클리크 마이닝에 기반한 새로운 커뮤니티 탐지 알고리즘 연구

양예선 [1], 펭소니 [1], 박두순 [1,*], 김석훈 [1,*] 이혜정 [2], 싯소포호트 [1]
[1]순천향대학교 소프트웨어융합학과
[2]순천향대학교 AI • SW 교육원
yangyxuan621@gmail.com, peng.sony61@gmail.com, parkds@sch.ac.kr, seokhoon@sch.ac.kr, hjlee100@sch.ac.kr, siet.sophort60@gmail.com

A Novel Study on Community Detection Algorithm Based on Cliques Mining

Yixuan Yang[1], Sony Peng[1], Doo-Soon Park[1], Seok-Hoon Kim[*,1], HyeJung Lee[2], Sophort Siet[1]
[1]Dept. of Software Convergence, Soonchunhyang University
[2] Institute for Artificial Intelligence and Software, Soonchunhyang University

Abstract

Community detection is meaningful research in social network analysis, and many existing studies use graph theory analysis methods to detect communities. This paper proposes a method to detect community by detecting maximal cliques and obtain the high influence cliques by high influence nodes, then merge the cliques with high similarity in social network.

1. Introduction

Community structure is an important feature of complex networks. In simple terms, to find a community in a network is to divide similar nodes into a set, so that the interaction between the nodes in the set is stronger than the interaction between the nodes out of the set, that is, the interaction between nodes within the same community is stronger. Meanwhile, the links in a community are relatively dense, and the links between different communities are relatively sparse [1].

However, the multiple social attributes of users in social networks lead to the fact that users can belong to multiple communities at the same time, so the community discovery algorithm based on overlapping clustering is better. Discovering high-quality communities is helpful for understanding real complex networks.

By analyzing the existing research, there have been five types of overlapping community discovery algorithms, namely clique filtering algorithm, local extended community discovery algorithm, fuzzy overlapping community discovery algorithm, edge community discovery algorithm and tag propagation algorithm.

Among them, mining cliques to detect communities is a representative method. As a basic cohesive subgraph model, clique is widely used to reveal the dense community structure of graphs[2]. Clique, the complete subgraph of graph, is an important structure in graph theory. The name derives from the representation of cliques of people in social networks [3].

Cliques mining [4] is a hot and interesting topic in graph theory. For the cliques mining, there have been several studies, such as detecting maximal cliques (maximal complete subgraph), maximum cliques (a clique is maximum only if it is not a subgraph of another), balanced cliques (a clique has same positive and negative weighted edges in signed graph).

In our work, we intend to search the community in the social networks by mining maximal cliques and merging similar cliques, according to consider the similarity of overlapping cliques.

2. Related Work

One particular application of cliques is to detect and mining communities in social networks, such as in [5, 6], or equivalently, the cohesive groups in the graph of social network [7]. Das et al. [8] utilized a parallel graph algorithm to enumerate maximal cliques of big-scale biological networks; Cheng et al. [9] first proposed an efficient

*Corresponding author: Doo-Soon Park, Seok-Hoon Kim
Acknowledgments: This research was supported by the National Research Foundation of Korea (No. NRF-2022R1A2C1005921) and BK21 FOUR (Fostering Outstanding Universities for Research) (No.5199990914048)

partition-based approach to deal with the problem of processing large graphs with limited memory for maximal cliques enumeration algorithms.

In 2005, Gergely et al. proposed the Clique percolation method (CPM) algorithm [7]. Its core idea is to find overlapping communities based on k-maximal cliques. A fully connected subgraph composed of k nodes is called k-maximal group, and the value of k is given by researchers. CPM is designed to find adjacent k-maximal cliques. Due to the full connectivity between the internal nodes of maximal cliques, a community structure with dense interior and sparse exterior can be formed, which is an ideal community structure. Adjacent k-maximal cliques are the overlapping community structures that the CPM looks for.

In 2015, Zhang et al. proposed a new overlapping community discovery method, called the MOHCC algorithm [10]. The algorithm combines the Coupling Strength [11] proposed by Wang as the objective function to merge of the maximal cliques, so as to obtain the best hierarchical division.

For the detection of maximal cliques, there is a famous and classical method named as Bron-Kerbosch method[12], which is a recursive backtracking algorithm and proposed by Coenraad Bron and Joep Kerbosch in 1973.

3. Problem Statement

Before presenting our problem addressed in this paper, in order to improve the understanding, several fundamental definitions in social networks and maximal cliques will be provided at first.

Definition 1 (Social Network) A Social Network can be represented as a graph G=(V, E), with V indicating the set of nodes, $V = \{v_1, v_2, v_3, \ldots, v_n\}$, E indicating the set of edges $E = \{e_{ij} | i,j \in V\}$.

Definition 2 (Clique) [2] Let G = (V, E) be a Social Network. A clique in G is a subset $C \subseteq V$ such that for any two nodes vi , vj ∈ C, there exists an edge (vi , vj) ∈ E. The **maximal clique** is the subset of a complete subgraph C cannot add any more nodes.

Problem Definition: Given a social network G, our research aims to detect the community in the social network G by detecting maximal cliques and merge the similar cliques.

$$CL = cliques(G) \quad (1)$$

$$\Omega = merger(CL) \quad (2)$$

CL is the set of all maximal cliques in G, and Ω is the community set which merge the similar cliques in set CL .

4. Proposed Approach

In this section, we will propose the approach of detecting absolute fairness maximal balanced cliques in the Signed Attribute Social Network.

The step of our method is:

1. Detect the maximal cliques by BK (Bron-Kerbosch) method.
2. Sort the influence of nodes by PageRank and get top-k nodes.
3. Calculate the similar with cliques which contain the top-k nodes and merge eligible cliques.

In the step 1, the BK (Bron-Kerbosch) method is a traditional algorithm for detecting maximal cliques in the field of graph theory. The processing of BK is described and explained in lines 4-11 of Algorithm below.

In the step 2, we use the formulation to compute the influence of nodes by PageRank, t is a threshold and usually the value is 0.85:

$$Inf(Vi) = t \sum_{Vl \in F(Vi)} \frac{Inf(Vl)}{N(Vl)} + \frac{1-t}{N} \quad (3)$$

For the similarity of cliques, we give the formulation to compute when the value δ >= 0.8, the community modularity reaches the optimal value:

$$\delta_{C_p,C_q} = \frac{|C_p \cap C_q|}{\min(|C_p|,|C_q|)} \quad (4)$$

The process of the approach is described as a pseudo-code as follows.

Algorithm:	
1	Input G=(V, E), k
2	Initial sets R, P, X, InfluenceNum, NodesList, C as
3	empty sets
4	Bron-Kerbosch(R, P, X) // Step 1: BK method
5	if P and X are both empty:
6	report R as a maximal clique
7	for each vertex v in P:
8	Bron-Kerbosch(R.union({v}), P.intersection
9	(N(v)), X.intersection(N(v)))
10	P ← P\ {v}
11	X← X.union ({v})
12	For Vi in V: // Step 2: Sort the influence of nodes
13	InfluenceNum ← Inf(Vi)
14	NodesList ← sorted(InfluenceNum)
15	For n in NodeList in range(k): // Step 3: Compute the
16	similarity within different cliques
17	Search Cp, Cq contain n in set R:
18	If the Similartiy(Cp, Cq) > 0.8:
	Ω← merge(Cp, Cq)
13	**Return Ω**

<Figure 1> The overview of our method.

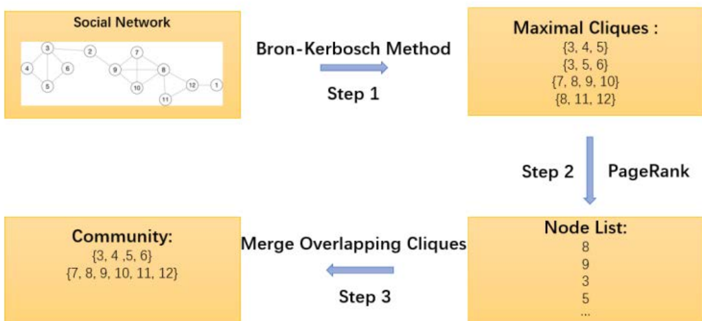

Figure 1 illustrates a case study of our method. First, we give a social network graph with 12 nodes and 17 edges. In the first step, we mine 4 maximal cliques through the Bron-Kerbosch algorithm, which are: {3, 4, 5}, {3, 5, 6}, {7, 8, 9, 10} and {8, 11, 12}; Then, use the PageRank algorithm to rank the influence of the nodes to obtain the top-k nodes: 8. 9, 3, 5 …; Finally, use the formulation (4) to calculate the similarity of different cliques, and compare the results with the threshold (0.8), thereby merging eligible cliques.

5. Conclusion

This paper mainly studies the problem of detecting community in social networks. We use Bk method to detecting maximal cliques, search the high influence nodes by PageRank method, and calculate the similarity of cliques, then to merge the cliques with high similarity by the formulation in the paper.

Reference

[1] Newman M E J. The Structure and function of complex networks. SIAM Review, 2003, 45(2): 167−256

[2] Yixuan Y., Fei H, Doo-Soon P., et al. Modelling Prevention and Control Strategies for COVID-19 Propagation with Patient Contact Networks[J]. Human-centric Computing and Information Sciences, 2021, 11.

[3] Luce R D, Perry A D. A method of matrix analysis of group structure[J]. Psychometrika, 1949, 14(2): 95-116.

[4] Yixuan Y., Fei H, BeiBei P., et al, "Dynamic Maximal Cliques Detection and Evolution Management in Social Internet of Things: A Formal Concept Analysis Approach," in IEEE Transactions on Network Science and Engineering, doi: 10.1109/TNSE.2021.3067939.

[5] Palla G, Derényi I, Farkas I, et al. Uncovering the overlapping community structure of complex networks in nature and society[J]. nature, 2005, 435(7043): 814-818.

[6] Kumpula J M, Kivelä M, Kaski K, et al. Sequential algorithm for fast clique percolation[J]. Physical review E, 2008, 78(2): 026109.

[7] Wasserman S, Faust K. Social network analysis: Methods and applications[J]. 1994.

[8] Das A, Sanei-Mehri S V, Tirthapura S. Shared-memory parallel maximal clique enumeration[C]//2018 IEEE 25th International Conference on High Performance Computing (HiPC). IEEE, 2018: 62-71.

[9] Cheng J, Zhu L, Ke Y, et al. Fast algorithms for maximal clique enumeration with limited memory[C]//Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. 2012: 1240-1248.

[10] Zhang Z, Wang Z. Mining overlapping and hierarchical communities in complex networks[J]. Physica A: Statistical Mechanics and its Applications, 2015, 421: 25-33.

[11] Wang L. Using the relationship of shared neighbors to find hierarchical overlapping communities for effective connectivity in IoT[C]//2011 6th International Conference on Pervasive Computing and Applications. IEEE, 2011: 400-406.

[12] Akkoyunlu E A. The enumeration of maximal cliques of large graphs[J]. SIAM Journal on Computing, 1973, 2(1): 1-6.