European Journal of Operational Research 263 (2017) 390–400

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Discrete Optimization

On biconnected and fragile subgraphs of low diameter

Oleksandra Yezerska[a], Foad Mahdavi Pajouh[b], Sergiy Butenko[a,*]

[a] Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX 77843-3131, United States
[b] Department of Management Science and Information Systems, University of Massachusetts Boston, Boston, MA 02125–3393, United States

ARTICLE INFO

Article history:
Received 7 April 2016
Accepted 11 May 2017
Available online 17 May 2017

Keywords:
2-clubs
Biconnected 2-clubs
Fragile 2-clubs
Combinatorial branch-and-bound
Branch-and-cut
Robust network clusters

ABSTRACT

An *s*-club is a subset of vertices inducing a subgraph with a diameter of at most *s*. It is commonly used to characterize network clusters in applications for which easy reachability between group members is of high importance. In this paper, we study two special cases of the 2-club model – a *biconnected 2-club*, and a *fragile* (not biconnected) *2-club*, respectively. We investigate certain properties of both models, propose a combinatorial branch-and-bound algorithm to find a *maximum biconnected 2-club*, and design a polynomial-time algorithm for finding a *maximum fragile 2-club* in a given graph. In addition, we formulate the maximum biconnected 2-club problem as a linear 0–1 program and solve this formulation by a branch-and-cut approach where some nontrivial constraints are applied in a lazy fashion. Finally, numerical results obtained using the proposed algorithms on a test-bed of randomly generated instances and real-life graphs are also provided.

1. Introduction

Cluster detection is one of the prominent problems in network analysis. It allows to study network properties and structure, reveal hidden patterns, and understand the functions of the underlying systems. One of the earliest mathematical formalizations of a cluster was clique (Luce & Perry, 1949). In graph-theoretical terms, it is a subset of pairwise adjacent vertices. A clique is considered the "perfect" cluster as it possesses all the desirable properties of a cluster, such as highest vertex degree (familiarity), highest number of edges (connectivity, robustness), and shortest pairwise distances (reachability).

However, such structure is too restrictive and most real-life networks can be better analyzed by detecting graph-theoretical clique relaxations, clusters where some direct connections between vertices do not exist. These clique relaxation models are defined by relaxing one or some of the clique properties. Different types of these relaxation models have been proposed in the literature and have found many applications in various domains (Pattillo, Youssef, & Butenko, 2013).

In this work, we focus on a distance-based relaxation of a clique, an *s*-club. Given a graph $G = (V, E)$, let the open neighborhood of a vertex i, which is the set of all vertices adjacent to i in G, be denoted by $N_G(i)$. The set $N_G(i) \cup \{i\}$ is referred to as the closed neighborhood of i in G and is denoted by $N_G[i]$. The distance $d_G(u, v)$ between vertices u and v in G is the length of a shortest path between them. For any set $S \subseteq V$, the subgraph induced by S is defined as $G[S] = (S, E \cap (S \times S))$. An *s*-club S (Mokken, 1979) is a subset of vertices that induces a subgraph $G[S]$ with $diam(G[S]) \leq s$, where $diam(G[S]) = \max_{u,v \in S} d_{G[S]}(u, v)$. Similar to an *s*-club, but less restrictive, is a structure called an *s*-clique. Formally, an *s*-clique S is a subset of vertices such that $\max_{u,v \in S} d_G(u, v) \leq s$. The structural design of an *s*-club is very common and important for various transportation and telecommunication networks, where fast and easy access to all entities is crucial, but designing a clique is too expensive.

Even though *s*-clubs provide easy reachability, they cannot always guarantee high connectivity and robustness. For example, an extreme case of a 2-club – a *star*, which is a graph with one "hub" vertex connected to the rest of the vertices that are not directly connected to each other, is apt to complete disconnection due to the failure of the hub (Pattillo, Youssef, & Butenko, 2013). Therefore, in the situations where failures are anticipated, it is important to design or detect clusters not only of small diameter, but also with additional constraints on their connectivity. To address this concern, several approaches based on the notions of vertex connectivity, heredity, and robustness can be used.

Recall that the *(vertex) connectivity* $\kappa(G)$ of a graph G is determined by the smallest number of vertices that need to be removed from the graph to make it disconnected or trivial (i.e., a one-vertex graph). In particular, the vertex connectivity of a complete graph on n vertices is $n - 1$. The vertex connectivity also corresponds to the smallest number of vertex-disjoint paths between a pair of

* Corresponding author.
E-mail addresses: yaleksa@tamu.edu (O. Yezerska), foad.mahdavi@umb.edu (F. Mahdavi Pajouh), butenko@tamu.edu (S. Butenko).

http://dx.doi.org/10.1016/j.ejor.2017.05.020
0377-2217/

vertices. A graph G is said to be *k-connected* if $\kappa(G) \geq k$. An s-club S is called

- a *k-connected s-club* if $\kappa(G[S]) \geq k$;
- a *k-hereditary s-club* (Pattillo, Youssef, & Butenko, 2013) if $diam(G[S \backslash S']) \leq s$ for any $S' \subset S$ such that $|S'| \leq k$;
- an *R-robust s-club* (Veremyev & Boginski, 2012) if there are at least R vertex-disjoint paths of length at most s between all pairs of vertices in $G[S]$.

To highlight the differences between the three models, consider the complete bipartite graph $K_{3,3}$. The set of vertices of this graph is a 3-connected 2-club since we need to remove at least 3 vertices to disconnect the residual graph. It is also a 2-hereditary 2-club (we can remove up to two vertices and still preserve the diameter of two) and a 1-robust 2-club (there is only one path of length no more than two between pairs of vertices in different parts). Hence, k-connectivity is the least restrictive of the three requirements used to enforce connectivity of an s-club.

To the best of our knowledge, only one of the three models above, the R-robust s-club, has been studied from an optimization perspective (Veremyev & Boginski, 2012). In this paper, we start investigating the k-connected s-club model by considering a special case of $k = s = 2$, that is, *biconnected* 2-clubs. In addition, we study the 2-clubs that are not biconnected, which we call *fragile*. Next, we formally define these two structures.

Definition 1. Given a simple graph $G = (V, E)$, a set $S \subseteq V$ is a biconnected 2-club if $diam(G[S]) \leq 2$ and $\kappa(G[S]) \geq 2$.

Definition 2. Given a simple graph $G = (V, E)$, a set $S \subseteq V$ is a fragile 2-club if $diam(G[S]) \leq 2$ and $\kappa(G[S]) = 1$.

The maximum s-club problem, which is to find an s-club of the largest cardinality in the graph, is NP-hard for any fixed s (Bourjolly, Laporte, & Pesant, 2002), even when restricted to graphs of diameter $s + 1$ (Balasundaram, Butenko, & Trukhanov, 2005). It is also NP-hard to test s-club maximality for any fixed integer $s \geq 2$ (Mahdavi Pajouh & Balasundaram, 2012). We show that a maximum fragile 2-club can be detected in polynomial time, while finding a maximum biconnected 2-club is still hard.

The maximum s-club problem has been approached with polyhedral methods (Almeida & Carvalho, 2012; Balasundaram, Butenko, & Trukhanov, 2005; Bourjolly, Laporte, & Pesant, 2002; Carvalho & Almeida, 2011; Mahdavi Pajouh, Balasundaram, & Hicks, 2016; Veremyev & Boginski, 2012), heuristics (Bourjolly, Laporte, & Pesant, 2000; Kahruman-Anderoglu, Buchanan, Butenko, & Prokopyev, 2016; Shahinpour & Butenko, 2013a) and exact algorithms (Bourjolly, Laporte, & Pesant, 2002; Hartung, Komusiewicz, & Nichterlein, 2015; Mahdavi Pajouh & Balasundaram, 2012; Moradi & Balasundaram, 2015); see (Shahinpour & Butenko, 2013b) for a recent survey. Bourjolly, Laporte, and Pesant (2000) proposed three constructive heuristic methodologies for the maximum s-club problem. CONSTELLATION and DROP heuristics have also been utilized as a lower bound estimator within the exact branch-and-bound frameworks in Bourjolly, Laporte, and Pesant (2002); Mahdavi Pajouh and Balasundaram (2012). A heuristic based on the variable neighborhood search (Mladenović & Hansen, 1997) was developed by Shahinpour and Butenko (2013a). The authors used it as a lower bound estimator within the combinatorial branch-and-bound algorithm developed by Mahdavi Pajouh and Balasundaram (2012), as well as part of a hybrid algorithm. Recently, Moradi and Balasundaram (2015) proposed a lazy-fashioned branch-and-cut algorithm that utilizes an s-clique formulation and cuts off all incumbent solutions that are not s-clubs.

The maximum biconnected 2-club problem (i.e., to find a largest biconnected 2-club in a graph) is NP-hard, as discussed later in this article. However, we present a necessary and sufficient condition for a subset S of vertices to form a fragile 2-club, which can be checked in linear time ($O(|E(G[S])|)$, where $E(G[S])$ is the edge set of the subgraph induced by S). Using this condition and additional related theoretical results, we adapt the combinatorial branch-and-bound framework for finding a maximum s-club developed by Mahdavi Pajouh and Balasundaram (2012) to detect a maximum biconnected 2-club. Intractability of the maximum biconnected 2-club problem further motivates developing integer programming techniques to solve this problem. To this aim, we will formulate this problem as a linear 0–1 program and solve this formulation by a lazy-fashioned branch-and-cut approach. We will also show that the maximum fragile 2-club problem is polynomial-time solvable and design the corresponding solution procedure.

The remainder of this paper is organized as follows. Section 2 discusses the computational complexity of the maximum biconnected 2-club problem and presents some structural properties of biconnected and fragile 2-clubs. These properties are used in developing algorithms for the maximum fragile 2-club and the maximum biconnected 2-club problems in Sections 3 and 4, respectively. In particular, a polynomial-time algorithm for finding a maximum fragile 2-club in a graph is given in Section 3. Furthermore, a combinatorial branch-and-bound algorithm and a lazy-fashioned branch-and-cut approach based on a linear 0–1 programming formulation for the maximum biconnected 2-club problem are presented in Section 4. Results of the computational experiments testing the performance of the proposed algorithms on a testbed of randomly generated and real-life instances are reported in Section 5. Concluding remarks and directions for future research are summarized in Section 6.

2. Computational complexity and structural properties

In this section, we address the computational complexity of the maximum biconnected 2-club problem and show that the decision version of this problem is NP-complete. We also further study the biconnectivity property in 2-clubs and present some characterizations of this property. The theoretical results developed here will be used to propose exact algorithms for finding a largest biconnected and a largest fragile 2-club in a graph.

2.1. Computational complexity of detecting a maximum biconnected 2-club

The decision version of the maximum biconnected 2-club is given by a graph $G = (V, E)$, a positive integer c, and a question that asks if there exists a biconnected 2-club of size at least c in graph G. The following theorem addresses the computational complexity of this decision problem and shows its intractability.

Proposition 1. *The decision version of the maximum biconnected 2-club problem is NP-complete.*

Proof. The same construction as used by Balasundaram, Butenko, and Trukhanov (2005) to show the NP-completeness of the maximum s-club problem for graphs of diameter at least s can be utilized to prove this statement. According to Balasundaram, Butenko, and Trukhanov (2005) and assuming $s = 2$, given an instance of the maximum clique problem $< G, k >$, we construct $G' = (V', E')$, such that $V' = V \cup E$ and $E' = E_1 \cup E_2$, where $E_1 = \{(v, e) : v \in V, e \in E, v \text{ is incident to } e\}, E_2 = \{(e_1, e_2) : e_1, e_2 \in E, e_1 \neq e_2\}$. It can be shown that G has a clique of size at least k if and only if G' has a biconnected 2-club of size at least $k + |E|$. □

2.2. Characterizing biconnectivity in 2-clubs

According to Menger's Theorem (Diestel, 2005; Menger, 1927), the local connectivity between two vertices is determined by