Discrete Optimization 9 (2012) 84–97

Contents lists available at SciVerse ScienceDirect

Discrete Optimization

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

On inclusionwise maximal and maximum cardinality k-clubs in graphs

F. Mahdavi Pajouh, B. Balasundaram *

322 Engineering North, School of Industrial Engineering & Management, Oklahoma State University, Stillwater, OK 74078, United States

ARTICLE INFO

Article history:
Received 14 February 2010
Received in revised form 3 February 2012
Accepted 4 February 2012
Available online 24 February 2012

Keywords:
Clique
k-club
Exact combinatorial algorithms
Graph-based data mining
Social network analysis

ABSTRACT

A k-club is a distance-based graph-theoretic generalization of a clique, originally introduced to model cohesive social subgroups in social network analysis. The k-clubs represent low diameter clusters in graphs and are appropriate for various graph-based data mining applications. Unlike cliques, the k-club model is nonhereditary, meaning every subset of a k-club is not necessarily a k-club. In this article, we settle an open problem establishing the intractability of testing inclusion-wise maximality of k-clubs. This result is in contrast to polynomial-time verifiability of maximal cliques, and is a direct consequence of its nonhereditary nature. We also identify a class of graphs for which this problem is polynomial-time solvable. We propose a distance coloring based upper-bounding scheme and a bounded enumeration based lower-bounding routine and employ them in a combinatorial branch-and-bound algorithm for finding maximum cardinality k-clubs. Computational results from using the proposed algorithms on 200-vertex graphs are also provided.

1. Introduction

Given a simple, undirected graph $G = (V, E)$ of order $n = |V|$ and size $m = |E|$, the subgraph induced by $S \subseteq V$ is denoted by $G[S] = (S, E \bigcap (S \times S))$. A subset $S \subseteq V$ is called a clique if $G[S]$ is complete and it is called an independent set if $G[S]$ is edgeless. The maximum clique problem is to find a clique of maximum cardinality, referred to as the clique number of G, denoted by $\omega(G)$. An early application of cliques was in social network analysis (SNA) where they were used to model "tightly knit" social subgroups referred to as cohesive subgroups [1]. Since, there have been applications of the clique model in diverse fields such as coding theory, pattern recognition, fault diagnosis, bioinformatics, computational finance, and telecommunication in addition to SNA [2–5]. However, cliques also have certain drawbacks that motivated the development of graph-theoretic clique relaxations in SNA. Cliques require all possible edges to exist between a group of vertices for the group to be considered cohesive. This requirement was found to be overly restrictive leading to relaxations based on degree [6], distance [7–9], and edge density [10,11,5]. The clique provides the best possible guarantees for three desirable properties of a tight cluster, namely, high degree, short pairwise distances, high vertex and edge connectivity. In applications where cohesiveness is well described by any one of the three properties alone, the clique model becomes unduly restrictive. This article studies a distance-based generalization of cliques known as *k-clubs* which model low diameter clusters in graphs. The diameter of $G = (V, E)$ is given by $\text{diam}(G) = \max_{u,v \in V} d_G(u, v)$, where $d_G(u, v)$ is the length of a shortest path (in number of edges) between vertices u and v in G. The diameter of G is said to be infinite if there does not exist a path between a pair of vertices.

Definition 1. A k-clique is a subset $S \subseteq V$ for which $d_G(u, v) \leq k$ for all $u, v \in S$.

* Corresponding author. Tel.: +1 405 744 6055.
E-mail addresses: mahdavi@okstate.edu (F. Mahdavi Pajouh), baski@okstate.edu (B. Balasundaram).

1572-5286/$ – see front matter
doi:10.1016/j.disopt.2012.02.002

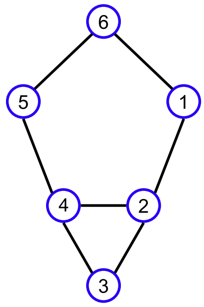

Fig. 1. 2-clique vs. 2-club.

Definition 2. A k-club is a subset $S \subseteq V$ for which $d_{G[S]}(u, v) \leq k$ for all $u, v \in S$. Equivalently, S is a k-club if $\text{diam}(G[S]) \leq k$.

Both k-cliques and k-clubs describe a clique when $k = 1$ and are relaxations when $k \geq 2$. Note that every k-club in G is also a k-clique in G, but the converse is not necessarily true when $k \geq 2$. In a k-clique S, there can exist two vertices $u, v \in S$ such that $d_G(u, v) \leq k$, but $d_{G[S]}(u, v) > k$. For the graph in Fig. 1, vertices $\{1, 2, 3, 4, 5\}$ form a 2-clique which is not a 2-club.

The k-clique model was introduced in SNA by Luce [7] as a distance-based relaxation of cliques. But the definition allows for a k-clique to utilize vertices outside the k-clique in the shortest paths. Since this was not desirable in a cohesive subgroup, the k-club model was introduced to bound pairwise distances in the induced graph instead of the original graph [8,9]. We denote the cardinality of a maximum k-clique by $\widetilde{\omega}_k(G)$, referred to as the k-clique number of G. The k-club number of a graph is the cardinality of a largest k-club in that graph which is denoted by $\overline{\omega}_k(G)$. Clearly, $\omega(G) \leq \overline{\omega}_k(G) \leq \widetilde{\omega}_k(G)$ for every positive integer k. Further, every k-club (k-clique) is also a $k + 1$-club ($k + 1$-clique) by definition. Note that k-cliques are equivalent to cliques through a simple observation concerning power graphs. Given a graph $G = (V, E)$, the k-th power of G is denoted by $G^k = (V, E^k)$ where $E^k = \{(u, v) : d_G(u, v) \leq k\}$. So $S \subseteq V$ is a k-clique in G if and only if S is a clique in G^k, and $\widetilde{\omega}_k(G) = \omega(G^k)$. Given the modeling drawback of k-cliques and their equivalence to cliques in power graphs, they have received much less attention in the literature.

1.1. Previous work and our contributions

The problem of finding a maximum clique is NP-hard [12] and it is NP-hard to approximate within any factor $n^{1-\epsilon}$ for any $\epsilon > 0$ [13]. Further, finding a clique parameterized by solution size is not known to be fixed-parameter tractable, and it is in fact a basic $W[1]$-hard problem [14]. Although k-clubs generalize cliques, not all these complexity results extend to k-clubs. Bourjolly et al. [15] established the NP-hardness of the maximum k-clique and k-club problems and developed an exact branch-and-bound algorithm for the latter. Balasundaram et al. [16] independently established the NP-hardness of the problems for every fixed positive integer k, and in addition showed that the problems remain NP-hard even when restricted to graphs of fixed diameter. Hence, the maximum k-clique and k-club problems are known to be NP-hard for every fixed k even on graphs of diameter $k+1$. Bourjolly et al. also studied greedy construction and elimination heuristics for the problem in [17]. Pertinently, Butenko and Prokopyev [18] have shown that recognizing whether there is a gap between $\bar{\omega}_k(G)$ and $\bar{\omega}_l(G)$ for $k \neq l$ is NP-hard. This allowed them to show that for $k \geq 2$, there does not exist a polynomial time algorithm to find a k-club in G of size strictly larger than $\Delta(G) + 1$ on graphs with $\bar{\omega}_k(G) > \Delta(G) + 1$, unless $P = NP$. Note that $\Delta(G)$ is the maximum vertex degree in G, and a vertex of maximum degree along with all its neighbors forms a k-club in G for any $k \geq 2$.

The maximum k-club problem for fixed $k \geq 2$ was shown to be inapproximable within a factor of $n^{\frac{1}{3}-\epsilon}$ for any $\epsilon > 0$, if $NP \neq ZPP$ [19], which has been strengthened to $n^{\frac{1}{2}-\epsilon}$ recently under the assumption $P \neq NP$ [20]. Algorithms for the maximum k-club problem with approximation factors of $n^{\frac{1}{2}}$ for even k and $n^{\frac{2}{3}}$ for odd k have also been proposed recently [20]. The problem is also fixed-parameter tractable when parameterized by solution size as demonstrated in [21] where a $O((s-2)^s s! s^3 n + mn)$ time algorithm is presented to find a k-club of size s in G. The approximability and fixed-parameter tractability of the problem are in contrast to the results obtained for cliques.

As noted in [16], complete bipartite graphs are edge-critical 2-clubs. Therefore, finding a maximum 2-club in a bipartite graph amounts to finding a largest order biclique, which in trees corresponds to the vertex of maximum degree and its neighbors. Recently, it was shown that the maximum 2-club problem can be solved on bipartite graphs in $O(n^5)$, and the maximum k-club problem can be solved on trees and interval graphs in $O(nk^2)$ and $O(n^2)$, respectively [22]. Furthermore, it is also shown to be polynomial-time solvable on graphs with bounded tree-width or clique-width [22].

The 2-club polytope which is the convex hull of incidence vectors of 2-clubs in a graph, is studied in [16] and a family of facet defining inequalities are identified. This line of research is also explored in [23] where additional strong valid inequalities for the 2-club polytope are identified. While the k-club problem has a compact binary integer programming