European Journal of Operational Research 239 (2014) 349–362

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Discrete Optimization

Finding maximum subgraphs with relatively large vertex connectivity

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

[a] Munitions Directorate, Air Force Research Laboratory, Eglin AFB, FL, USA
[b] Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL, USA
[c] Department of Industrial Engineering, University of Pittsburgh, Pittsburgh, PA, USA

ARTICLE INFO

Article history:
Received 8 June 2013
Accepted 26 May 2014
Available online 4 June 2014

Keywords:
Vertex connectivity
Clique relaxations
Computational complexity
Mixed integer programming
2-Club

ABSTRACT

We consider a clique relaxation model based on the concept of relative vertex connectivity. It extends the classical definition of a k-vertex-connected subgraph by requiring that the minimum number of vertices whose removal results in a disconnected (or a trivial) graph is proportional to the size of this subgraph, rather than fixed at k. Consequently, we further generalize the proposed approach to require vertex-connectivity of a subgraph to be some function f of its size. We discuss connections of the proposed models with other clique relaxation ideas from the literature and demonstrate that our generalized framework, referred to as f-vertex-connectivity, encompasses other known vertex-connectivity-based models, such as s-bundle and k-block. We study related computational complexity issues and show that finding maximum subgraphs with relatively large vertex connectivity is *NP*-hard. An interesting special case that extends the R-robust 2-club model recently introduced in the literature, is also considered. In terms of solution techniques, we first develop general linear mixed integer programming (MIP) formulations. Then we describe an effective exact algorithm that iteratively solves a series of simpler MIPs, along with some enhancements, in order to obtain an optimal solution for the original problem. Finally, we perform computational experiments on several classes of random and real-life networks to demonstrate performance of the developed solution approaches and illustrate some properties of the proposed clique relaxation models.

1. Introduction

Let $G = (V, E)$ be a simple undirected graph with a set of n vertices (nodes) V and a set of edges $E \subset V \times V$. Two vertices v and v' are *connected* if G contains a path between them; graph G is *connected* (*disconnected*) if all its vertices are pairwise connected (there exists a pair of vertices that are not connected). The *vertex connectivity* of a graph G, referred to as $\kappa(G)$, is defined as the minimum number of vertices of G whose removal results in either a disconnected graph or a trivial graph (i.e., consisting of exactly one vertex) (Kirousis, Serna, & Spirakis, 1993; Pattillo, Youssef, & Butenko, 2013). We say that graph G is *k-vertex-connected* if its vertex connectivity is at least k, i.e., $\kappa(G) \geqslant k$. Vertex connectivity and k-vertex connectivity of a given graph can be verified in polynomial time (Galil, 1980). Since vertex connectivity is among the fundamental graph properties, there is a considerable body of work on this topic, see Galil (1980); Kammer and Täubig (2005); Kirousis et al. (1993); Matula (1969); Reif and Spirakis (1985) and references therein.

The *longest distance* (or, equivalently, the length of a *longest shortest path*) between all pairs of vertices in G is referred to as the *diameter* of G, i.e., $diam(G) = \max_{v,v' \in G} d_G(v, v')$, where $d_G(v, v')$ denotes the length of a *shortest path* between v and v' in G. The *density* $\rho(G)$ of graph $G = (V, E)$ is defined as the ratio of the number of edges $|E|$ to the maximum possible number of edges in a graph with $|V|$ vertices, i.e., $\rho(G) = |E|/\binom{|V|}{2}$. Denote by $N_G(v)$ the set of all neighbors of $v \in V$ in G, i.e., $v' \in N_G(v)$ for all $(v, v') \in E$. Then the *degree* of v in G is given by $deg_G(v) = |N_G(v)|$.

A graph $G = (V, E)$ is *complete* if all its pairs of vertices are connected by an edge. For any subset of vertices S, $G[S] = (S, (S \times S) \cap E)$ denotes the subgraph induced by S on G. A *clique* C is a subset of V such that the subgraph $G[C]$ induced by C in G is complete (Luce & Perry, 1949). The *maximum clique problem* is to find a clique of maximum cardinality in G, see, e.g., Bomze, Budinich, Pardalos, and Pelillo (1999). This problem is *NP*-hard (and its decision version is *NP*-complete) (Garey & Johnson, 1979).

A clique is a very intuitive and simple concept of a cohesive subgraph with numerous important applications (Bomze et al., 1999; Butenko & Wilhelm, 2006). Cliques possess a number of "ideal" cohesiveness properties (Pattillo, Youssef, et al., 2013): each vertex is connected to all other vertices, a clique has maximum possible

* Corresponding author. Address: 1048 Benedum Hall, University of Pittsburgh, Pittsburgh, PA 15261, USA. Tel.: +1 412 624 9833; fax: +1 412 624 9831.
E-mail address: droleg@pitt.edu (O.A. Prokopyev).

http://dx.doi.org/10.1016/j.ejor.2014.05.041
0377-2217/© 2014 Elsevier B.V. All rights reserved.

edge density as well as edge and vertex connectivity, the distance between any pair of vertices is one, etc. However, in many practical scenarios cliques are overly restrictive graph structures. Thus, a number of clique relaxation models have been introduced in the literature, see, e.g., Balasundaram, Butenko, Hicks, and Sachdeva (2011); Balasundaram, Butenko, and Trukhanov (2005); Veremyev and Boginski (2012a).

A *unifying taxonomic framework* proposed by Pattillo, Youssef, et al., 2013 demonstrates that all existing clique relaxation models in the literature are based on relaxing some of the *elementary clique-defining properties*, namely, distance, diameter, domination, degree, density and connectivity. These relaxations are further classified into *absolute* and *relative* ones. One illustrative example of an absolute (*diameter-based*) relaxation is an *s-club*, which is defined as a subset $S \subseteq V$ such that the subgraph $G[S]$ induced by S in G has diameter of at most s, i.e., $diam(G[S]) \leqslant s$, where s is a fixed positive integer (Balasundaram et al., 2005; Veremyev & Boginski, 2012a). Clearly, requiring $s = 1$ results in a clique, while relaxing $s \geqslant 2$ defines a subgraph with somewhat less restrictive diameter requirements. The problem of finding maximum s-clubs is known to be *NP*-hard for any fixed $s \geqslant 2$ (Balasundaram et al., 2005).

Another absolute clique relaxation model using vertex connectivity is a *k-block* (Pattillo, Youssef, et al., 2013), which is defined as a subset $S \subseteq V$ such that the subgraph $G[S]$ induced by S in G has vertex connectivity of at least k, i.e., $\kappa(G[S]) \geqslant k$. In contrast to the computationally hard clique relaxation model above, finding a maximum 1-block is a polynomially solvable problem, as it corresponds to finding the largest connected component of a graph. Similarly, maximum 2-connected and 3-connected components can be found in $O(|V| + |E|)$ time (Kammer & Täubig, 2005); furthermore, for any fixed $k > 3$, finding maximum k-connected components can be performed in $O(2^k|V|^3)$ time (Pattillo, Youssef, et al., 2013).

A classical example of a relative (*edge-based*) clique relaxation model is a *γ-quasi-clique* defined as a subset $S \subseteq V$ such that the subgraph $G[S]$ induced by S in G has an edge density of at least γ, i.e., $\rho(G[S]) = |(S \times S) \cap E| / \binom{|S|}{2} \geqslant \gamma$, where $\gamma \in [0,1]$ is a fixed constant parameter (Abello, Resende, & Sudarsky, 2002). Obviously, $\gamma = 1$ corresponds to a clique, while $0 \leqslant \gamma < 1$ defines subgraphs with smaller edge densities. The problem of finding maximum γ-quasi-cliques is known to be *NP*-hard for any fixed $\gamma \in (0,1]$, see Pattillo, Veremyev, Butenko, and Boginski (2013); Pattillo, Youssef, et al. (2013); Uno (2010).

The γ-quasi-clique is, probably, the most well-known relative clique relaxation model with a number of important applications in biological, social, telecommunication and financial areas (Abello, Pardalos, & Resende, 1999; Abello et al., 2002; Matsuda, Ishihara, & Hashimoto, 1999; Mahdavi Pajouh, Miao, & Balasundaram, 2014; Uno, 2010). For example, in graph-theoretical models that are built upon some real-life (e.g., experimental) data, measurement errors and noisy observations often result in missing "links" (i.e., edges). Using cliques for capturing and representing *dense* clusters of closely related (either through cohesiveness or "tightness") functional elements within such networked systems can be impractical and too idealistic since large cliques rarely occur in natural systems. Consequently, techniques based on γ-quasi-clique models can be applied to address such issues. However, Pattillo, Youssef, et al., 2013 pointed out that other relative clique relaxations ideas should also be studied. Specifically, the concept of relative vertex connectivity was identified as an interesting research direction that is worth exploring. In this paper, we further investigate this issue by using the following definition of a clique relaxation model:

Definition 1 (*γ-relative-vertex-connected subgraph*). Given a graph $G = (V, E)$ and a fixed parameter $\gamma \in [0, 1]$, a subgraph $G[S], S \subseteq V$, is called *γ-relative-vertex-connected* (or relative γ-vertex-connected) if the minimum number of vertices, whose removal disconnects $G[S]$ (or results in a trivial subgraph with exactly one vertex), is at least $\gamma(|S| - 1)$.

Parameter γ can be viewed as the minimum *fraction, or percentage,* of vertices that need to be removed ("destroyed") in order to disconnect $G[S]$ or obtain a subgraph with exactly one vertex. Note that if S is a clique in G, then $G[S]$ has maximum possible vertex connectivity, i.e., $\kappa(G[S]) = |S| - 1$. Conversely, $\gamma = 1$ implies that S is a clique in G.

One important observation is that any γ-relative-vertex-connected subgraph is also a γ-quasi-clique; thus, desirable properties of γ-quasi-cliques (i.e., edge-density) are preserved in γ-relative-vertex-connected subgraphs. Indeed, the degree of any vertex in $G[S]$ is at least $\gamma(|S| - 1)$; otherwise, if there exists a vertex with a smaller degree, then deletion of all its neighbors results in a disconnected vertex, which violates the definition of a γ-relative-vertex-connected subgraph. Therefore, the number of edges in $G[S]$ is at least $\gamma(|S| - 1) \cdot |S|/2$ implying that $G[S]$ is also a γ-quasi-clique. However, there exist γ-quasi-cliques that do not induce γ-relative-vertex-connected subgraphs. In particular, a γ-quasi-clique may be a *disconnected graph*, which is often mentioned as the key disadvantage of this relative clique relaxation model.

In this paper we consider the problem of finding a maximum (in terms of cardinality $|S|, S \subseteq V$) subgraph $G[S]$ that is γ-relative-vertex-connected. We refer to the decision version of this problem as the γ-RELATIVE-VERTEX-CONNECTED subgraph problem. In Section 2.1 we show for any fixed $\gamma \in (0, 1]$ the problem remains *NP*-complete (and its optimization version is *NP*-hard). Note that $\gamma = 0$ corresponds to a polynomially solvable case as any graph G is 0-relative-vertex-connected, while $\gamma = 1$ reduces to the classical maximum clique problem.

Next, we further generalize the concept of γ-relative-vertex-connectivity with the following[1]:

Definition 2 (*f-vertex-connected subgraph*). Given a graph $G = (V, E)$ and a function $f(\cdot)$ such that $f : \mathbb{Z}_{>0} \to \mathbb{R}_+$, a subgraph $G[S], S \subseteq V$, is called f-vertex-connected if the minimum number of vertices, whose removal disconnects $G[S]$ (or results in a trivial graph with exactly one vertex) is at least $f(|S|)$.

For a fixed $f(\cdot)$, define the decision version of the f-VERTEX-CONNECTED subgraph problem as follows: given graph $G = (V, E)$ and positive integer k, the question is whether G contains an f-vertex-connected subgraph of size at least k. The optimization version of the f-VERTEX-CONNECTED subgraph problem consists of finding a maximum (in terms of cardinality, $|S|$) f-vertex-connected subgraph $G[S], S \subseteq V$.

First, observe that $f(|S|) = |S| - 1$ corresponds to the CLIQUE problem, while $f(|S|) = \gamma(|S| - 1), \gamma \in (0, 1]$, reduces to the γ-RELATIVE-VERTEX-CONNECTED subgraph problem (see Definition 1). Similarly, if $f(|S|) = k$, where k is a fixed positive integer, then the problem corresponds to finding a maximum k-block, e.g., the largest connected component for $k = 1$. The proposed generalization encompasses another known vertex-connectivity-based model, namely, *s-bundle* (Pattillo, Youssef, et al., 2013), which is defined as $S \subseteq V$ that induces a subgraph with vertex connectivity at least $|S| - s$, i.e., $\kappa(G[S]) \geqslant |S| - s$, where s is a fixed positive integer. Any 1-bundle is a clique and an s-bundle is also an f-vertex-connected subgraph for $f(|S|) = |S| - s$. It is known that finding a maximum s-bundle is also *NP*-hard (Pattillo, Youssef, et al., 2013).

The discussion above highlights the fact that identifying maximum subgraphs remains a difficult combinatorial optimization problem in general graphs as long as the required cohesiveness property is relatively strict (e.g., clique, s-bundle,

[1] We use $\mathbb{Z}_{>0}$ to denote the set of all strictly positive integers.