A Hybrid Evolutionary Algorithm for the Diversified Top-k Weight Clique Search Problem (Student Abstract)

Jun Wu , Minghao Yin*

School of Information Science and Technology, Northeast Normal University, China
Changchun, 130117, China
{wuj342, ymh}@nenu.edu.cn

Abstract

The diversified top-k weight clique (DTKWC) search problem is an important generalization of the diversified top-k clique search problem, which extends the DTKC search problem by taking into account the weight of vertices. This problem involves finding at most k maximal weighted cliques that cover maximum weight of vertices with low overlapping in a given graph. In this study, a mixed integer linear program constraint formulation is proposed to model DTKWC search problem and an efficient hybrid evolutionary algorithm (HEA-D) based on some heuristic strategies is proposed to tackle it. Experiments on two sets of 110 graphs show that HEA-D outperforms the state-of-art methods.

Introduction

Given an undirected graph $G = (V, E)$, a clique is a subset of the graph G, where any two vertices are adjacent. A clique is not covered by any other clique from G, so that this clique is a maximal clique. The diversified top-k clique (DTKC) search problem is to find k maximal cliques to occupy as many vertices as possible, where k is a parameter that requires to be provided. The diversified top-k weight clique (DTKWC) search problem extends the DTKC search problem by taking into account the weight of vertices, and it attempts to find k maximal weight cliques that are not only informative but also large in the sense that they are together cover the most weight in the graph. This problem can be verified as an NP-hard problem inspired by Wu et al. (2020).

Recently, considerable attentions have been paid to solve diversified top-k cohesive group problems on large graphs. And they can be very well applied to practical applications, such as the influential community in social networks, motif discovery in molecular biology, and advertisement putting on TV programmes. In the literature, there are few methods for diversified top-k cohesive group problems. Such as Yuan et al. (2015; 2016) propose the concept of DTKC and then provide some approximate algorithms for it. Wu et al. (2020) provide a state-of-the-art local search algorithm to solve the DTKC search problem. Zhou et al. (2021) propose a weighted version of DTKC, called DTKWC, and solve it by encoding this problem into the weighted partial MaxSAT (WPMS) problem with WPMS solvers. Wu and Yin (2021) introduce a new network model, named DTKSP, and solve it by a local search method. However, due to the NP-hardness of the DTKWC search problem, the exact algorithm failed to encode large even medium-sized graphs to WPMS. Furthermore, to the best of our knowledge, the powerful population-based evolutionary approach has not been investigated for the diversified top-k cohesive group problems. In this work, we present an effective hybrid evolutionary algorithm called HEA-D (hybrid evolutionary algorithm for DTKWC) and extensive experiments are carried out on two benchmarks to evaluate it.

Constraint Formulation for the DTKWC Search Problem

Given a graph $G = (V, E, w)$, and a fixed integer k, the DTKWC search problem is to compute a set C, such that each $c \in C$ is a maximal weight clique, $|C| \leq k$, and the total weight of C is maximized. Let x_{ih} be the binary variable such that $x_{ih} = 1$ if the vertex v_i is allocated to the h^{th} maximal weighted clique, and $x_{ih} = 0$ otherwise. Likewise, let y_i be a binary variable associated with vertex v_i such that $y_i = 1$ if v_i in a maximal weighted clique, $y_i = 0$ otherwise. w_i denotes the weight of vertex v_i. $\bar{E}$ is a set of the edges in the complement graph $\bar{G} = (V, \bar{E}, w)$. The DTKWC search problem can be expressed by the following mathematical formulation.

$$Maximize \quad f(G) = \sum_{i=1}^{|V|} y_i w_i \tag{1}$$

Subject to:

$$x_{ih} + x_{jh} \leq 1, \forall (v_i, v_j) \in \bar{E}, h \in [1, k] \tag{2}$$

$$y_i = \begin{cases} 0, & \text{if } \sum_{h=1}^{k} \sum_{i=1}^{|V|} x_{ih} = 0, \\ 1, & \text{otherwise.} \end{cases} \tag{3}$$

$$x_{ih} \in \{0, 1\}, \forall i \in [1, |V|], \forall h \in [1, k] \tag{4}$$

In the above formulation, the objective function (1) is to maximize the function $f(G)$, in other words, maximize the total weight of the selected maximal weight cliques. The intuitions of other formulas above are given next.

*Corresponding author.

Algorithm 1: The main framework of the HEA-D algorithm

Input: a weighted graph $G(V, E, w)$, one integer k, population size p and other parameters of the algorithm
Output: a set C^* containing at most k maximal weight cliques

1: $Pop \leftarrow PoolInitialize(p)$;
2: $C^* \leftarrow \arg\max_{C \in Pop} W(C)$; /* $W(C)$ is the total weight of vertices in C */
3: **while** stopping condition is not met **do**
4: Randomly select two solutions C_1 and C_2 from Pop;
5: $C_0 \leftarrow Crossover(C_1, C_2)$;
6: $C_0 \leftarrow LocalOptimization(C_0)$;
7: **if** $f(C_0) > f(C^*)$ **then**
8: $C^* \leftarrow C_0$;
9: **end if**
10: $Pop \leftarrow UpdatePool(Pop, C_0)$;
11: **end while**
12: **return** C^*;

1. Constraint (2) is guaranteed that there is an edge between every two vertices in a clique.
2. Constraint (3) means that the value of $y_i = 1$ if there exists at least one $x_{ih} = 1$ ($v_i \in V$ and $1 \leq h \leq k$). Otherwise, the value of $y_i = 0$.
3. Constraint (4) indicates that x_{ih} ($v_i \in V$ and $1 \leq h \leq k$) can only be assigned a value of 0 or 1.

Therefore, maximizing the value $f(G)$ in Equation (1) is the optimal solution of DTKWC search problem.

Hybrid Evolutionary Algorithm

In this section, we propose a hybrid evolutionary algorithm for DTKWC search problem, called HEA-D. The main framework of HEA-D is shown in Algorithm 1, that is a powerful memetic framework for solving difficult optimization problems. HEA-D has strong global exploration and local expansion capabilities, benefiting from two main components in the memetic framework, the crossover operator and the local optimization procedure.

HEA-D starts with an initial population of at most p ($p \geq 2$) individuals (line 1). After that, HEA-D performs a number of generations to evolve the population until the given stopping condition is met. In the paper, the condition is a cutoff time limit (lines 3-11). During each generation, HEA-D selects two parent solutions C_1 and C_2 from the initial population Pop. A clique-based crossover operator is used to recombine C_1 and C_2 to generate an offspring solution (line 5). Generally, the quality of this offspring solution is not good enough. Thus, a simulated annealing based local optimization procedure is performed to improve the quality (line 6). Finally, the improved offspring solution is used to update the population according to adopted updating rule (line 10).

Experimental Evaluation

We evaluate HEA-D on a wide range of real-world benchmarks, including 8 advertisement putting graphs which are a set of real size TV programmes provided by SAPPRFT ($size \times 10000$ is the sum of viewers of selected programmes), and 102 large real-world graphs which are transformed from the unweighted graphs from Network Data Repository[1] with the weights generated as $w(v_i) = (i \ mod \ 200) + 1$. For advertisement putting, we used the optimal results obtained from the WPMS solvers. For large real-world graphs, we compared it with the CPLEX solver (version 12.9) which uses the mathematical model presented in this paper, and a weighted version of the state-of-the-art algorithm for DTKC search problem called TOPKWCLQ. For each graph, HEA-D and TOPKWCLQ are performed on 10 independent runs with a cutoff time (600 s). For CPLEX solver, a cutoff time of one hour is used.

Benchmark	k	CPLEX		TOPKWCLQ		HEA-D	
		#Better	#Equal	#Better	#Equal	#Better	#Equal
real-world graphs (102)	10	0	13	0	49	53	49
	20	0	3	0	39	63	39
	30	0	4	0	31	70	32
	40	0	4	0	28	74	28
	50	0	4	0	28	74	28

Table 1: Summary of comparison between CPLEX, TOPKWCLQ and HEA-D on real-world graphs. #Better denotes the number of graphs where two algorithms find the best objective value. #Equal denotes the number of graphs where an algorithm can find equal objective values with another one.

For each graph in our experiments, the parameter k is set to 10, 20, 30, 40, and 50, respectively. Thus, we totally have 550 DTKWC search problem instances. The summary results are shown in Table 1. Both of WPMS solvers and HEA-D can find the optimal results on advertisement putting instances within less than 10^{-3} second, and the results are not reported here. For real-world instances, HEA-D performs better than CPLEX and TOPKWCLQ on all instances. In particular, HEA-D established new lower bounds for 334 out of the 550 instances.

References

Wu, J.; Li, C.; Jiang, L.; Zhou, J.; and Yin, M. 2020. Local search for diversified top-*k* clique search problem. *Comput. Oper. Res.*, 116: 104867.

Wu, J.; and Yin, M. 2021. Local Search for Diversified Top-k s-plex Search Problem (Student Abstract). *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(18): 15929–15930.

Yuan, L.; Qin, L.; Lin, X.; Chang, L.; and Zhang, W. 2015. Diversified top-k clique search. In *2015 IEEE 31st International Conference on Data Engineering*, 387–398. IEEE.

Yuan, L.; Qin, L.; Lin, X.; Chang, L.; and Zhang, W. 2016. Diversified top-k clique search. *The VLDB Journal*, 25(2): 171–196.

Zhou, J.; Li, C.; Zhou, Y.; Li, M.; Liang, L.; and Wang, J. 2021. Solving diversified top-k weight clique search problem. *Sci. China Inf. Sci.*, 64(5).

[1]http://www.graphrepository.com/index.php