} x_k \geqslant R(x_i + x_j - 1) \quad i,j \in V, i < j \tag{25}$$

$$x_i \in \{0, 1\} \quad i \in V \tag{4}$$

where $a_{ij} = 1$ if $(i,j) \in E$ and $a_{ij} = 0$ otherwise.

For *k* > 2, to guarantee that there are at least *R* internally node-disjoint paths between any pair of nodes it is necessary to identify the inner nodes of each path with more than two edges. In Veremyev and Boginski (2012), the $w_{ij}^{(l)}$ variables that represent the number of distinct paths with *l* edges ($l = 2,\ldots,k$) that link each pair of nodes *i* and *j* ($i,j = 1,\ldots,|V|$) are defined by recursion, and by its very nature a definition by recursion is incompatible with the identification of inner nodes path by path. For that reason, for *k* > 2 Veremyev and Boginski (2012) considered a relaxation of the *R*-robust *k*-club concept, obtained by replacing the original condition "*R* internally node-disjoint paths" by the relaxed condition "*R* distinct paths".

The relaxed *R*-robust *k*-club problem is formulated in Veremyev and Boginski (2012) with binary node variables x_i ($i = 1,\ldots,|V|$) and integer path variables $w_{ij}^{(l)}$ ($i,j = 1,\ldots,|V|; l = 2,\ldots,k$) as follows:

$$\max \quad Z = \sum_{i \in V} x_i$$

$$\text{s.t.} \quad a_{ij} + \sum_{l=2}^{k} w_{ij}^{(l)} \geqslant R(x_i + x_j - 1) \tag{26}$$

$$w_{ij}^{(2)} \leqslant \sum_{k \in (N_i \cap N_j)} x_k + |V|(2 - x_i - x_j) \tag{27}$$

$$w_{ij}^{(2)} \geqslant \sum_{k \in (N_i \cap N_j)} x_k - |V|(2 - x_i - x_j) \tag{28}$$

$$w_{ij}^{(2)} \leqslant |V|x_i \quad w_{ij}^{(2)} \geqslant -|V|x_i \tag{29}$$

$$w_{ij}^{(2)} \leqslant |V|x_j \quad w_{ij}^{(2)} \geqslant -|V|x_j \text{ and for } l = 3,\ldots,k, \tag{30}$$

$$w_{ij}^{(l)} \leqslant \sum_{k \in N_i} w_{kj}^{(l-1)} + |V|^{l-1}(1 - x_i) \tag{31}$$

$$w_{ij}^{(l)} \geqslant \sum_{k \in N_i} w_{kj}^{(l-1)} - |V|^{l-1}(1 - x_i) \tag{32}$$

$$w_{ij}^{(l)} \leqslant |V|^{l-1}x_i \quad w_{ij}^{(l)} \geqslant -|V|^{l-1}x_i \tag{33}$$

$$x_i \in \{0, 1\} \tag{4}$$

$$w_{ij}^{(l)} \in Z^+ \tag{34}$$

where $a_{ij} = 1$ if $(i,j) \in E$, $a_{ij} = 0$ otherwise, and $i, j = 1, \ldots, |V|$.

Constraints (26) impose that each pair of selected nodes must be linked by at least *R* paths with at most *k* edges. Constraints (27)–(30) define the number of two-edge paths that link each pair of nodes, and constraints (31)–(33) define the number of paths with three or more edges that link each pair of nodes. Constraints (34) define the path variables as nonnegative and integer.

While for *k* = 2 the conditions "*R* internally node-disjoint paths" and "*R* distinct paths" are equivalent, for *k* > 2 the latter is considerably weaker than the former, since distinct paths may share even all inner nodes. This means that the error and attack tolerance properties of *R*-robust *k*-clubs are not guaranteed by conditions (26)–(34) and (4). For *k* = 3, the error and attack tolerance properties of *R*-robust *k*-clubs can be guaranteed by adapting formulation [F_EC], as described next.

Given a pair of nodes *i* and *j* not linked by an edge in *G*, constraints (22) and (23) impose that at most one chain that includes node *p* and at most one chain that includes node *q* can be selected to link them in the induced subgraph. Furthermore, in formulation [F_EC], variables y_{pq}^{ij} are defined only for $(p,q) \in E$ such that $p \in (N_i \setminus N_j)$ and $q \in (N_j \setminus N_i)$. Thus, chains represented by variables y_{pq}^{ij} are internally node-disjoint with two-edge chains that link *i* and *j* (which are represented by their central nodes). In short, all that is missing to model *R*-robust 3-clubs is a set of variables to identify each chain with three edges that links pairs of nodes *i* and *j*, linked by an edge in *G*.

For each pair of nodes *i* and *j* adjacent in *G*, let us define variables y_{pq}^{ij}, with $p \in N_i \setminus (N_j \cup \{j\})$ and $q \in N_j \setminus (N_i \cup \{i\})$. Each variable is associated with one edge in $E_{ij}^1 = \{(p,q) \in E : p \in N_i \setminus (N_j \cup \{j\}), q \in N_j \setminus (N_i \cup \{i\})\}$.

To obtain a valid formulation for the *R*-robust 3-club problem, we substitute constraints

$$\sum_{r \in (N_i \cap N_j)} x_r + \sum_{(p,q) \in E_{ij}} y_{pq}^{ij} \geqslant R(x_i + x_j - 1) \quad \{i,j\} \in \mathscr{N}^3 \tag{37}$$

for constraints (21) and enlarge formulation [F_EC] with the new constraints

$$\sum_{r \in (N_i \cap N_j)} x_r + \sum_{(p,q) \in E_{ij}^1} y_{pq}^{ij} \geqslant (R-1)(x_i + x_j - 1) \quad (i,j) \in E \tag{38}$$

$$y_{pq}^{ij} \leqslant x_i \quad y_{pq}^{ij} \leqslant x_j \quad (i,j) \in E \quad (p,q) \in E_{ij}^1 \tag{39}$$

$$\sum_{q:(p,q) \in E_{ij}^1} y_{pq}^{ij} \leqslant x_p \quad (i,j) \in E \quad p \in N_i \setminus (N_j \cup \{j\}) \tag{40}$$

$$\sum_{p:(p,q) \in E_{ij}^1} y_{pq}^{ij} \leqslant x_q \quad (i,j) \in E \quad q \in N_j \setminus (N_i \cup \{i\}) \tag{41}$$

$$y_{pq}^{ij} \in \{0, 1\} \quad (i,j) \in E \quad (p,q) \in E_{ij}^1 \tag{42}$$

Conditions (37) and (38) impose that each pair of nodes must be linked by at least *R* selected chains with at most three edges. Conditions (39)–(41) impose that all nodes in selected chains that link adjacent nodes are included in the solution and that these chains are internally node-disjoint.

To summarize, an optimal solution for

$$[\text{F\_EC} - R] \quad \max \left\{ \sum_{i \in V} x_i : (12), (18), (22), (23), (4), (20), (37) - (42) \right\}$$

is a maximum 3-club that satisfies the additional condition that every pair of its nodes is linked by at least *R* internally node-disjoint chains.

8. Conclusions

In this paper, we have presented a comparative study of the LP relaxations of integer formulations for the *k*-club problem. The
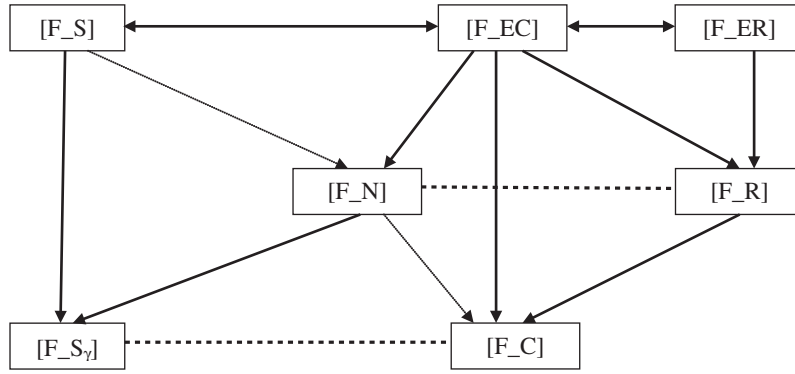
Fig. 5. Connections among formulations for the 3-club problem.

comparisons involved models known from the literature and three new models. Some models are valid for any $k > 1$ whereas some other were designed for the case $k = 3$.

For the case $k = 3$, the connections among all these formulations are synthesised in Fig. 5. For generic formulations [F_∗] and [F_•], [F_∗] → [F_•] means that the LP relaxation of [F_∗] is stronger than the LP relaxation of [F_•], and [F_∗] ↔ [F_•] means that [F_∗] and [F_•] are LP-equivalent. Thick arrows represent the relations derived in this study and thin arrows represent relations derived in Almeida and Carvalho (2012). A dashed line between [F_∗] and [F_•] means that there is no relation between their LP relaxations.

For $k = 4$, we showed that [F_R]$_{LP}$ is not stronger than [F_C]$_{LP}$.

Based on the insight obtained by comparing [F_EC] and [F_R] for $k = 3$, we devised [F_ER] for the k-club problem and showed that it dominates [F_R], for any $k > 1$. In addition, we have derived a formulation for the maximum R-robust 3-club problem. To the best of our knowledge, this is the first formulation in the literature for the maximum R-robust 3-club problem.

Acknowledgements

Appendix A

Consider graph $G = (V, E)$ depicted in Fig. A.1 and $k = 4$.

Suppose that $\bar{x}_i = \bar{x}_j = 1$, $\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = \frac{1}{5}$, and $\bar{x}_4 = \frac{1}{20}$.

In [F_C]$_{LP}$ there are six variables y_t, associated with chains $(i,1,j)$, $(i,1,2,j)$, $(i,3,2,j)$, $(i,3,2,4,j)$, $(i,1,2,4,j)$, and $(i,3,2,1,j)$. Consider any $(\bar{X}, \bar{Y})$. To be a solution for [F_C]$_{LP}$, it must comply
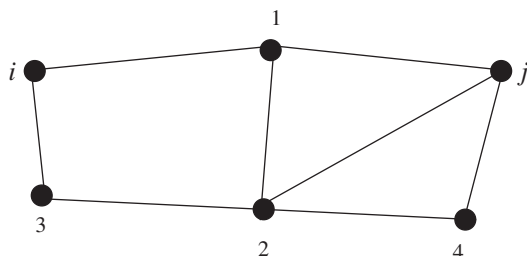

Fig. A.1. Graph to illustrate that [F$_R$] does not dominate [F$_C$] for k=4.

with $\sum_{t=1}^{6} y_t \leqslant \frac{9}{10}$, but then constraint (2) for the pair $\{i,j\}$, is violated.

Consider now $(\bar{X}, \bar{V})$, with,

$$\bar{v}_{ab}^{(2)} = \min \left\{ \bar{x}_a, \quad \bar{x}_b, \quad \sum_{k \in (N_a \cap N_b)} \bar{x}_k \right\} \text{ for all } a, b \in V$$

$$\bar{v}_{ab}^{(l)} = \min \left\{ \bar{x}_a, \quad \sum_{k \in N_a} \bar{v}_{kb}^{(l-1)} \right\} \text{ for all } a, b \in V, l = 3, 4$$

Then, $\sum_{l=2}^{4} \bar{v}_{ij}^{(l)} = \frac{1}{5} + \frac{2}{5} + \frac{2}{5} = 1$, which satisfies constraint (6) in [F_R]$_{LP}$, for the pair $\{i,j\}$.

Appendix B. proof of Proposition 6

Proposition 6. *For k = 3, Z[F_EC]$_{LP}$ = Z[F_ER]$_{LP}$.*

Proof. Suppose that $(\bar{X}, \bar{V})$ is feasible for [F_ER]$_{LP}$. Let us show that it is possible to build from it a feasible solution $(\bar{X}, \bar{Y})$ for [F_EC]$_{LP}$. We shall assume, without loss of generality, that

$$\bar{v}_{ab}^{(2)} = \min \left\{ \bar{x}_a, \bar{x}_b, \sum_{k \in (N_a \cap N_b)} \bar{x}_k \right\} \text{ for all } a, b \in V$$

$$\bar{v}_{ab}^{(3)} = \min \left\{ \bar{x}_a, \bar{x}_b, \sum_{k \in (N_a \setminus N_b)} \bar{v}_{kb}^{(2)}, \sum_{k \in (N_b \setminus N_a)} \bar{v}_{ka}^{(2)} \right\} \text{ for all } a, b \in V$$

For packing constraints and for chains with two edges, the rational of the proof is the same rationale used in Section 6.2.

Let us now address chains with three edges that link nodes i and j. Let $\widetilde{C}_{ij}^{=3} = \left\{ t = (i, p, q, j) \in C_{ij}^{=3} : p \notin N_j, q \notin N_i \right\}$ be the set of chains associated with variables $y_{pq}^{ij} (\{i,j\} \in \mathcal{N}^3, p \in A_{ij}, q \in B_{ij})$ in [F_EC]$_{LP}$.

Consider a pair $\{i,j\} \in \mathcal{N}^3$, chosen arbitrarily. Let $\mathcal{D}_{ij}$ be a layered digraph with seven layers:

$$L_1 = \{i\}, \ L_2 = \{i'\}, \ L_3 = A_{ij}, \ L_4 = \left\{ t : t \in \widetilde{C}_{ij}^{=3} \right\}, \ L_5 = B_{ij},$$
$$L_6 = \{j'\}, \text{ and } L_7 = \{j\}$$

(where i' and j' are copies of i and j, respectively) whose set of arcs is

$$\mathcal{A}_{ij} = \{(i, i')\} \cup \{(i', p) : p \in A_{ij}\} \cup$$
$$\cup \left\{ (p, t) : p \in A_{ij}, t \in \widetilde{C}_{ij}^{=3}, \text{ and } p \text{ is a node of } t \right\} \cup$$
$$\cup \left\{ (t, q) : q \in B_{ij}, t \in \widetilde{C}_{ij}^{=3}, \text{ and } q \text{ is a node of } t \right\} \cup$$
$$\cup \{(q, j') : q \in B_{ij}\} \cup \{(j', j)\}$$

and arc capacities are defined by

$$
\text{cap}\,(u,v) = \begin{cases}
\bar{x}_i & \text{if} \quad u=i & \text{and} \quad v=i' \\
\bar{v}^{(2)}_{pj} & \text{if} \quad u=i' & \text{and} \quad v=p \in A_{ij} \\
\bar{x}_p & \text{if} \quad u=p \in A_{ij} & \text{and} \quad v=t \in \widetilde{C}^{=3}_{ij} \text{ and } p \text{ is a node of } t \\
\bar{x}_q & \text{if} \quad u=t \in \widetilde{C}^{=3}_{ij} & \text{and} \quad v=q \in B_{ij} \text{ and } q \text{ is a node of } t \\
\bar{v}^{(2)}_{qi} & \text{if} \quad u=q \in B_{ij} & \text{and} \quad v=j' \\
\bar{x}_j & \text{if} \quad u=j' & \text{and} \quad v=j
\end{cases}.
$$

Note that each edge $(p,q) \in E_{ij}$ is represented in $\mathcal{D}_{ij}$ by exactly one node $t \in L_4$. This node will be denoted by $\varphi(p,q)$.

Suppose that f is a maximum $i-j$ flow in $\mathcal{D}_{ij}$.

By construction, $\text{value}\,(f) = \sum_{\substack{p \in A_{ij} \\ t \in C^{=3}_{ij}}} f_{pt} = \sum_{\substack{q \in B_{ij} \\ t \in C^{=3}_{ij}}} f_{tq}$. Also by construction,

$$
\sum_{\substack{u \in L3 \\ v \in L4}} \text{cap}\,(u,v) \geqslant \sum_{p \in A_{ij}} \bar{x}_p \geqslant \sum_{p \in A_{ij}} \min\left\{ \bar{x}_p, \bar{x}_j, \sum_{k \in B_{ij}} \bar{x}_k \right\} = \sum_{p \in A_{ij}} \bar{v}^{(2)}_{pj}
$$

and

$$
\sum_{\substack{u \in L4 \\ v \in L5}} \text{cap}\,(u,v) \geqslant \sum_{q \in B_{ij}} \bar{x}_q \geqslant \sum_{q \in B_{ij}} \min\left\{ \bar{x}_q, \bar{x}_i, \sum_{k \in A_{ij}} \bar{x}_k \right\} = \sum_{q \in B_{ij}} \bar{v}^{(2)}_{qi}.
$$

Therefore, $\text{value}\,(f) = \min\left\{ \bar{x}_i, \quad \sum_{p \in A_{ij}} \bar{v}^{(2)}_{pj}, \quad \sum_{q \in B_{ij}} \bar{v}^{(2)}_{qi}, \bar{x}_j \right\} = \bar{v}^{(3)}_{ij}$.

If we set $\bar{y}^{ij}_{pq} = f_{p,\varphi(p,q)}$ for all $(p,q) \in E_{ij}$, then $\text{value}\,(f) = \sum_{(p,q) \in E_{ij}} \bar{y}^{ij}_{pq}$.

For the pair $\{i,j\}$, all constraints (22) and (23) are satisfied, due to node flow conservation in L_3 and L_5. Since constraints (21) for $\{i,j\}$ are also satisfied, and the pair $\{i,j\} \in \mathcal{N}^3$ was chosen arbitrarily, solution $(\overline{X},\overline{Y})$ is feasible for [F_EC]$_{LP}$.

Conversely, suppose that $(\overline{X},\overline{Y})$ is feasible for [F_EC]$_{LP}$. By repeating the arguments used in the proof of Proposition 4, we conclude that $(\overline{X},\overline{V})$ is feasible for [F_ER]$_{LP}$, where

$$
\bar{v}^{(2)}_{ab} = \min\left\{ \bar{x}_a, \bar{x}_b, \sum_{k \in (N_a \cap N_b)} \bar{x}_k \right\} \text{ for all } a,b \in V
$$

$$
\bar{v}^{(3)}_{ab} = \min\left\{ \bar{x}_a, \bar{x}_b, \sum_{k \in (N_a \setminus N_b)} \bar{v}^{(2)}_{kb}, \sum_{k \in (N_b \setminus N_a)} \bar{v}^{(2)}_{ka} \right\} \text{ for all } a,b \in V \qquad \square
$$

References

Alba, R. D. (1973). A graph-theoretic definition of a sociometric clique. *Journal of Mathematical Sociology, 3*, 113–126.

Alidaee, B., Glover, F., Kochenberger, G., & Wang, H. (2007). Solving the maximum weight clique problem via unconstrained quadratic programming. *European Journal of Operational Research, 181*, 592–597.

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

Balas, E. (2005). Projection, lifting and extended formulation in integer and combinatorial optimization. *Annals of Operations Research, 140*, 125–161.

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

Błażewicz, J., Formanowicz, P., & Kasprak, M. (2005). Selected combinatorial problems of computational biology. *European Journal of Operational Research, 161*, 585–597.

Boesch, F. T., Harary, F., & Kabell, J. A. (1981). Graphs as models of communication network vulnerability: Connectivity and persistence. *Networks, 11*, 57–63.

Boginski, V., Butenko, S., & Pardalos, P. M. (2006). Mining market data: A network approach. *Computers & Operations Research, 33*, 3171–3184.

Bomze, I. M., Budinich, M., Pardalos, P. M., & Pelillo, M. (1999). The maximum clique problem. In D.-Z. Du & P. M. Pardalos (Eds.), *Handbook of combinatorial optimization* (pp. 1–74). Dordrecht, The Netherlands: Kluwer Academic Publishers.

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*, 21–28.

Butenko, S., & Wilhelm, W. E. (2006). Clique detection models in computational biochemistry and genomics. *European Journal of Operational Research, 173*, 1–7.

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

Gouveia, L., & Voss, S. (1995). A classification of formulations for the (time-dependent) traveling salesman problem. *European Journal of Operational Research, 83*, 69–82.

Magnanti, T. L., & Wolsey, L. A. (1995). Optimal trees. In M. O. Ball, T. L. Magnanti, C. L. Monma, & G. L. Nemhauser (Eds.). *Handbooks in operations research and management science* (Vol. 7, pp. 503–615). Elsevier.

Mahdavi Pajouh, F., & Balasundaram, B. (2012). On inclusionwise maximal and maximum cardinality *k*-clubs in graphs. *Discrete Optimization, 9*, 84–97.

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

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

Nemhauser, G. L., & Wolsey, L. A. (1988). *Integer and combinatorial optimization*. New York: John Wiley.

Oncan, T., Altinel, I. K., & Laporte, G. (2009). A comparative analysis of several asymmetric traveling salesman problem formulations. *Computers & Operations Research, 36*, 637–654.

Sorensen, M. (2004). New facets and branch and cut algorithm for the weighted clique problem. *European Journal of Operational Research, 154*, 57–70.

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