European Journal of Operational Research 226 (2013) 9–18

Contents lists available at SciVerse ScienceDirect

European Journal of Operational Research

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

Discrete Optimization

On clique relaxation models in network analysis

Jeffrey Pattillo [a], Nataly Youssef [b], Sergiy Butenko [c,*]

[a] Department of Mathematics, Texas A&M University, College Station, TX 77843-3131, United States
[b] Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, United States
[c] Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX 77843-3131, United States

ARTICLE INFO

Article history:
Received 9 August 2012
Accepted 19 October 2012
Available online 7 November 2012

Keywords:
Clique relaxations
Maximum clique problem
Social network analysis
Cohesive subgroups
Biological networks

ABSTRACT

Increasing interest in studying *community structures*, or *clusters* in complex networks arising in various applications has led to a large and diverse body of literature introducing numerous graph-theoretic models relaxing certain characteristics of the classical clique concept. This paper analyzes the elementary clique-defining properties implicitly exploited in the available *clique relaxation* models and proposes a taxonomic framework that not only allows to classify the existing models in a systematic fashion, but also yields new clique relaxations of potential practical interest. Some basic structural properties of several of the considered models are identified that may facilitate the choice of methods for solving the corresponding optimization problems. In addition, bounds describing the cohesiveness properties of different clique relaxation structures are established, and practical implications of choosing one model over another are discussed.

1. Introduction

Initially proposed by Luce and Perry (1949) as a model of a *cohesive subgroup* (*cluster*) within the context of social network analysis, a *clique* refers to a "tightly knit" set of elements (referred to as "actors" and described by *vertices* in graph-theoretic representation of a network), in which every pair of actors shares some common attribute. In other words, all elements of a clique are "directly connected" to each other. This allows for perfect familiarity and reachability between members of a clique. Moreover, removal of any element of a clique results in a slightly smaller clique and does not impact the perfectly-tied structure of the group, making cliques ideal in terms of robustness as well. Thus, the clique model possesses idealized cohesiveness properties within a group of actors it describes. However, requiring all possible links to exist may prove to be rather restrictive for many applications, where interaction between members of the group needs not be direct and could be successfully accomplished through intermediaries.

To overcome the impracticalities stemming from the clique's overly conservative nature, alternative graph-theoretic models have been introduced in the literature. The *s-clique* model, first introduced by Luce (1950), relaxes the requirement of direct interaction. Associating the number of intermediary links with the graph-theoretic notion of distance, the *s*-clique definition requires vertices within the group to be at most *s*-distant. Since intermediaries may not be a part of the *s*-clique itself, Alba (1973) proposed a definition of the so-called *sociometric clique of diameter s*, more commonly known as *s-club* (Mokken, 1979), requiring the existence of connections solely through intermediaries belonging to the group. Clubs guarantee easy reachability, however, they do not fare well in terms of other cohesiveness properties. For example, star graphs, i.e., graphs in which one "hub" vertex is directly linked to all other vertices, with no direct links between them, possess a 2-club structure and suffer from a low familiarity and a high vulnerability to the incident of a hub dysfunction.

The latter observation drew the attention towards the necessity in some applications to consider clique-like models emphasizing high level of familiarity and robustness. In particular, Barnes (1968) adopted the notion of *edge density* to address familiarity within a group. More recently this concept was formalized under the so-called γ-*quasi-clique* model (Abello et al., 2002) that ensures a certain minimum ratio γ of the number of existing links to the maximum possible number of links within the group. Seidman (1983) argues that edge density is a rather averaging property and may result in a group with highly cohesive regions involving a high volume of direct interactions, coupled with very sparse regions, relying mostly on indirect interactions with the rest of the group. His observation led to defining a *k-core*, a concept restricting the minimum number of direct links an element must have with the rest of the cluster. While a *k*-core guarantees a certain minimum number *k* of neighbors within the group, the number of non-neighbors within the group may still be much higher than *k*, indicating a low level of familiarity within the group relative to its size.

* Corresponding author.
E-mail addresses: jeff.pattillo@gmail.com (J. Pattillo), youssefn@mit.edu (N. Youssef), butenko@tamu.edu (S. Butenko).

0377-2217/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ejor.2012.10.021

In an earlier work, Seidman and Foster (1978) proposed the notion of s-plex, controlling the number of non-neighbors that elements within the group are allowed to have. In addition to high level of familiarity within the group ensured by its definition for low values of s, an s-plex fares well with respect to robustness expressed in terms of *vertex connectivity*, which is the minimum number of vertices that need to be removed in order to disconnect the graph. Vertex connectivity has recently been linked to social cohesion in social network analysis literature (Moody and White, 2003), where it quickly became a central concept referred to as *structural cohesion*. Thus, the related notion of *k-connected subgraph*, which ensures that the group remains connected unless at least k elements are deleted, can be used as another natural model of a cohesive group. Consistent with the previous literature in graph theory, which defines a *block* to be a maximal connected subgraph that cannot be disconnected by removing a single vertex, we will call a subset of vertices inducing a k-connected subgraph a *k-block*.

Yet another model of a cluster was introduced recently in a study of protein interaction networks (Yu et al., 2006), where an *s-defective clique*, which differs from a clique by at most s missing edges, was used to predict protein interactions. Some of the more recent cluster models proposed in the literature appear to be "hybrids" enforcing a mix of desired group properties. For instance, the (λ, γ)-*quasi-clique* model (Brunato et al., 2008), in addition to requiring the group to be a γ-quasi-clique, sets a lower bound λ on the fraction of the elements that each member of the group must neighbor. In another example, the *k-robust s-club* model requires an s-club to have at least k distinct paths of length at most s between any two vertices (Veremyev and Boginski, 2012), which implies that the s-club preserves its diameter even if up to k elements are removed from the set.

Note that all concepts mentioned as alternatives to clique in the previous paragraph were defined using a parameter, s; k; γ; or λ. Moreover, for $s = 1$ ($s = 0$ for an s-defective clique); $k = n - 1$; $\gamma = 1$; and $\lambda = 1$, where n is the number of vertices in the group being defined, each of the above definitions describes a clique. Hence, defining each of these concepts for an *arbitrary* value of the corresponding parameter yields a *generalization* of the notion of a clique, since it includes the clique definition as a special case. On the other hand, defining any of the concepts above for a *fixed* value of the corresponding parameter, i.e., positive integer s or $k > 1$; real γ and $\lambda \in (0, 1)$, provides a *clique relaxation* (Kosub, 2005; McClosky, 2010).

The described clique relaxation concepts, as well as numerous other similar definitions have emerged in an ad hoc and somewhat spontaneous fashion and were motivated by cluster–detection problems arising in a wide variety of applications. Furthermore, some clique relaxation models have been reinvented under different nomenclature. Despite the obvious practical importance of these models, little work has been done towards establishing theoretical and algorithmic foundations for studying the clique relaxations in a systematic fashion. As a result, applied researchers seeking an appropriate model of a cluster in their application of interest may quickly get overwhelmed by the wide range of models available in the literature. This paper aims to start filling this gap by proposing a taxonomy classifying the previously defined clique relaxations under a unified framework. More specifically, we build on the elementary graph-theoretic properties of cliques to provide a hierarchically ordered classification of clique relaxation models. We complement the taxonomy by deriving bounds on the cohesiveness properties guaranteed by the so-called *canonical* clique relaxations, defined later. The established bounds are proved to be sharp, thus providing rigorously grounded guidelines for practitioners in selecting a cluster model most suited for a particular application. The proposed taxonomy also helps to unveil some structural properties of the considered models that may facilitate the choice of methods for solving the corresponding optimization problems. In addition, it uncovers potential horizons for developing and analyzing new clique relaxation models.

The remainder of this paper is organized as follows. After furnishing the definitions and notations used throughout the paper in Section 2, we describe the proposed taxonomy of clique relaxation models in Section 3. Some basic structural properties of the considered clique relaxations and their implications for choosing the appropriate approaches to solving the corresponding optimization problems are discussed in Section 4. A comprehensive and rigorous analysis of guaranteed cohesiveness properties for the canonical clique relaxation structures is given in Section 5. Section 6 discusses some practical considerations motivated by findings from this analysis, and Section 7 concludes the paper. Finally, Appendix A contains some background information from extremal graph theory and provide proofs of some of the technical results presented in the paper.

2. Definitions and notations

A simple undirected graph $G = (V, E)$, is defined by the set of vertices V and the set of edges E connecting pairs of vertices. If $(v, v') \in E$, the two vertices v and v' in G are called *adjacent* or *neighbors*, and the edge (v, v') is said to be *incident* to v and v'. The set of all neighbors of a vertex v in G is denoted by $N_G(v)$, and its cardinality $|N_G(v)|$ is called the degree of v in G and is denoted by $deg_G(v)$. The minimum and the maximum degree of a vertex in G are denoted by $\delta(G)$ and $\Delta(G)$, respectively. A graph $G' = (V', E')$ is a *subgraph* of $G = (V, E)$ if $V' \subseteq V$ and $E' \subseteq E$. Given a subset of vertices $S \subseteq V$, the *subgraph induced by S*, $G[S]$, is obtained by deleting all vertices in $V \backslash S$ and the edges incident to at least one of them. A *path* of length r between vertices v and v' in G is a subgraph of G defined by an alternating sequence of distinct vertices and edges $v \equiv v_0, e_0, v_1, e_1, \ldots, v_{r-1}, e_{r-1}, v_r \equiv v'$ such that $e_i = (v_i, v_{i+1}) \in E$ for all $1 \leqslant i \leqslant r - 1$. Two vertices v and v' are *connected* in G if G contains at least one path between v and v'. A graph is *connected* if all its vertices are pairwise connected and *disconnected* otherwise. The *distance* between two connected vertices v and v' in G, denoted by $d_G(v, v')$, is the shortest length of a path between u and v in G. The largest distance among the pairs of vertices in G defines the diameter of the graph, $diam(G) = \max_{v,v' \in V} d_G(v, v')$. The *connectivity* or *vertex connectivity* $\kappa(G)$ of G is given by the minimum number of vertices whose deletion yields a disconnected or a trivial graph. The *density* $\rho(G)$ of G is the ratio of the number of edges to the total number of possible edges, i.e., $\rho(G) = |E| / \binom{|V|}{2}$.

A subset of vertices $D \subseteq V$ is called a dominating set in G if every vertex in the graph is either in D or has a neighbor in D. A *complete graph* is a graph that contains all possible edges and is denoted by K_n, where n is its number of vertices. The *complement* $\overline{G}$ of $G = (V, E)$ is defined by $\overline{G} = (V, \overline{E})$, where $\overline{E}$ is such that $E \cap \overline{E} = \emptyset$ and $K_{|V|} = (V, E \cup \overline{E})$. A *clique* C is a subset of vertices $C \in V$ such that the induced subgraph $G[C]$ is complete. The size of a largest clique in G is referred to as the *clique number* of G and is denoted by $\omega(G)$.

Next, some of the well known clique relaxation models, which are central for this study and were already mentioned in the previous section, are formally defined. We assume that the constants s and k are positive integers and $\lambda, \gamma \in (0, 1]$ are reals. In all definitions below, S is assumed to be a subset of vertices in $G = (V, E)$.

Definition 1 (*s-clique*). S is called an s-clique if $d_G(v, v') \leqslant s$, for any $v, v' \in S$.

Definition 2 (*s-club*). S is an s-club if $diam(G[S]) \leqslant s$.

Definition 3 (*s-plex*). S is an s-plex if $\delta(G[S]) \geqslant |S| - s$.

Definition 4 (*s-defective clique*). S is an s-defective clique if $G[S]$ contains at least $\binom{|S|}{2} - s$ edges.

Definition 5 (*k-core*). S is a k-core if $\delta(G[S]) \geqslant k$.

Definition 6 (*k-block*). S is a k-block if $\kappa(G[S]) \geqslant k$.

Definition 7 (γ-*quasi-clique*). S is a γ-quasi-clique if $\rho(G[S]) \geqslant \gamma$.

Definition 8 ((λ, γ)-*quasi-clique*). S is a (λ, γ)-quasi-clique if $\delta(G[S]) \geqslant \lambda(|S| - 1)$ and $\rho(G[S]) \geqslant \gamma$.

Definition 9 (*k-hereditary s-club*). S is a k-hereditary s-club if $diam(G[S \backslash S']) \leqslant s$ for any $S' \subset S$ such that $|S'| \leqslant k$.

It should be noted that, in general, depending on the choice of k and a graph instance G, a nonempty k-core or k-block may not exist in G. This observation has led to the introduction of the notion of *graph degeneracy* based on the concept of a k-core. Namely, a graph is called *d-degenerate* if it does not contain a nonempty k-core for $k > d$. The *degeneracy* of G is the smallest d for which G is d-degenerate, which is the same as the largest k for which G has a nonempty k-core.

3. A taxonomy of clique relaxation models

Most of the elementary graph concepts, such as degree, distance, diameter, density, connectivity, and domination, can be used to derive alternative, equivalent definitions of a clique. We state this observation formally in the following proposition, which is trivial to verify.

Proposition 1. *A subset of vertices C is a clique in G if and only if one of the following conditions hold:*

(a) $d_G(v, v') = 1$, *for any* $v, v' \in C$;
(b) $diam(G[C]) = 1$;
(c) $D = \{v\}$ *is a dominating set in* $G[C]$, *for any* $v \in C$;
(d) $\delta(G[C]) = |C| - 1$;
(e) $\rho(G[C]) = 1$;
(f) $\kappa(G[C]) = |C| - 1$.

In the remainder of this paper, we refer to the conditions specified in the above proposition as *elementary clique-defining properties*. These properties are summarized in Table 1, together with the corresponding graph concepts defining each property. The rows of the table are split into two parts, with the first part corresponding to the parameters whose value is set to the lowest possible value in the clique definition (distance, diameter, size of a set guaranteeing domination), and the second part containing the parameters required to have the highest possible value for the set of a *given size* (degree, density, and connectivity).

Aiming to derive a *minimal* set of *simple* rules based on the elementary clique-defining properties that would allow us to obtain the known clique relaxation models in a systematic fashion, we examine the relation of Definitions 1–9 to the alternative clique definitions summarized in Table 1. It becomes apparent that each of the defined clique relaxation models essentially relaxes at least one of the elementary clique-defining properties according to some simple rules that can be classified into two broad categories.

Table 1
Alternative clique definitions based on elementary clique-defining properties.

Parameter	Definition
Distance	Vertices are distance *one* away from each other
Diameter	Vertices induce a subgraph of diameter *one*
Domination	Every *one* vertex forms a dominating set
Degree	Each vertex is connected to *all* vertices
Density	Vertices induce a subgraph that has *all* possible edges
Connectivity	*All* vertices need to be removed to obtain a disconnected induced subgraph

Namely, some relaxations are created by providing an upper bound on the extent to which an elementary clique-defining property is allowed to be violated, while others aim to ensure the presence of an elementary clique-defining property that characterizes a clique of a *given minimum size*. Each of these two cases is elaborated in more detail in one of the following two subsections.

3.1. Restricting violation of an elementary clique-defining property

3.1.1. Increasing a parameter that has the lowest possible value in a clique

In the scenarios described in the first three rows of Table 1, we obtain a clique relaxation model by increasing a parameter that was set to the lowest possible value in an alternative clique definition. Such models are created by naturally replacing *one* in one of the elementary clique-defining properties with (*at most*) s. In particular, instead of requiring the (upper bound on the) diameter of the induced subgraph to be equal to *one*, an s-club relaxes this requirement to allow a diameter *at most s*. Similarly, by replacing *one* with *at most s* in the elementary clique-defining properties based on distance and domination, we obtain definitions of s-clique and s-plex, respectively. In the case of s-plex, we use the fact that S is an s-plex if and only if any subset of s vertices from S forms a dominating set in $G[S]$ (Seidman and Foster, 1978).

3.1.2. Reducing a parameter that has the highest possible value in a clique of a given size

Note that, while we were able to define s-plex by relaxing an upper bound on the number of vertices ensuring domination, the original definition of s-plex was based on restricting the number of non-neighbors that a vertex can have within the group (Seidman and Foster, 1978). This definition naturally corresponds to allowing, for every vertex, s exceptions (including the vertex itself) in the degree-based definition of a clique. Namely, we just replace *all* by *all but s* in the degree-based definition of a clique to obtain the s-plex definition. Similarly, the density-based clique definition yields the s-defective clique model. By applying the same logic to the clique definition based on connectivity, we obtain a new clique relaxation model, which we propose to call an *s-bundle*.

Definition 10 (*s-bundle*). A subset S of vertices is called an s-bundle if $\kappa(G[S]) \geqslant |S| - s$.

The s-bundle model with a small value of $s > 1$ may prove to be a useful alternative to a clique (which can be equivalently defined as a 1-bundle) in applications emphasizing the robustness of a cluster.

3.2. Ensuring the presence of an elementary clique-defining property

In the last three rows of Table 1, we replace the overly restrictive requirement of a clique definition to have the highest possible value for a parameter (assuming that the size of a set is given) by, instead, imposing a fixed lower bound on that parameter. In such cases, we replace *all* in one of the elementary clique-defining

properties with (*at least*) k. For example, a k-core, does not require each vertex to be connected to *all*, but to *at least* k other vertices. Likewise, we can obtain the definition of a k-block by relaxing the connectivity-based definition of a clique in the same fashion. Similarly, we could define an analogous concept corresponding to the density-based definition of a clique. Namely, we could introduce a clique relaxation model for a subset of vertices inducing a subgraph with at least k edges. However, it is not clear if such a model would present any practical value; therefore, we do not investigate it any further in this paper. Instead, we study its *relative* counterpart, γ-quasi-clique, as will be discussed in the next subsection. It should be noted that, unlike the relaxations described in the previous subsection, the clique relaxation models based on setting a fixed lower bound on a parameter can potentially result in degeneracy (i.e., a structure of this type may be empty if the value of k is set too high for a given graph).

3.3. Absolute and relative relaxations

As suggested by the example of γ-quasi-clique, size-relative or, simply, *relative* clique relaxations is another category of models that needs to be considered. Thus, it makes sense to refer to the above-described categories that use the absolute parameter values (s or k) as *absolute*. We can generate the relative clique relaxation models from the absolute models by replacing s or k by $\gamma|S|$ ($\gamma\binom{|S|}{2}$ in case of density), where $0 \leqslant \gamma \leqslant 1$. While the γ-quasi-clique is, perhaps, the most well known in this category, other relative size-dependent clique relaxations can be defined similarly. For instance, the relative version of s-club would guarantee the induced subgraph $G[S]$ to have a diameter at most $\gamma|S|$. Similarly, one could ensure that at least all but $\gamma|S|$ vertices need to be removed to disconnect the induced subgraph.

3.4. Standard and weak relaxations

In definitions of most of the clique relaxation models discussed above (s-clique being the only exception), we required the relaxed clique-defining properties to be satisfied within the *induced subgraph*. However, as the example of s-clique suggests, in some cases it is sufficient to require the same property to be satisfied within the *original graph* instead of the induced subgraph. In particular, this can be done in the situations involving the elementary clique-defining properties based on distance and connectivity, both of which can be defined through paths. In the case of connectivity, Menger's theorem (Diestel, 1997) asserts that a graph is k-connected if and only if there are at least k vertex-independent paths (i.e., paths with no common internal vertex) between any two of its vertices. Thus, by requiring the conditions on pairwise distances and connectivity to hold in the whole graph rather than the subgraph induced by a cluster's vertices, we allow the paths in the corresponding definitions to pass through vertices outside of the cluster. As a result, we obtain a relaxation with weaker cohesiveness properties. We will refer to such relaxations as *weak*, while the relaxations that require the relaxed clique-defining property to be satisfied in the induced subgraph will be called *standard*. For example, an s-club is a standard relaxation, while an s-clique is its weak counterpart and could be alternatively called a *weak s-club*. Similarly, we could define a *weak k-block* as a subset of vertices such that there are at least k vertex-independent paths between any two of its vertices in the original graph.

3.5. Structural and statistical relaxations

In a recent survey of locally dense structures used in network analysis, Kosub (2005) distinguished between *structural* clique relaxations, such as s-plexes and k-cores, and their *statistical* counterparts, in which a certain desirable property is required to be satisfied on average over all group members. An example of a statistically dense group is *densest subgraph*, which is a subset of vertices that maximizes the average degree of a vertex in the corresponding induced subgraph. According to Kosub (2005), "In general, statistically dense groups reveal only few insights into the group structure". Thus, in the remainder of this paper we concentrate on studying the structural clique relaxation models. An interested reader can easily develop the corresponding statistical clique relaxation concepts. We remark, however, that edge density is a structural property that is averaging in nature; therefore, the quasi-clique model can be thought of as a statistical clique relaxation as well as structural.

3.6. Order of a clique relaxation

Calling the clique itself a zero-order clique relaxation, the aforementioned clique-like objects, which were obtained by relaxing only one clique-defining property, are referred to as *first-order clique relaxations*. Higher-order clique relaxations can be defined by relaxing multiple clique-defining properties simultaneously. The second-order relaxations would correspond to relaxing two elementary clique-defining properties at the same time. For instance, the (λ, γ)- quasi-clique, based on relaxing both degree and density requirements, is a second-order relaxation. While any pair of properties can be enforced simultaneously in order to define a second-order model, in some cases requiring an extra property may be redundant. For example, as we will discuss in Section 5, an s-plex usually has a low diameter and a high connectivity to start with, hence it makes little sense to combine it with diameter or connectivity-based relaxations. On the other hand, if ensuring two of the relaxed clique properties is insufficient to guarantee the desired cohesiveness, one may relax more than two elementary clique-defining properties at a time to obtain relaxations of a higher order.

3.6.1. Hereditary higher-order relaxations

While a higher-order relaxation can be created by enforcing several relaxed clique-defining properties simultaneously, one of the properties, connectivity, can also be *embedded* into a definition of a clique relaxation. As an example, a k-hereditary s-club S can be viewed as a second-order clique relaxation structure defined by embedding k-connectivity into the definition of an s-club. Unlike its *simple* second-order counterpart, which would be defined as a subset of vertices S such that $\kappa(G[S]) \geqslant k$ and $diam(G[S]) \leqslant s$ and could be called *k-connected s-club*, the k-hereditary s-club requires that not only does the s-club S induce a k-connected subgraph, but also that removal of up to k vertices still preserves the s-club property. The property of k-heredity, which will be discussed in the next section, is embedded within the structure defined by other properties involved in the definition of a hereditary higher order relaxation, which makes it fundamentally different from the simple higher order relaxations that combine multiple properties in a straightforward fashion.

3.7. Additional elementary clique-defining properties and canonical models

The list of elementary clique-defining properties presented above is, by no means, exhaustive and is restricted to the concepts that appeared in various important applications in the literature. To illustrate the diversity of clique relaxation models covered by the proposed taxonomy, we will mention several additional elementary clique-defining properties. These properties are based on the classical graph-theoretic notions that are very closely related to the clique concept. Namely, a subset I of vertices is called an *independent set* if the corresponding induced subgraph $G[I]$ has

no edges. The *independence number* $\alpha(G)$ is the size of a largest independent set in G. Obviously, I is an independent set in G if and only if I is a clique in $\overline{G}$. A subset C of vertices is called a *vertex cover* if each edge in G is incident to at least one vertex in C. The *vertex cover number* $\tau(G)$ is the minimum size of a vertex cover in G. Note that C is a vertex cover if and only if $V\backslash C$ is an independent set. Given a positive integer k, a *proper k-coloring* of G is a partition of the set of vertices V into k non-overlapping independent sets $I_1,\ldots,I_k$, each of which defines a different *color class*. The minimum value of k for which a proper k-coloring exists is called the *chromatic number* of G and is denoted by $\chi(G)$. Similarly, the *clique cover problem* is to find a minimum k for which there exists a partition of the set V of vertices into k non-overlapping cliques, and the corresponding value of k is called the *clique cover number* and is denoted by $\bar{\chi}(G)$. It is easy to check that $\bar{\chi}(G) = \chi(\overline{G})$. The next concept is the analog of graph connectivity defined with respect to edges. More specifically, the *edge-connectivity* $\lambda(G)$ is the minimum number of edges that need to be removed in order to disconnect the graph. The following proposition, which is trivial to check, states the elementary clique-defining properties based on the concepts just defined.

Proposition 2. *A subset of vertices C is a clique in G if and only if one of the following conditions hold:*

(g) $\alpha(G[C]) = 1$;
(h) $\tau(G[C]) = |C| - 1$;
(i) $\chi(G[C]) = |C|$;
(j) $\bar{\chi}(G[C]) = 1$;
(k) $\lambda(G[C]) = |C| - 1$.

The reader can easily derive the corresponding clique relaxations based on the rules outlined above. It is not clear whether the resulting models will be of use in any applications. Therefore, in the remainder of this paper we mostly will concentrate on studying the clique relaxation models that were originally motivated by important applications and, thus, are of proven practical value. To be specific, the models of interest are s-club, s-plex, k-core, γ-quasi-clique and k-block. We treat these models as the *canonical* models for the corresponding graph invariants used to formulate the elementary clique-defining properties. Thus, s-club is the canonical clique relaxation model for diameter; s-plex – for domination; k-core – for degree; γ-quasi-clique – for density; and k-block – for connectivity. All of the canonical models, except for quasi-clique, are absolute clique relaxation models. We selected quasi-clique over s-defective clique to represent a density-based relaxation in this study due to two reasons. First, the concept of density is traditionally discussed as a relative measure by definition; and second, γ-quasi-clique is by far more widely represented in the literature. Note that the distance property for standard clique relaxations is equivalent to the same property for the diameter, since we limit the analysis to induced subgraphs.

To illustrate the definitions of the canonical clique relaxations, as well as their necessity, consider an example arising in the analysis of protein interaction networks, where an important problem is to determine the protein complexes responsible for biological processes of interest (Levy et al., 2006). Protein complexes have been found to come in a variety of structures, many of which appear to be well described by various clique relaxation models. Five such structures, together with the names of the corresponding protein complexes, as well as a clique relaxation model each of them is best described with, are shown in Fig. 1. As we start to explore the structure of each clique relaxation, it will become apparent that we have matched each protein complex with the clique relaxation most equipped to find it within a protein interaction network. This illustrates the importance of each considered relaxation, as different settings require different structures.

4. Optimization problems

In most application scenarios dealing with clique relaxation models, one is interested in computing *large* clusters of a certain type. While typically multiple large clusters (partitioning into clusters), not necessarily largest possible, are of practical interest, the maximum size of a clique relaxation of a given kind quantifies the *global cohesiveness* of the analyzed network in terms of the considered clique relaxation model of a cohesive subgroup. Moreover, it provides the tight upper bound on the size of clusters of the considered type that exist in the network, and hence facilitates computing such clusters. Thus, we are interested in issues associated with the corresponding optimization problems. The purpose of this section is to point out structural properties of different types of clique relaxation models that may facilitate the process of selecting computational techniques that would be appropriate for solving the corresponding optimization problems.

First, let us formally define the general optimization problem for a clique relaxation model. Let RELAXED CLIQUE refer to a subset of vertices that satisfies the definition of an arbitrary clique relaxation concept. The following definitions are general and can be adopted for a particular clique relaxation model by replacing the term RELAXED CLIQUE with the name of the corresponding structure (i.e., s-club, s-plex, etc.).

Definition 11. A subset of vertices S is called a maximal RELAXED CLIQUE if it is a RELAXED CLIQUE and is not a proper subset of a larger RELAXED CLIQUE.

Definition 12. A subset of vertices S is called a maximum RELAXED CLIQUE if there is no larger RELAXED CLIQUE in the graph. The maximum RELAXED CLIQUE problem asks to compute a maximum RELAXED CLIQUE in the graph, and the size of a maximum RELAXED CLIQUE is called the RELAXED CLIQUE number.

Most of the discussion in this section is centered around the concept of *heredity*, which could be thought of as a dynamic property, since it describes the characteristics of a graph undergoing a change, i.e., vertex addition or removal. Heredity is defined with respect to a graph property Π and is formally introduced next.

Definition 13 (*Heredity*). A graph property Π is said to be *hereditary on induced subgraphs*, if for any graph G with property Π the deletion of any subset of vertices does not produce a graph violating Π.

The presence of heredity on induced subgraphs implies certain properties that may help streamlining the study of the corresponding optimization problems. In particular, it turns establishing the computational intractability of the problem into a simple exercise of checking several basic facts about the property Π. Namely, a property Π is called *nontrivial* if it is true for a single-vertex graph and is not satisfied by every graph, and Π is called *interesting* if there are arbitrarily large graphs satisfying Π. The following general complexity result is due to Yannakakis (1978).

Theorem 1 Yannakakis, 1978. *The problem of finding the largest-order induced subgraph not violating property Π that is nontrivial, interesting and hereditary on induced subgraphs is NP-hard.*

In addition, heredity on induced subgraphs is the foundational property for some of the most successful combinatorial algorithms for the maximum clique problem (Carraghan and Pardalos, 1990; Östergård, 2002), which can be generalized to solve any other maximum RELAXED CLIQUE problem based on relaxed clique-defining properties that are hereditary on induced subgraphs. By analyzing the taxonomy introduced in Section 3, we can conclude that the only models that fall within this category are the standard, absolute

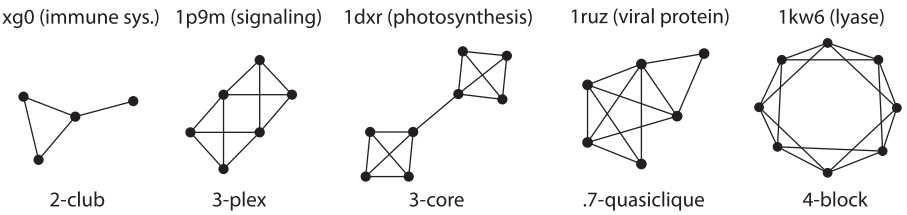

Fig. 1. Examples of protein complexes corresponding to different clique relaxation models.

clique relaxation models obtained by restricting violation of a clique-defining property and based on reducing a parameter that has the highest possible value in a clique of a given size. These are the models described in the second paragraph of subSection 3.1, namely, s-plex, s-defective clique, and s-bundle. Hence, the corresponding optimization problems are NP-hard and can be solved by adopting the combinatorial algorithms for the maximum clique problem proposed earlier (Carraghan and Pardalos, 1990; Östergård, 2002). The presence of the heredity property also suggests that these problems are good candidates for solving by methods based on polyhedral combinatorics, as was already demonstrated for two of these models, s-plex (Balasundaram et al., 2011) and s-defective clique (Sherali and Smith, 2006). Moreover, computing maximal RELAXED CLIQUE is trivial in this case, as maximality is guaranteed whenever the current solution cannot be expanded by adding any single vertex from outside.

Even though the properties defining other first-order clique relaxation models do not posses heredity, they have closely related characterizations that can also be utilized in designing solution methods. We propose to define these dynamic properties of *weak heredity*, *quasi-heredity*, and *k-heredity* as follows.

Definition 14 (*Weak heredity*). A graph property Π is said to be *weakly hereditary*, if for any graph $G = (V, E)$ with property Π all subsets of V demonstrate the property Π in G.

Definition 15 (*Quasi-heredity*). A graph property Π is said to be *quasi-hereditary*, if for any graph $G = (V, E)$ with property Π and for any size $0 \leqslant r < |V|$, there exists some subset $R \subset S$ with $|R| = r$, such that $G[S \backslash R]$ demonstrates property Π.

Definition 16 (*k-Heredity*). A graph property Π is said to be *k-hereditary on induced subgraphs*, if for any graph G with property Π the deletion of any subset of vertices with up to k vertices does not produce a graph violating Π.

Note that weak heredity considers whether a certain property is still applicable for all subsets in the original graph, as opposed to heredity on the induced subgraph. On the other hand, quasi-heredity essentially requires the existence of a sequence of vertices such that their removal in this sequence preserves, at every step of the vertex removal process, the property in the remaining subgraph. However, property Π may not exist for every subset R of vertices removed from S. Also, observe that heredity implies both weak heredity and quasi-heredity, whereas the latter two do not appear to have any definitive relation.

The weak heredity property holds for s-cliques and weak k-blocks, both of which are weak clique relaxation models. The weak heredity property allows to reduce the corresponding clique relaxation structures to cliques in auxiliary graphs. Thus, the numerous algorithms developed for the maximum clique problem, can be directly applied to auxiliary graphs in order to solve the optimization problems dealing with the weak clique relaxations. In the case of s-clique, the auxiliary graph is given by the *power graph*. Given a graph $G = (V, E)$, its t-th power graph $G^t = (V, E^t)$ has the same set of vertices V and the set of edges E^t that connects pairs of vertices that are distance at most t from each other in G. Obviously, S is an s-clique in G if and only if S is a clique in G^s. Similarly, for the weak k-block, we can define an auxiliary graph $G(k) = (V, E(k))$, where $(v, v') \in E(k)$ if and only if there are at least k vertex-independent paths between v and v' in G. Then, again, S is a weak k-block in G if and only if S is a clique in $G(k)$.

The definition of quasi-heredity was motivated by the observation that this property holds for the γ-quasi-clique model, since the iterative removal of the lowest degree vertex will preserve at least the same density in the induced subgraphs at every step (Pattillo et al., 2013). The presence of this property suggests that developing heuristics based on greedy sequencing of vertices may prove effective in practice (Glover and Kochenberger, 2002). Finally, the k-heredity property is what we enforce in hereditary higher-order clique relaxations discussed in the previous section. Not surprisingly, the first hereditary second-order relaxation studied involves s-clubs, which do not posses any type of heredity considered if $s > 1$. This is demonstrated by a cycle of length $2s + 1$; its set of vertices is an s-club that contains no s-club of size $s + 2, \ldots, 2s$.

On an optimistic note, two of the discussed maximum RELAXED CLIQUE problems, the maximum k-core and the maximum k-block, can be solved in polynomial time. More specifically, all maximal k-cores can be computed in $O(|E||V|\log|V|)$ time (Kosub, 2005); biconnected and tri-connected components can be found in $O(|V| + |E|)$ time (Kammer and Täubig, 2005), while the only known algorithms for computing k-connected components for $k > 3$ are based on identifying all k-cutsets (subsets of k vertices that, if removed, disconnect the graph) in the graph. Such procedures require $O(2^k|V|^3)$ time and, hence, become expensive for high values of the constant k.

5. Cohesiveness properties of standard first-order clique relaxation models

The hierarchical classification proposed in Section 3 allows to define a wide variety of relaxations with different levels of proximity to the clique structure. However, care must be vested while investing in higher-order relaxations. This requires an in-depth understanding of the properties that first-order relaxations have to offer in terms of the group structure. For instance, it may not be worth restricting an additional property for some first-order relaxation if its structure automatically guarantees good bounds on the desired property. This observation motivates the current section, in which we provide a study of the various structural properties guaranteed by canonical clique relaxations. To better understand the similarities and differences between the canonical relaxations, this section aims to develop sharp worst-case bounds that could be ensured for each of the relaxed elementary clique-defining properties.

Several results of this nature are well-known in graph theory, in particular, in its branch called *extremal graph theory* (Bollobás, 1978), and are summarized in Appendix A. In the remainder of this section, we study the cohesiveness properties of the canonical clique relaxation models, with the emphasis being placed on *sharpness* of the corresponding bounds. Namely, for each value of the parameter used to define a RELAXED CLIQUE structure and for each size

of a RELAXED CLIQUE, we aim to provide an example of a graph on which a worst-case bound for a given elementary clique-defining property is achieved.

In the case of s-club, the cohesiveness properties of interest and their sharpness are trivial to establish, as described in the following statement.

Proposition 3 (*Cohesiveness properties of s-clubs*). *An s-club S satisfies the following conditions:*

(a) $diam(G[S]) \leqslant s$;
(b) *Any* $D \subseteq S$ *such that* $|D| \geqslant |S| - 1$ *is a dominating set in G[S];*
(c) $\delta(G[S]) \geqslant 1$;
(d) $\kappa(G[S]) \geqslant 1$;
(e) $\rho(G[S]) \geqslant \frac{2}{|S|}$.

All these bounds are achieved when S induces a star graph and hence are sharp.

Next we mention some known results for s-plexes that are directly related to the discussion that follows. The diameter and connectivity of a graph $G = (V, E)$ whose vertex set V forms an s-plex are known to satisfy the following conditions (Seidman and Foster, 1978; Kosub, 2005):

$$diam(G[S]) \leqslant 2 \text{ if } s < (|V| + 2)/2; \tag{1}$$

$$diam(G[S]) \leqslant 2s - |V| + 2 \text{ if } s \geqslant (|V| + 2)/2 \text{ and } G \text{ is connected}; \tag{2}$$

$$\kappa(G) \geqslant |V| - 2s + 2. \tag{3}$$

The following proposition states that an s-plex inducing a connected subgraph is also an s-club.

Proposition 4. *If S is an s-plex in G and G[S] is connected then diam (G[S]) ⩽ s.*

Proof. Consider the shortest path between the two most distant vertices v and v' in $G[S]$. This shortest path contains exactly one neighbor of v, since a shorter path could have been obtained otherwise. Now, since v has at most $s - 1$ non-neighbors in S, the path between v and v' is of length at most s, consisting of one neighbor of v and $s - 1$ non-neighbors of v, including v'. □

Note that the bound above is achieved on a set of $s + 1$ vertices of a path of length s, however, it is not sharp for an s-plex of an arbitrary size. A sharp bound on the diameter of an s-plex, which also implies bound (1) and yields a strict improvement of bound (2), is given in the following proposition characterizing the cohesiveness properties of an s-plex.

Proposition 5 (*Cohesiveness properties of s-plexes*). *An s-plex S satisfies the following conditions:*

(a) *If G[S] is connected then* $diam(G[S]) \leqslant d'_s$, *where*

$$d'_s = \max\left\{\left\lceil \frac{|S|}{|S| - s + 1}\right\rceil, 3\left(\left\lfloor \frac{|S| - z}{|S| - s + 1}\right\rfloor - 1\right) + z, z \in \{0, 1, 2\}\right\}. \tag{4}$$

(b) *Any* $D \subseteq S$ *such that* $|D| \geqslant s$ *is a dominating set in G[S];*
(c) $\delta(G[S]) \geqslant |S| - s$;
(d) $\kappa(G[S]) \geqslant |S| - 2s + 2$;
(e) $\rho(G[S]) \geqslant 1 - \frac{s-1}{|S|-1}$.

All these bounds are sharp.

Proof. Bound (a) follows from Lemma 1 in Appendix A by observing that a k-core of a fixed size $|S|$ is also an s-plex with $s = |S| - k$. Properties (b) and (c) are equivalent and are used as alternative definitions of an s-plex (Seidman and Foster, 1978), while (e) trivially follows from (c). Bound (d) is the same as (3) and is known to be sharp (Seidman and Foster, 1978). An extremal example is a graph on n vertices consisting of three complete graphs, $H_1 = K_{n-2s+2}$, and $H_2 = H_3 = K_{s-1}$, with each vertex of H_2 and H_3 connected to each vertex of H_1. Note that (d) implies that an s-plex is connected when its size exceeds $2(s - 1)$. □

Proposition 6 (*Cohesiveness properties of k-cores*). *A k-core S satisfies the following conditions:*

(a) *If G[S] is connected then* $diam(G[S]) \leqslant d'_k$, *where*

$$d'_k = \max\left\{\left\lceil \frac{|S|}{k+1}\right\rceil, 3\left(\left\lfloor \frac{|S| - z}{k+1}\right\rfloor - 1\right) + z, z \in \{0, 1, 2\}\right\}. \tag{5}$$

(b) *Any* $D \subseteq S$ *such that* $|D| \geqslant |S| - k$ *is a dominating set in G[S];*
(c) $\delta(G[S]) \geqslant k$;
(d) $\kappa(G[S]) \geqslant 2k + 2 - |S|$;
(e) $\rho(G[S]) \geqslant \frac{k}{|S|-1}$.

All these bounds are sharp.

Proof. Bound (a) is established in Lemma 1 in Appendix A. Bounds (b), (d), and (e) follows directly from the corresponding properties of Proposition 5 by observing that a fixed k-core S is an s-plex with $s = |S| - k$. □

Proposition 7 (*Cohesiveness properties of k-blocks*). *A k-block S satisfies the following conditions:*

(a) $diam(G[S]) \leqslant \left\lfloor \frac{|S|-2}{k} + 1 \right\rfloor$;
(b) *Any* $D \subseteq S$ *such that* $|D| \geqslant |S| - k$ *is a dominating set in G[S];*
(c) $\delta(G[S]) \geqslant k$;
(d) $\kappa(G[S]) \geqslant k$;
(e) $\rho(G[S]) \geqslant \frac{k}{|S|-1}$.

All these bounds are sharp.

Proof. Bound (a) and its sharpness are shown in Lemma 2 in Appendix A. Knowing that a k-block is also a k-core, any set of size at least $|S| - k$ is a dominating set. This bound is indeed sharp, since a k-connected subgraph could contain a clique of size $|S| - 1$ with an additional vertex adjacent to exactly k vertices from the clique. In this special case, excluding more than k vertices from the set of vertices would no longer guarantee that the additional vertex is dominated by the set of remaining vertices. The same example can be used to prove (c). Since the degree of each vertex in a k-connected subgraph is at least k, there are at least $\frac{k|S|}{2}$ edges, yielding the density of at least $\frac{k}{|S|-1}$. The bound is sharp on k-regular k-connected graphs (Hsu and Luczak, 1994). □

Proposition 8 (*Cohesiveness properties of γ-quasi-cliques*). *A γ-quasi-clique S satisfies the following bounds, each of which is sharp:*

(a) *If G[S] is connected, then* $diam(G[S]) \leqslant d_\gamma$, *where*

$$d_\gamma = \left\lfloor |S| + \frac{1}{2} - \sqrt{\gamma|S|^2 - (2 + \gamma)|S| + \frac{17}{4}} \right\rfloor. \tag{6}$$

(b) *There is no* $t < |S|$ *guaranteeing that any* $D \subseteq S$ *such that* $|D| \geqslant t$ *is a dominating set in* $G[S]$*;*

(c) $\delta(G[S]) \geqslant \left\lceil \gamma\binom{|S|}{2} - \binom{|S|-1}{2} \right\rceil$;

(d) $\kappa(G[S]) \geqslant \left\lceil \gamma\binom{|S|}{2} - \binom{|S|-1}{2} \right\rceil$;

(e) $\rho(G[S]) \geqslant \lceil \gamma\binom{|S|}{2} \rceil / \binom{|S|}{2}$.

Proof. Bound (a) is proved in Lemma 3 in Appendix A. To prove (b), note that for a γ-quasi-clique S, $\gamma\binom{|S|+1}{2} \leqslant \binom{|S|}{2}$ holds for a large enough $|S|$. A γ-quasi-clique could then consist of an independent vertex accompanied by a large enough clique S. In this case, the smallest t guaranteeing that any subset of size t is a dominating set is $t = |S|$. Knowing that the minimum possible degree is no less than the graph's connectivity, bound (c) on the minimum degree for γ-quasi-cliques can be deduced from the lower bound on connectivity (d), which is established next. Let $a = \gamma\binom{|S|}{2} - \binom{|S|-1}{2}$ define the number of edges necessary beyond $K_{|S|-1}$ to achieve density γ. By definition, any γ-quasi-clique S comprises $\gamma\binom{|S|}{2}$ edges. $G[S]$ can then be represented as $K_{|S|}$ missing $\binom{|S|}{2} - \gamma\binom{|S|}{2}$ edges. Since $\binom{|S|}{2} = |S| - 1 + \binom{|S|-1}{2}$, $G[S]$ is $K_{|S|}$ missing $|S| - 1 + \binom{|S|-1}{2} - \gamma\binom{|S|}{2} = |S| - 1 - a$ edges. $K_{|S|}$ being $(|S|-1)$-connected, the removal of $(|S| - 1 - a)$ edges could destroy at most $(|S| - 1 - a)$ vertex-independent paths. Thus, $G[S]$ has at least $|S| - 1 - (|S| - 1 - a) = a$ vertex-independent paths between any two vertices. By Menger's theorem, $\kappa(G[S]) \geqslant a \equiv \gamma\binom{|S|}{2} - \binom{|S|-1}{2}$. To show that this bound is sharp, let us consider a clique of size $|S| - 1$ and a single vertex. Connecting this vertex to $a = \gamma\binom{|S|}{2} - \binom{|S|-1}{2}$ vertices in the clique results in a γ-quasi-clique of size $|S|$. Connectivity of the corresponding graph is equal to the number of edges connecting that single vertex to the clique, i.e., $\gamma\binom{|S|}{2} - \binom{|S|-1}{2}$. □

All the bounds developed above in this section are summarized in Table 2. It should be noted that the cohesiveness properties of weak clique relaxation structures are not nearly as strong as of their standard counterparts. For example, consider the *s*-clique model, which exhibits weak heredity and hence offers an attractive alternative to the *s*-club model from the computational perspective. We can construct graphs containing *s*-cliques that are independent sets. Even if we require an *s*-clique S to induce a connected subgraph, we still cannot guarantee that $diam(G[S]) < |S| - 1$.

6. Practical considerations

Table 2 can be very useful in identifying which clique relaxation is particularly fit for a given application. To choose the appropriate model of a cluster, the essential cohesiveness properties should be identified and candidates for grouping chosen using the appropriate columns of the table. Note that the remaining columns should then be considered, because extraneous cohesiveness requirements may exist and result in valid groups being dismissed for failing to demonstrate the extra structure. In the discussion that follows, we attempt to highlight the important characteristics for each clique relaxation in Table 2. We demonstrate applications for which each clique relaxation appears to be particularly fit because of its characteristics. It is important to note that, when using clique relaxation models to analyze a real-life complex system, one should be cautious with making conclusions regarding the system's behavior based solely on the structural characteristics of the network describing the system, as making such conclusion requires in-depth understanding of domain-specific functions (Alderson, 2008).

The *s*-clique and *s*-club relaxations were designed to guarantee easy reachability between the nodes in a network. A unique feature of these relaxations is their minimal requirements for degree, dominating set size, density, and connectivity. These clique relaxations are particularly adept when data should be clustered with low diameter, but also low density. The *s*-clubs have had success in clustering topically related information on the internet to facilitate faster searches for this reason (Terveen et al., 1999). The internet, along with numerous other networks, demonstrates preferential attachment, meaning new edges tend to appear at nodes that already have high degree (Faloutsos et al., 1999; Doyle et al., 2005). Sets of nodes with low diameter, but also low density, permeate such graphs and often should be grouped despite the sparsity of the corresponding induced subgraph. When this is the case, *s*-clubs or *s*-cliques are the appropriate choices. To decide between the two models, one needs to keep in mind that the *s*-club model possesses stronger cohesiveness properties, while the *s*-clique relaxation has the weak heredity property, and computing *s*-cliques can be reduced to detecting cliques in the *s*th power of a graph, making the numerous algorithms developed for the maximum clique problem directly applicable.

The *s*-plex model is unique in that it ensures nearly every property in Table 2 to an extent (assuming that *s* is small relative to the size of the group of interest). It was specifically introduced in social network analysis literature as an alternative to *s*-clique and *s*-club with more guaranteed structure because the internal structure of low-diameter graphs was "poorly understood" (Seidman and Foster, 1978). Accordingly, it is often useful in applications where cliques are desired but a few missing edges are tolerated, perhaps caused by errors in data collection. Because it ensures a high level of interaction by *all* members (assuming low *s* values), the *s*-plex tends to demonstrate uniform density and substantial symmetry. This makes it particularly adept at identifying protein complexes in protein interaction networks (Luo et al., 2009), where, according to Levy et al. (2006), 85% of complexes currently in the database demonstrate symmetry. In addition, the *s*-plex model may serve as an attractive alternative to cliques in several scenarios arising in computational biochemistry and genomics (Butenko and Wilhelm, 2006; Strickland et al., 2005).

The key property of the *k*-core relaxation is that the corresponding optimization problem is solvable in polynomial time. It has proven a useful tool for pruning a graph in order to find cliques and clique relaxations where a lower bound is known on the degree of the vertices in the induced subgraph (Abello et al., 1999). In some large-scale, sparse instances, the resulting scale reduction is sufficient to be able to compute the maximum clique in the residual graph (Boginski et al., 2005). In addition, *k*-core has been used to detect molecular complexes and predict protein functions (Altaf-Ul-Amin et al., 2003; Bader and Hogue, 2003; Rual et al., 2005).

The *k*-block is specifically defined to guarantee that communication can survive breakdowns in the network. It is often referred to as a "survivable" or "redundant" network in applied fields and is more often used in design rather than analysis of a network. It has been proposed as an alternative to density-based relaxations for identifying complexes in protein interaction networks (Habibi et al., 2010). Recently, it gained popularity in social network

Table 2
Bounds on guaranteed cohesiveness of canonical clique relaxations. The expressions for d'_s, d'_k and d_γ are given in Eqs. (4)–(6), respectively. The bounds on diameter of s-plex, k-core, and γ-quasi-clique are given assuming that $G[S]$ is connected.

$S \subseteq V$ Clique	Diameter "one"	Dominating set "one"	Minimum degree "all"	Connectivity "one"	Edge density "all"
s-Club	$\boxed{s}$	$\lvert S\rvert - 1$	1	1	$\frac{2}{\lvert S\rvert}$
s-Plex	d'_s	$\boxed{s}$	$\lvert S\rvert - s$	$\lvert S\rvert - 2s + 2$	$1 - \frac{s-1}{\lvert S\rvert - 1}$
k-Core	d'_k	$\lvert S\rvert - k$	$\boxed{k}$	$2k + 2 - \lvert S\rvert$	$\frac{k}{\lvert S\rvert - 1}$
k-Block	$\left\lfloor \frac{\lvert S\rvert - 2}{k} + 1 \right\rfloor$	$\lvert S\rvert - k$	k	$\boxed{k}$	$\frac{k}{\lvert S\rvert - 1}$
γ-Quasi-clique	d_γ	$\lvert S\rvert$	$\left\lceil \gamma\binom{\lvert S\rvert}{2} - \binom{\lvert S\rvert - 1}{2} \right\rceil$	$\left\lceil \gamma\binom{\lvert S\rvert}{2} - \binom{\lvert S\rvert - 1}{2} \right\rceil$	$\boxed{\gamma}$

analysis literature, where k-connectivity is referred to as structural cohesion (Moody and White, 2003). Further research on uses for this clique relaxation could prove extremely valuable, especially in applications where network survivability is key. In addition, in applications where robustness of a cluster is most crucial, the s-bundle concept defined in this paper could provide an attractive alternative to k-block. By noting that a fixed s-bundle $|S|$ is a k-block with $k = |S| - s$, we can easily obtain the inequalities characterizing the cohesiveness properties of s-bundles from Proposition 7 by replacing k with $|S| - s$ in the corresponding expressions. One can conclude that s-bundle represents a more cohesive structure than a k-block for most realistic choices of k and s.

Quasi-cliques, like the s-plex model, demonstrate a high level of interaction between all members. This inevitably results in numerous other properties, as was true with s-plex. What makes it different from s-plex, however, is that the connections within the group are not as structured and, depending on size, no minimum degree is required. This makes it useful in data mining applications where high density sets should be grouped regardless of structure. In addition to being employed in computational biology (Bhattacharyya and Bandyopadhyay, 2009; Matsuda et al., 1999), quasi-cliques were successfully used to mine massive sets of telecommunications data in order to find a good way of organizing it (Abello et al., 1999; Abello et al., 2002). A heuristically defined relaxation called *paraclique*, which is very similar to quasi-clique, proved useful in mining biological data for functional relationships between attributes (Perkins and Langston, 2009). This approach yielded cohesive subgroups that dwarfed the largest cliques and helped reveal relationships previously missed due to a small subset of missing edges.

7. Conclusion

We introduced a taxonomy of clique relaxations that encompasses many of the popular models studied in the literature and establishes foundations for a systematic study of the corresponding optimization problems and their applications. The paper opens the door for many interesting research directions that can be undertaken in exploring the existing, as well as newly identified clique relaxation models. In particular, the established bounds on cohesiveness properties of the canonical clique relaxation models should help to identify higher-order relaxations that are worth investigating. Exploring the proposed directions for solving the considered optimization problems computationally is of significant practical interest. The relationship between optimization problems dealing with absolute and relative relaxations corresponding to the same elementary clique-defining property is an interesting related question to study. In addition, examining network clustering techniques based on various clique relaxation structures is another direction to explore.

Acknowledgements

This research was partially supported by the US Department of Energy Grant DE-SC0002051 and US Air Force Office of Scientific Research Award Nos. FA9550-09-1-0154 and FA9550-12-1-0103.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ejor.2012.10.021.

References

Abello, J., Pardalos, P., Resende, M., 1999. On maximum clique problems in very large graphs. In: Abello, J., Vitter, J. (Eds.), External Memory Algorithms and Visualization, DIMACS Series on Discrete Mathematics and Theoretical Computer Science, vol. 50. American Mathematical Society, pp. 119–130.
Abello, J., Resende, M., Sudarsky, S., 2002. Massive quasi-clique detection. In: Rajsbaum, S. (Ed.), LATIN 2002: Theoretical Informatics. Springer-Verlag, London, pp. 598–612.
Alba, R., 1973. A graph-theoretic definition of a sociometric clique. Journal of Mathematical Sociology 3, 113–126.
Alderson, D.L., 2008. Catching the "network science" bug: insight and opportunity for the operations researcher. Operations Research 56 (5), 1047–1065.
Altaf-Ul-Amin, M., Nishikata, K., Koma, T., Miyasato, T., Shinbo, Y., Arifuzzaman, M., Wada, C., et al., 2003. Prediction of protein functions based on k-cores of protein-protein interaction networks and amino acid sequences. Genome Informatics 14, 498–499.
Bader, G.D., Hogue, C.W.V., 2003. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 4 (2).
Balasundaram, B., Butenko, S., Hicks, I., 2011. Clique relaxations in social network analysis: the maximum k-plex problem. Operations Research 59, 133–142.
Barnes, J., 1968. Networks and political process. In: Swart, M. (Ed.), Local-Level Politics. Aldine, Chicago, pp. 107–130.
Bhattacharyya, M., Bandyopadhyay, S., 2009. Mining the largest quasi-clique in human protein interactome. In: IEEE International Conference on Artificial Intelligence Systems. IEEE Computer Society, Los Alamitos, CA, USA, pp. 194–199.
Boginski, V., Butenko, S., Pardalos, P., 2005. Statistical analysis of financial networks. Computational Statistics & Data Analysis 48, 431–443.
Bollobás, B., 1978. Extremal Graph Theory. Academic Press, New York.
Brunato, M., Hoos, H., Battiti, R., 2008. On effectively finding maximal quasi-cliques in graphs. In: Maniezzo, V., Battiti, R., Watson, J. (Eds.), Proc. 2nd Learning and Intelligent Optimization Workshop, LION 2, LNCS, vol. 5313. Springer Verlag.
Butenko, S., Wilhelm, W., 2006. Clique-detection models in computational biochemistry and genomics. European Journal of Operational Research 173, 1–17.
Carraghan, R., Pardalos, P., 1990. An exact algorithm for the maximum clique problem. Operations Research Letters 9, 375–382.
Diestel, R., 1997. Graph Theory. Springer-Verlag, Berlin.
Doyle, J.C., Alderson, D.L., Li, L., Low, S., Roughan, M., Shalunov, S., Tanaka, R., Willinger, W., 2005. The "robust yet fragile" nature of the internet. Proceedings of the National Academy of Sciences 102 (41), 14497–14502.
Faloutsos, M., Faloutsos, P., Faloutsos, C., 1999. On power-law relationships of the internet topology, In: Proceedings of the ACM-SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication Cambridge, MA, pp. 251–262.
Glover, F., Kochenberger, G. (Eds.), 2002. Handbook of Metaheuristics. Springer, London.
Habibi, M., Eslahchi, C., Wong, L., 2010. Protein complex prediction based on k-connected subgraphs in protein interaction network. BMC Systems Biology 4, 129.

Hsu, D.F., Luczak, T., 1994. On the k-diameter of k-regular k-connected graphs. Discrete Mathematics 133 (1–3), 291–296.
Kammer, F., Täubig, H., 2005. Connectivity. In: Brandes, U., Erlebach, T. (Eds.), Network Analysis, LNCS, vol. 3418. Springer-Verlag, Berlin Heidelberg, pp. 143–177.
Kosub, S., 2005. Local density. In: Brandes, U., Erlebach, T. (Eds.), Network Analysis, LNCS, vol. 3418. Springer-Verlag, Berlin Heidelberg, pp. 112–142.
Levy, E.D., Pereira-Leal, J.B., Chothia, C., Teichmann, S.A., 2006. 3d complex: a structural classification of protein complexes. PLoS Computational Biology 2, e155.
Luce, R., 1950. Connectivity and generalized cliques in sociometric group structure. Psychometrika 15, 169–190.
Luce, R., Perry, A., 1949. A method of matrix analysis of group structure. Psychometrika 14, 95–116.
Luo, F., Li, B., Wan, X., Sheuermann, R., 2009. Core and periphery structures in protein interaction networks. BMC Bioinformatics, 10.
Matsuda, H., Ishihara, T., Hashimoto, A., 1999. Classifying molecular sequences using a linkage graph with their pairwise similarities. Theoretical Computer Science 210 (2), 305–325.
McClosky, B., 2010. Clique relaxations. In: Cochran, J.J., Cox, L.A., Keskinocak, P., Kharoufeh, J.P., Smith, J.C. (Eds.), Wiley Encyclopedia of Operations Research and Management Science. John Wiley & Sons Inc., pp. 650–657.
Mokken, R., 1979. Cliques, clubs and clans. Quality and Quantity 13, 161–173.
Moody, J., White, D.R., 2003. Structural cohesion and embeddedness: a hierarchical concept of social groups. American Sociological Review 68, 103–127.
Östergård, P.R.J., 2002. A fast algorithm for the maximum clique problem. Discrete Applied Mathematics 120, 197–207.
Pattillo, J., Veremyev, A., Butenko, S., Boginski, V., 2013. On the maximum quasi-clique problem. Discrete Applied Mathematics 161, 244–257.
Perkins, A.D., Langston, M.A., 2009. Threshold selection in gene co-expression networks using spectral graph theory techniques. BMC Bioinformatics 10 (Suppl. 11), S4.
Rual, J.-F., Venkatesan, K., et al., 2005. Towards a proteome-scale map of the human proteinprotein interaction network. Nature 437, 1173–1178.
Seidman, S.B., 1983. Network structure and minimum degree. Social Networks 5, 269–287.
Seidman, S.B., Foster, B.L., 1978. A graph theoretic generalization of the clique concept. Journal of Mathematical Sociology 6, 139–154.
Sherali, H.D., Smith, J.C., 2006. A polyhedral study of the generalized vertex packing problem. Mathematical Programming 107 (3), 367–390.
Strickland, D.M., Barnes, E., Sokol, J.S., 2005. Optimal protein structure alignment using maximum cliques. Operations Research 53, 389–402.
Terveen, L., Hill, W., Amento, B., 1999. Constructing, organizing, and visualizing collections of topically related, web resources. ACM Transactions on Computer–Human Interaction 6, 67–94.
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.
Yannakakis, M., 1978. Node-and edge-deletion NP-complete problems. In: STOC '78: Proceedings of the 10th Annual ACM Symposium on Theory of Computing. ACM Press, New York, NY, p. 253264.
Yu, H., Paccanaro, A., Trifonov, V., Gerstein, M., 2006. Predicting interactions in protein networks by completing defective cliques. Bioinformatics 22, 823–829.