Information and Computation 256 (2017) 160–173

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

Finding connected k-subgraphs with high density ☆,☆☆

Xujin Chen [a], Xiaodong Hu [a], Changjun Wang [b,*]

[a] *Institute of Applied Mathematics, AMSS, Chinese Academy of Sciences, Beijing 100190, China*
[b] *Beijing Institute for Scientific and Engineering Computing, Beijing University of Technology, Beijing 100124, China*

ARTICLE INFO

Article history:
Received 30 August 2015
Available online 10 July 2017

Keywords:
Densest k-subgraphs
Heaviest k-subgraphs
Connectivity
Approximation algorithms
Interval graphs

ABSTRACT

Given an edge-weighted connected graph G on n vertices and a positive integer $k \le n$, a subgraph of G on k vertices is called a k-subgraph in G. We design combinatorial approximation algorithms for finding a connected k-subgraph in G such that its weighted density is at least a factor $\Omega(\max\{1/k, k^2/n^2\})$ of the maximum weighted density among all k-subgraph in G (which are not necessarily connected), where $\max\{1/k, k^2/n^2\} \ge n^{-2/3}$ implies an $O(n^{2/3})$-approximation ratio. We obtain improved $O(n^{2/5})$-approximation for unit weights. These particularly provide the first non-trivial approximations for the heaviest/densest connected k-subgraph problem on general graphs. We also give $O(\sqrt{n}\log n)$-approximation for the problem on general weighted interval graphs.

1. Introduction

Let $G = (V, E)$ be a connected simple undirected graph with n vertices, m edges, and nonnegative edge weights. The (*weighted*) *density* of G is defined as its average (weighted) degree. Let $k \le n$ be a positive integer. A subgraph of G is called a *k-subgraph* if it has exactly k vertices. A k-subgraph of G with the maximum density (weighted density) among all k-subgraphs of G is referred to as a *densest* (*heaviest*) k-subgraph of G. The *densest k-subgraph problem* (DkSP) is to find a densest k-subgraph of G. If the k-subgraph requires to be connected, then the problem is referred as to the *densest connected k-subgraph problem* (DCkSP). Both DkSP and DCkSP have their weighted generalizations, the *heaviest k-subgraph problem* (HkSP) and the *heaviest connected k-subgraph problem* (HCkSP), which ask for a heaviest k-subgraph of G and a heaviest connected k-subgraph of G, respectively. Identifying k-subgraphs with high (weighted) densities is a useful primitive, which arises in diverse applications – from social networks, to protein interaction graphs, to the world wide web, etc. While dense/heavy subgraphs can give valuable information about interactions in these networks, the additional connectivity requirement turns out to be natural in various scenarios. One of typical examples is searching for a large community.

It is worth pointing out that HCkSP is a generalization of HkSP. To see it, we note that each HkSP instance on graph G corresponds to an equivalent HkSP on the complete graph G' which obtained from G by adding zero-weight edges between pairs of nonadjacent vertices.

Related work. An easy reduction from the maximum clique problem shows that DkSP, DCkSP and their weighted generalizations are all NP-hard in general. The NP-hardness remains even for some very restricted graph classes such as (unweighted)

☆ Research supported in part by NNSF of China under Grant No. 11531014 and 11601022, and by the Open Fund of Beijing Institute for Scientific and Engineering Computing.
☆☆ A preliminary version of this article was included in the TAMC 2015 conference proceedings.
* Corresponding author.
E-mail address: wcj@bjut.edu.cn (C. Wang).

http://dx.doi.org/10.1016/j.ic.2017.07.003
0890-5401/© 2017 Elsevier Inc. All rights reserved.

chordal graphs, triangle-free graphs, comparability graphs [9], and bipartite graphs of maximum degree three [14]. Moreover, HkSP and HCkSP remain NP-hard in the metric case, i.e., the edge weights of complete graphs satisfy the triangle inequality [29], and the binary weighted case for cographs and split graphs [9]. The complexity status of DkSP on planar graphs and interval graphs has been open for three decades [9], while the NP-hardness of DCkSP on planar graphs has been established by reducing from the connected vertex cover problem [18].

Most literature on finding dense subgraphs focus on the versions without requiring subgraphs to be connected. For DkSP and its generalization HkSP, narrowing the large gap between the lower and upper bounds on the approximability is an important challenge. On the negative side, the decision problem version of DkSP, in which one is asked if there is a k-subgraph with more than h edges, is NP-complete even if h is restricted by $h \le k^{1+\varepsilon}$ [4]. Feige [11] showed that computing a $(1+\varepsilon)$-approximation for DkSP is at least as hard as refuting random 3-SAT clauses for some $\varepsilon > 0$. Khot [19] showed that there does not exist any polynomial time approximation scheme (PTAS) for DkSP assuming NP does not have randomized algorithms that run in sub-exponential time. Recently, constant factor approximations in polynomial time for DkSP have been ruled out by Raghavendra and Steurer [28] under Unique Games with Small Set Expansion conjecture, and by Alon et al. [1] under certain "average case" hardness assumptions. On the positive side, considerable efforts have been devoted to finding good quality approximations for HkSP. Improving the $O(n^{0.3885})$-approximation of Kortsarz and Peleg [21], Feige et al. [13] proposed a combinatorial algorithm with approximation ratio $O(n^{\delta})$ for some $\delta < 1/3$. The latest algorithm of Bhaskara et al. [6] provides an $O(n^{1/4+\varepsilon})$-approximation in $n^{O(1/\varepsilon)}$ time. If allowed to run for $n^{O(\log n)}$ time, their algorithm guarantees an approximation ratio of $O(n^{1/4})$. The $O(n/k)$-approximation algorithm by Asahiro et al. [5] is remarkable for its simple greedy removal method. Linear and semidefinite programming (SDP) relaxation approaches have been adopted in [12,16,30] to design randomized rounding algorithms, where Feige and Langberg [12] obtained an approximation ratio somewhat better than n/k, while the algorithms of Srivastav and Wolf [30] and Han et al. [16] outperform this ratio for a range of values $k = \Theta(n)$. On the other hand, the SDP relaxation methods have a limit of $n^{\Omega(1)}$ for DkSP as shown by Feige and Seltser [14] and Bhaskara et al. [7].

For some special cases in terms of graph classes, values of k and optimal objective values, better approximations have been obtained for DkSP and HkSP. Arora et al. [3] gave a PTAS for the restricted DkSP where $m = \Omega(n^2)$ and $k = \Omega(n)$, or each vertex of G has degree $\Omega(n)$. Kortsarz and Peleg [21] approximated DkSP with ratio $O((n/k)^{2/3})$ when the number of edges in the optimal solution is larger than $2\sqrt{k^5/n}$. Demaine et al. [10] developed a 2-approximation algorithm for DkSP on H-minor-free graphs, where H is any given fixed undirected graph. Using a greedy approach, Liazi et al. [25] designed a 3-approximation algorithm for DkSP on the class of chordal graphs. Chen et al. [8] showed that DkSP admits $O(\sigma)$-approximation for the graphs that have polynomial time computable σ-quasi elimination order. This implies constant factor approximations for a large family of intersection graphs, including chordal graphs, circular-arc graphs, claw-free graphs and disk graphs. Several PTAS have been designed for DkSP on unit disk graphs [8], interval graphs [26], and a subclass of chordal graphs [24].

The work on approximating heaviest/densest connected k-subgraphs are relatively very limited. To the best of our knowledge, the existing polynomial time algorithms deal only with special graphical topologies, including:

- $O(kn + m)$-time 4-approximation [29] and $O(n^2 + k\log k)$-time 2-approximation [17] for the metric HkSP and HCkSP, where the underlying graph G is complete, and the connectivity is trivial;
- $O(k^2 n)$-time exact algorithms for HkSP and HCkSP on trees by dynamic programming [9,27];
- polynomial time exact algorithms for DkSP and DCkSP on h-trees, cographs and split graphs [9], DkSP, where the optimal solutions found for DkSP are guaranteed to be connected; and
- $O(k^4 n)$-time exact algorithm for DCkSP on interval graphs whose clique graphs are simple paths [23].

Among the well-known relaxations of DkSP and HkSP is the problem of finding a (connected) subgraph (without any cardinality constraint) of maximum weighted density. It is strongly polynomial time solvable using max-flow based techniques [15,22]. Andersen and Chellapilla [2] and Khuller and Saha [20] studied two relaxed variants of DkSP/HkSP for finding a densest/heaviest subgraph with at least or at most k vertices. The former variant was shown to be NP-hard even in the unweighted case, and admit 2-approximations in the weighted setting. The approximation of the latter variant was proved to be as hard as that of DkSP/HkSP up to a constant factor.

Our results. Given the interest in finding heaviest/densest connected k-subgraphs from both the theoretical and practical point of view, a better understanding of the problems is an important challenge for the field. In this paper, we design $O(mn\log n)$ time combinatorial approximation algorithms for finding

- a connected k-subgraph of G whose weighted density (equivalently, total edge weight) is at least a factor $\Omega(\max\{k^2/n^2, 1/k\})$ of that of the heaviest k-subgraph of G (which is not necessarily connected), and
- a connected k-subgraph of G whose density (equivalently, number of edges) is at least a factor $\Omega(n^{-2/5})$ of that of the densest k-subgraph of G (which is not necessarily connected).

These particularly provide the first non-trivial approximations for the heaviest/densest connected k-subgraph problem on general graphs: $O(\min\{k, n^2/k^2\})$-approximation for HCkSP (note $\min\{k, n^2/k^2\} \le n^{2/3}$) and $O(\min\{n^{2/5}, n^2/k^2\})$-