On k-club and k-clique numbers in graphs

Sergiy Butenko

Department of Industrial and Systems Engineering, Texas A&M University
236E Zachry Engineering Center, College Station, TX 77843-3131, USA

Oleg Prokopyev

Department of Industrial Engineering, University of Pittsburgh
1048 Benedum Hall, Pittsburgh, PA 15261, USA

Abstract

For a simple undirected graph and a given positive integer k, a k-club is a subset of vertices that induces a subgraph of diameter at most k, and the k-club number $\bar{\omega}_k(G)$ is the cardinality of a largest k-club in G. In this paper we first prove that for given positive integers k and l, $k \neq l$, the problem of recognizing whether there is a gap between $\bar{\omega}_k$ and $\bar{\omega}_l$ is NP-hard. Then we use this result to show that for $k \geq 2$, unless $P = NP$, one cannot design a polynomial-time algorithm that would detect a k-club of size $> \Delta(G) + 1$ in any graph G with $\bar{\omega}_k(G) > \Delta(G) + 1$, where $\Delta(G)$ denotes the maximum degree of a vertex in G. The same results hold for the maximum k-clique problem as well.

Key words: k-club, k-clique, graphs, discrete optimization, heuristics

---

1 Introduction

Consider a simple undirected graph $G = (V, E)$ with the set of vertices V and the set of edges E. For a subset W of vertices, we will denote by $G[W]$ the subgraph induced by W, *i.e.*, $G[W] = (W, E \cap (W \times W))$. For a vertex $i \in V$, $N(i) = \{j \in V : (i, j) \in E\}$ is the neighborhood of i and $\deg(i) = |N(i)|$ is the degree of vertex i. The largest degree of a vertex in G is denoted by $\Delta(G)$. Let $\mathrm{diam}(G)$ denote the diameter of graph G, $\mathrm{diam}(G) = \max\limits_{i,j \in V} \mathrm{dist}(i, j)$, where

---

Email addresses: `butenko@tamu.edu` (Sergiy Butenko), `prokopyev@engr.pitt.edu` (Oleg Prokopyev).

$\mathrm{dist}(i,j)$ is the length of a shortest path between i and j in G (measured in the number of edges). Given a positive integer k, a subset of vertices $C \subseteq V$ is called a k-clique if for any $i,j \in C$: $\mathrm{dist}(i,j) \leq k$. A subset of vertices $C \subseteq V$ is called a k-club if $\mathrm{diam}(G[C]) \leq k$. Note that any k-club in G is also a k-clique, however the converse is not true, since all shortest paths between a pair of vertices in a k-clique C can contain vertices outside C. All results mentioned or proved below in this paper for k-clubs are also true for k-cliques and can be stated by simply replacing the term "k-club" with "k-clique". Therefore, for the sake of simplicity, we will use the term "k-club" throughout the paper to describe such results.

The maximum k-club problem, which is to find the largest k-club in a graph, is NP-hard for any fixed k, even if restricted to the graphs of diameter at most $k+1$ [1]. The reader is referred to the classical book [4] for the complexity-theoretic background and to a recent paper [1] for a discussion on history and applications of the notions of k-club and k-clique. Denote by $\bar{\omega}_k(G)$ the k-club number of G, which is the number of vertices in a largest k-club in G. Clearly, for $l < k$ we have

$$\bar{\omega}_l(G) \leq \bar{\omega}_k(G).$$

Note that for $l = 1$ an l-club is a clique and $\bar{\omega}_1(G) \equiv \omega(G)$, where $\omega(G)$ is the clique number of G. In this case, for $k \geq 2$, we have an obvious inequality:

$$\omega(G) \leq \Delta(G) + 1 \leq \bar{\omega}_k(G),$$

in which, similarly to the famous Sandwich Theorem [6], a polynomially-computable value is sandwiched between two values that are NP-hard to compute. In an interesting related paper [2], Busygin and Pasechnik have shown that it is NP-hard to check whether $\bar{\chi}(G) - \alpha(G) = 0$, where $\bar{\chi}(G)$ denotes the cardinality of a minimum clique partitioning in G and $\alpha(G)$ is the independence number of G. This is in contrast to the fact that checking whether a graph G is perfect (i.e., for every subset S of vertices we have $\bar{\chi}(G[S]) = \alpha(G[S])$) is polynomial [3]. The Busygin-Pasechnik result implies that any polynomially-computable parameter that lies between $\alpha(G)$ and $\bar{\chi}(G)$ will provide the "best" upper bound for the independence number in the sense that no other polynomially computable bound can be provably better for all graphs where this bound can be improved. In particular, the Lovász theta is one such polynomially computable bound [5].

The remainder of this paper is organized as follows. In Section 2 we will show that it is NP-hard to recognize whether $\bar{\omega}_l(G) = \bar{\omega}_k(G)$. In Section 3 we will discuss the implications this result has on potential performance of heuristics for the maximum k-club problem with $k \geq 2$. Finally, Section 4 concludes the paper.

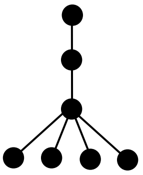

Fig. 1. A sample broom graph B_4^3.

2 Complexity of gap recognition

Theorem 2.1 *Let positive integer constants* k *and* l*,* $l < k$ *be given. The problem of checking whether* $\bar{\omega}_l(G) = \bar{\omega}_k(G)$ *is* NP*-hard.*

Proof. Assume that there is a polynomial time algorithm $\mathcal{A}_{kl}(G)$ that, given a graph G , correctly answers the question "Is $\bar{\omega}_l(G) = \bar{\omega}_k(G)$?" with either "yes" or "no". Next we analyze the two possible cases.

(i) The answer given by $\mathcal{A}_{kl}(G)$ is "no". Then we run the following polynomial-time algorithm to compute $\bar{\omega}_k(G)$:

```
0. $G' = G,\ i = 1$;
1. while the answer of $\mathcal{A}_{kl}(G)$ is "no" repeat
       $G' = G' \cup C_i$, where $C_i$ is the clique on $i$ vertices;
       $i = i + 1$.
2. return $i$.
```

Note that starting from $i = \bar{\omega}_l(G)$ each next iteration of this algorithm will increase $\bar{\omega}_l(G')$ by 1 without changing $\bar{\omega}_k(G') = \bar{\omega}_k(G)$. The algorithm will terminate when $i = \bar{\omega}_l(G') = \bar{\omega}_k(G') = \bar{\omega}_k(G)$.

(ii) The answer given by $\mathcal{A}_{kl}(G)$ is "yes". Again, we want to design a polynomial time algorithm for computing $\bar{\omega}_k(G)$. Since k is a constant, we can check if $\bar{\omega}_k(G) < k$ by examining all subsets of size $< k$ in $O(n^k)$ time. Thus, we are interested only in the case where $\bar{\omega}_k(G) \geq k$. Denote by B_s^h the (h, s)-*broom graph*, which consists of a path of h vertices and s more vertices connected to one of the endpoints of this path. As an example, Figure 1 shows the broom graph B_4^3. Obviously, the diameter of B_s^k is equal to k if $k \geq 1$, therefore, for any $l < k$, B_s^k is a k-club, but not an l-club. We can use the broom graphs to compute $\bar{\omega}_k(G)$ as follows.

```
0. $G' = G,\ i = 1$;
1. while the answer of $\mathcal{A}_{kl}(G)$ is "yes" repeat
       $G' = G' \cup B_i^k$;
       $i = i + 1$.
2. return $i + k - 1$.
```

Hence, we have shown that if $\mathcal{A}_{kl}(G)$ was a polynomial-time algorithm, then we would be able to compute $\bar{\omega}_k(G)$ in polynomial time. The result follows from NP-hardness of computing $\bar{\omega}_k(G)$ for any fixed k. $\square$

3 On heuristics for the maximum k-club problem

Next we discuss the implications of the above complexity result for the case when $l = 1$ and $k \geq 2$. Recall that $\omega(G) \leq \Delta(G)+1 \leq \bar{\omega}_k(G)$ and observe that we can easily check whether $\omega(G) = \Delta(G)+1$ by checking, for each maximum degree vertex in G, if its neighbors form a clique. Hence, it is NP-hard to check whether $\bar{\omega}_k(G) = \Delta(G) + 1$. This implies that, unless $P = NP$, one cannot design a polynomial-time heuristic for the maximum k-club problem, which is provably better than the trivial approach consisting in picking a maximum degree vertex and all its neighbors as the output k-club. Indeed, existence of a polynomial-time algorithm that finds a k-club of size greater than $\Delta(G) + 1$ whenever $\bar{\omega}_k(G) > \Delta(G) + 1$ would imply that one can check in polynomial time whether $\bar{\omega}_k(G) = \Delta(G) + 1$. Thus, we obtain the following corollary.

Corollary 1 *Let k be a fixed integer, $k \geq 2$. Unless $P = NP$, there cannot be a polynomial time algorithm that finds a k-club of size greater than $\Delta(G) + 1$ whenever such a k-club exists in the graph.*

Assume that one needs to solve the maximum k-club problem in G for some fixed large k. If one is given a polynomially computable parameter $v_{l,k}(G)$ such that $\bar{\omega}_l(G) \leq v_{l,k}(G) \leq \bar{\omega}_k(G)$, it seems natural to expect a higher value of $v_{l,k}(G)$ to correspond to a higher the value of l. However, since for $l \geq 2$ we have

$$\omega(G) \leq \Delta(G) + 1 \leq \bar{\omega}_l(G) \leq v_{l,k}(G) \leq \bar{\omega}_k(G),$$

the above corollary implies that we cannot expect that $v_{l,k}$ will dominate $\Delta(G) + 1$, no matter how high the value of $l < k$ is.

In a special case when $k = 2l$, we can use $v_{l,k}(G) = \Delta(G^l) + 1$, where G^l is the l-th power of graph G, which is defined on the same set of vertices as G with edges connecting all pairs of vertices that are distance at most p from each other. It is easy to see that $\bar{\omega}_l \leq \Delta(G^l)+1 \leq \bar{\omega}_{2l}$ and $\bar{\omega}_l = \Delta(G^l)+1$ if and only if the neighbors of one of the vertices of degree $\Delta(G^l) + 1$ in G^l form a clique in G^l (which can be easily checked). Therefore, checking if $\Delta(G^l) + 1 = \bar{\omega}_{2l}$ is NP-hard. Note that for two positive integers p and r, such that $p > r > 1$, we have $\Delta(G) \leq \Delta(G^r) \leq \Delta(G^p)$ and

$$\omega(G) \leq \Delta(G) + 1 \leq \bar{\omega}_r(G) \leq \Delta(G^r) + 1 \leq \Delta(G^p) + 1 \leq \bar{\omega}_{2p}(G).$$

Since $\Delta(G^r) < \Delta(G^p)$ implies $\bar{\omega}_r(G) < \bar{\omega}_{2p}(G)$, the problem of checking whether $\bar{\omega}_r(G) = \bar{\omega}_{2p}(G)$ remains NP-hard even if restricted to graphs with

$\Delta(G^r) = \Delta(G^p)$.

4 Conclusion

Corollary 1 provided in the last section can be used as a reasonable theoretical justification of using a simple heuristic (based on finding a maximum-degree vertex) for the maximum k-club problem with fixed $k \geq 2$. It may also be viewed as an additional evidence of the problem's practical intractability. On the other hand, this should not prevent the practitioners from designing more sophisticated approaches for the maximum k-club problem, since the above result describes just the worst-case behavior of the heuristics. Simple greedy heuristics are often used to solve large-scale instances of NP-hard problems in practice, and similar complexity results for other problems, such as the maximum independent set (clique) or graph coloring, would be a good way to explain the choice of the approach from theoretical perspective. We are not aware of any other results of this type in the literature.

Acknowledgements

The research of the first author was partially supported by grants from CRDF and NSF.

References

[1] B. Balasundaram, S. Butenko, and S. Trukhanov. Novel approaches for analyzing biological networks. *Journal of Combinatorial Optimization*, 10:23–39, 2005.

[2] S. Busygin and D.V. Pasechnik. On NP-hardness of the clique partition – independence number gap recognition and related problems. *Discrete Mathematics*, 304:460–463, 2006.

[3] M. Chudnovsky, G. Cornuéjols, X. Liu, P. Seymour, and K. Vuskovic. Recognizing berge graphs. *Combinatorica*, 25:143–186, 2005.

[4] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-completeness.* W.H. Freeman and Company, New York, 1979.

[5] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization.* Springer-Verlag, Berlin, 2nd edition, 1993.

[6] D. E. Knuth. The sandwich theorem. *Electronic J. Combinatorics*, 1:1–48, 1994.