Research on collapsed \((\alpha, k)\)-NP-community on signed graph

Jiyuan Zhang, Yan Yang

School of Computer Science and Technology, Heilongjiang University, China
zjiyuan312569@163.com, yangyan@hlju.edu.cn

Abstract. In social network, the departure of critical users can lead to the disintegration of the entire community, i.e., lead a large number of other users to drop out. This paper aims to find the critical users in the community and prevent the whole community from disintegrating. Most of the other research work is carried out on the unsigned network. In this paper, we study this problem on a signed network and propose a new community model \((\alpha, k)\)-NP-community. Specifically, the nodes in the community model should satisfy the maximum subgraph with no less than \(\alpha k\) good friends (positive neighbor) and no more than \(k\) enemy (negative neighbor). We propose an effective algorithm to find \((\alpha, k)\)-NP-community. In order to identify the critical users of \((\alpha, k)\)-NP-community, we propose the problem of collapsed \((\alpha, k)\)-NP-community: given a signed graph \(G\), integers \(\alpha, k\), our goal is to find such nodes in graph \(G\), the departure of the best collapsers will lead to the smallest \((\alpha, k)\)-NP-community. We prove that the problem is NP-hard. Finally, we present two lemmas to apply the \(\alpha\)KCA algorithm, which effectively reduced the number of candidate nodes and accelerated the search process. The results of extensive experiments on four real datasets demonstrate the effectiveness and efficiency of our algorithms.

1. Introduction

With the rapid development of computer science, the concept of community has entered people's field of vision. How to maintain the stability of the whole community, that is, to prevent the disintegration of the community, has become a very popular task. To solve this problem, we have to mention the concept of user engagement. User engagement \([0][0][0]\) is a very practical concept in the community and has been mentioned frequently in recent years. Traditional community models such as \(k\)-core[0] and \(k\)-truss[0] are used to measure user engagement. We propose a new community model \((\alpha, k)\)-NP-community to measure user engagement. The user in community has two choices: one is to stay, the other is to leave. In \((\alpha, k)\)-NP-community, when the number of good friends of the user is no less than \(\alpha k\) and the number of enemy of the user is no more than \(k\), the user chooses to stay, otherwise he will choose to leave. For \((\alpha, k)\)-NP-community, the departure of one user in the community may lead to the departure of other users. The reason for leaving is that the condition of \((\alpha, k)\)-NP-community is not met (the degree of the positive edge is less than \(\alpha k\)). When the departure of a user causes the maximum number of users to leave, the size of final \((\alpha, k)\)-NP-community obtained is used to measure user engagement.

How can we accurately find the critical users in the \((\alpha, k)\)-NP-community? This paper puts forward the problem of collapsed \((\alpha, k)\)-NP-community, that is, given Graph \(G\), integer parameter \(\alpha, k\), our goal is to find node \(q\), when \(q\) is deleted, the level of community disintegrating will be the largest, which can be said that the largest number of users will follow this node to leave. Next, this paper takes a specific example to illustrate this problem:
As shown in Fig 1, Fig 1 is a signed network, the '+' edge represents a friend relationship between two users, and the '-' edge represents an enemy relationship between two users. The whole graph forms a social network of relationships. Assuming that User $u_4$ wants to invite his friends. The friends he wants to invited must satisfy the following conditions. First: The number of friends for each user must be no less than 2 ($\alpha=2$, $k=1$); Second: The number of enemy should be no more than 1 ($k=1$). This makes it easy to judge that $u_2$ and $u_6$ will not accept $u_4$’s invitation, and everyone else will meet the conditions. The remaining friends are only $u_1$, $u_3$, $u_5$, $u_7$, $u_8$, $u_9$. Assuming that User $u_5$ cannot come to the party for some reason. Because of the departure of friend $u_5$, users $u_1$ and $u_7$ do not meet the above mentioned conditions. Therefore, they will not participate in this party. That is, the community of the invited friends will have a certain extent of disintegration because of the departure of the user $u_5$.

At the same time, it is noted that the problem we are studying is different from the problem of influence maximization. In Figure 1, the user with the most good friends is User $u_9$, who has 5 friends, but the departure of User $u_9$ does not cause any one to leave.

The main contributions of this paper are as follows:
1) we study the collapsed community problem on signed network for the first time;
2) We propose a new community model ($\alpha$, $k$)-NP-community to measure the user engagement;
3) We propose two lemmas apply to the nkCA algorithm to solve the collapsed ($\alpha$, $k$)-NP-community problem;
4) The experiment proves the efficiency and effectiveness of our proposed algorithm.

2. Related work
Dense subgraph mining is a very interesting direction in social networks. Many different dense subgraph models are proposed. A more classic dense subgraph model is $k$-clique [0][0]. $k$-clique is a complete graph with $k$ vertices, $k(k-1)/2$ edges. But one drawback of $k$-clique is that its requirements are too strict. So there are subgraphs that are relatively loose but can also guarantee their cohesiveness. Several more relaxed forms of dense subgraphs are proposed, Such as: $k$-plex[0], quasi-clique[0], n-clan[0], n-club[0], $k$-core[0][0][0][0], $k$-truss[0], and maximal $k$-edge connected subgraph [0] and other models. These models greatly enrich the work of dense subgraph mining.

User engagement in social networks is a popular concept in recent years. $k$-core and $k$-truss are popular in social networks because of their own characteristics, and can measure participation in social networks. Luo et al. [0] studied the complexity of collapsed $k$-core problem. Bhawalkar [0] and others proposed anchoring the $k$-core problem to prevent the separation of social networks. This problem describes the finding of $b$ vertices outside the $k$-core, referred to as anchor vertices, so that $k$-core remains the largest. Zhang et al. [0] proposed a heuristic algorithm to solve this problem on the general graph. Anchored $k$-core problem is different from collapsed $k$-core [0]problem. The problem of dissolving the $k$-core is to study the problem on the $k$-core, and anchoring the $k$-core is to study the problem outside the $k$-core.
Similarly, Tie strength[0] [0] has also entered people's field of vision and is a very important concept in social networks. But k-core does not guarantee that its tie strength is very compact. Cohen then introduced the k-truss [0] model to describe tie strength.

Signed networks have ignited a lot of interest in recent years. For example, in the community detection, link prediction[0], recommendation system[0][0] and other directions have a wide range of applications. Our research has a very close relationship with community detecting [0]. We present a close model that uses a signed network to unfold our work. The new model we proposed excavated a deeper relationship within the community while ensuring cohesion.

3. (α, k)-NP-community On signed graph

3.1 Basic definition

Let G=(V, E) be an undirected signed graph, V denote vertex set, E denote edge set. |V|=n, |E|=m denote vertex number and edge number respectively. In figure G, for each edge e∈E is accompanied by a label, either '+' or '-'. An edge with a label of '+' is called a positive, and an edge with a label of '-' is called a negative. We use degu(+) to represent the number of positive neighbors of node u, that is, degu(+)=[Nu+] and degu(-) to represent the number of negative edges connected to node u, that is, degu(-)=[Nu-]. We use degN(u+) represent neighbor vertex degree of u. We record the edge connected to node u in G as E(u, G). We use deg(u(+), S), the positive degree of u in S, to represent the number of positive adjacent vertices of u in S. Next, this paper proposes a new community model to describe the cohesive subgraph, which is defined as follows:

Definition 1. ((α, k)-NP-community): Given an undirected signed graph G, a positive integer α, a positive integer k, an induced graph C that satisfies the following three constraints is called (α, k)-NP-community, where N stands for Negative and P stands for Positive:

- Positive edge constraint: for each node u∈C, degu(+)≥αk;
- Negative edge constraint: for each u∈C, degu(-)≤k;
- Maximum nature: the largest subgraph that satisfies the above constraint.

According to definition 1, positive edge constraint guarantees that each node in the community has at least αk positive neighbors, while negative edge constraint guarantees that each node in the community does not have too many negative neighbors. And the maximum nature satisfies the basic requirements of subgraph mining. According to definition 1, we give the definition of collapse (α, k)-NP-community.

Definition 2. (collapsed (α, k)-NP-community): Given a graph G and a node q, the collapsed (α, k)-NP-community, denoted by Cq(G), is the corresponding (α, k)-NP-community of G with vertices in A removed.

When node q is deleted, more vertices in graph G might be deleted as well due to the contagious of (α, k)-NP-community computation. These vertices are called followers of the collapsed vertices q, denoted by F(q, G).

Problem statement: given the graph G, positive integer α, k, the collapsed (α, k)-NP-community problem aims to find vertex q in G so that the resulting collapsed of (α, k)-NP-community is minimized when q is deleted; that is F(q, G) is maximized.

Example 2: In figure 1, we set α = 2, k = 1. We have node u5 which is the most collapsed node of the community, where Cu5(G)=\{u1, u2, u3, u6, u8, u9\}, F(u5)=\{u1, u7\}.

3.2 Complexity

Theorem 1: For any given positive integer α, k (k ≥ 1, α ≥ 1), the collapsed (α, k)-NP-community problem is NP-hard.

Proof: Because α≥1, k≥1, and because degu(+)≥degu(-), because we have to consider the number of good friends as much as possible, the number of general friends is as small as possible, for collapsed (α, k)-NP-community problem, we only need to consider the number of good friends, without having to consider the number of normal friends, that is, only need to consider degu(+)≥αk. This can be seen as a disintegration of the k-core problem, because the k-core problem is only
considered when the node is less than or equal to k, as a result of the collapsed k-core problem is NP-hard, so the collapsed (α,k)-NP-community problem is also NP-hard.

Theorem 2: If f(q)=|F(q)|, then f(q) is monotonic.

Proof: Suppose there is a set A' containing A. For each node u in F(A), when F(A') is calculated, each node in F(A) will still be deleted. Since the degree of the node in F(A) does not increase when the node in A \ A is deleted. It can be seen that f(q) is monotonic.

4. Solution

4.1 (α, k)-NP-community Algorithm
In this section, we describe how to solve the collapsed (α, k)-NP-community problem. We need to find (α, k)-NP-community from the signed graph G'. Algorithm 1 describes the process of finding (α, k)-NP-community.

Algorithm 1 describes the process of searching (α, k)-NP-community on a signed network.

Algorithm 1: (α, k)-NP-community computation (G, α, k)

Input: G: a signed graph, α, k: integer parameters
Output: (α, k)-NP-community

1 While exists u ∈ G with degu(+) < αk do
2   G := G \ {u ∪ E(u, G)}
3 for each v with degv(−) > k in G:
4   while (the number of negative edge of v)/(the positive edge number of adjacent node of v) is maximal
5     remove v from G
6   goto line1;
7 return (α, k)-NP-community

This algorithm input a scope signed graph G', integer parameters α, k, and finally output (α, k)-NP-community. In line 1-2 of algorithm 1, the subgraph satisfying the requirement of positive edge is found in the scope signed graph G'; in line 3-5, the node not satisfying the requirement of negative edge is deleted; in addition, the negative edge of P=(the number of negative edge of v)/(the positive edge number of adjacent node of v) is deleted, and the condition that the value of P is the largest is required to be satisfied. When the current node is deleted, the first line of the algorithm is returned and executed alternately in turn, and finally (α, k)-NP-community is obtained.

4.2 The collapsed algorithm
In this section, we query the node with the most followers on that community. A relative direct algorithm is to calculate the number of followers for each node in (α, k)-NP-community, and then compare the number of followers of all nodes, and finally get the node with the largest number of followers. Algorithm 2 describes this process.

Algorithm 2: αkGCA (G, α, k)

Input: (α, k)-NP-community graph G' under range constraint
Output: best of the collapsers

1 A := ∅;
2 For each u ∈ (α,k)-NP-community
3   Compute followers
4 U* = the best collapsers
5 A = A ∪ U*

Although algorithm 2 is simple and easy to think about, its algorithm complexity is high and its efficiency is low. Therefore, in order to improve the efficiency of the algorithm, we propose two lemmas to reduce the number of candidate collapse nodes.
Lemma 1: if a collapsed vertex x has at least one follower, if and only if the number of friends of the vertex x equals ak, or the number of friends of the user's neighbors equals ak.

Proof: proof by contradiction: in (α, k)-NP-community, suppose there exists a vertex u, deg_u(+) > ak, and its neighbor vertex degree deg_Nu(+) > ak has followers. If vertex u is attracted by another community for some reason, then the degree of all neighbor nodes of vertex u is at most reduced by 1, and the remaining nodes still satisfy the definition of (α, k)-NP-community, so vertex u does not have a follower. This contradicts the hypothesis, so the lemma is proved.

Lemma 2: if vertex x is a follower of vertex u and user v is a follower of user x, then user v must be a follower of user u. We can conclude that if x ∈ F(u), F(x) is contained in F(u).

Proof: x is a follower of u. It means that when u is attracted by other communities, x will disintegrate with u because it does not meet the conditions of (α, k)-NP-community. For the followers of vertex x, because the departure of user u will lead to the departure of user x, so the followers of user x are subset of the followers of user u, i.e., F(v) ⊆ F(u).

Based on lemma 1 and lemma 2, an αkCA algorithm is proposed. This algorithm is designed to efficiently find the best collapse. The algorithm first calculates the node set P with αk positive edge, i.e., P = {u: deg(u, (α, k)-NP-community)} = ak, and then records T = P ∪ N(P). Then, for each node in T, the number of its followers is calculated. Finally, the user with the largest follower is returned.

Algorithm 3 αkCA((α, k)-NP-community)
input : (α, k)-NP-community
Output: the best collapse
1. P = {u: deg(u, (α, k)-NP-community)} = ak
2. T = P ∪ N(P) (lemma1)
3. For each vertex in T:
4. Compute F(u, G);
5. T = T \ F(u, G) (lemma 2)
6. Return the best collapse;

5. Experiment
In this section, we conduct extensive experiments to prove the effectiveness and efficiency of our proposed algorithm. As we know, no one has studied the collapsed (α, k)-NP-community problem on signed graphs. The experiment is divided into two parts. The first part is to compare effectiveness of αkCA algorithm and αkGCA algorithm. The second part is to test the efficiency of α and k parameters.

All programs are implemented in Java in Windows10. All experiments are performed on a machine with Inter(R) Core (TM) 2.8GHz CPU and windows10 system.

5.1 Datasets
We use four real datasets to test our algorithm. The four datasets are Slashdot, Wiki, DBLP and Youtube. Slashdot and Wiki datasets are both signed networks. DBLP is a cooperation network, each node represents a author, each edge represent two authors worked at least one article, in order to create a signed network, as the two authors cooperation of no less than a given threshold r, we will set this edge to "+", when r is less than threshold, it will be set to "-". For Youtube datasets, Youtube is a social networking network. We randomly chose 30% of the edges to be "-" and the remaining edges to be "+". Slashdot, Youtube, and DBLP datasets are downloaded from the Stanford web datasets (http://snap.staford.edu). The Wiki is downloaded from the Koblenz network data collection (http://konect.uni-koblenz.de/).

| Table 1. Datasets |
|-------------------|
| Datasets | n=|V| | m=|E| | |E+| | |E-| | kmax |
|----------|---|---|---|---|---|---|
| Slashdot | 82,144 | 500,481 | 382,882 | 117,599 | 54 |
| Wiki     | 138,592 | 715,883 | 631,546 | 84,337 | 55 |
5.2 The effectiveness of the αkCA algorithm
This section analyzes the effectiveness of the αkCA algorithm. We set the query node to 4. Figure 2 compares the number of followers of users obtained by different methods. One of the methods (random) randomly selects a node as the user node to disintegrate, and the other method (Degree) finds the node with the largest degree in the T set as the disintegration node. Compare the two ways of finding the user node to the method of finding the user node in this paper, as shown in figure 2.

Fig 2: Number of the followers  Fig 3: Different dataset algorithm time comparison

We can see from Figure 4 that the random selection of the followers of the user node is the least. The selection of the most moderately disassembled user nodes in the T set is better than the random selection method, but the αkCA algorithm proposed by us is obviously superior.

5.3 Efficiency of the αkCA algorithm
We compare the speed of the αkCA algorithm with the αkGCA algorithm and express it by time. In this experiment, we set the query node to 4 and the parameters to α=4, k=2. The result is shown in Figure 3.

As can be seen from Figure 3, the performance of the αkCA algorithm is better than the αkGCA algorithm. This is obvious because the two lemmas proposed significantly reduce the number of candidate nodes, thus speeding up the process of finding collapser.

5.4 The impact of different parameters on algorithm efficiency
This section tests the efficiency of the two parameters α and k on the experimental results. The method we use is to fix one parameters and use the other parameter as a variable to test the running time of the algorithm. In figure 4, we set α=8, the runtime increases first and then decreases. In figure 5 we set k=8, the runtime decreases with a larger input of α, this is because the community size become fewer with a large k.

| Dataset | αkCA | Degree | Random |
|---------|------|--------|--------|
| DBLP    | 1314.050 | 5,362.414 | 1245.522 |
| Youtube | 1157827  | 2,987.624 | 2090338  |
6. Conclusions
In this paper, the problem of \((α, k)\)-NP-community disintegration is studied for the first time on the signed graph, i.e., finding the user node that damage the community most severe. We proved that the collapsed \((α, k)\)-NP-community problem is NP-hard. We proposed an algorithm for finding \((α, k)\)-NP-community, a LSA algorithm, and two collapsed algorithms, named, \(αkGCA\) algorithm and \(αkCA\) algorithm. Based on the \(αkGCA\) algorithm, we propose two lemmas to accelerate the search speed and reduce the number of candidate collapsed users. Our experiments on four real datasets demonstrate the effectiveness and efficiency of the algorithm.

References
[1] Malliaros F D, Vazirgiannis M. (2013) To stay or not to stay: modeling engagement dynamics in social graphs[J].
[2] Wu S, Sarma A D, Fabrikant A, et al. (2013) Arrival and departure dynamics in social networks[C]// Acm International Conference on Web Search & Data Mining. ACM.
[3] Bhawalkar K, Kleinberg J, Lewi K, et al. (2012) Preventing Unraveling in Social Networks: The Anchored $k$-$Core Problem[C]// Proceedings of the 39th international colloquium conference on Automata, Languages, and Programming - Volume Part II. Springer-Verlag.
[4] Seidman S B. (1983), Network structure and minimum degree[J]. Social Networks, 5(3):269.
[5] Cohen J. (2008) Trusses: Cohesive Subgraphs for Social Network Analysis[J]. National Security Agency Technical Report.
[6] Cheng J, Ke Y, Fu W C, et al. (2011) Finding Maximal Cliques in Massive Networks[J]. ACM Transactions on Database Systems, 36(4):21.
[7] Seidman SB (1983). Network structure and minimum degree[J]. Social Networks, 5(3):269.
[8] Abello J, Resende MG, Sudarsky S (2002) Massive quasi-clique detection, In: Latin American Symposium on Theoretical Informatics, Springer, pp. 598–612.
[9] Mokken RJ (1979) Cliques, clubs and clans. Quality & Quantity 13(2):161–173.
[10] Zhou R, Liu C, Yu J X, et al. (2012) Finding maximal k-edge-connected subgraphs from a large graph[C]// International Conference on Extending Database Technology.
[11] J.Luo, H.Molter, and O.Suchy, (2018) "A parameterized complexity view on collapsing k-core,"CoRR,vol.abs/1805.12453.
[12] Bhawalkar K, Kleinberg J, Lewi K, et al. (2012) Preventing Unraveling in Social Networks: The Anchored $k$-$Core Problem[C]// Proceedings of the 39th international colloquium conference on Automata, Languages, and Programming - Volume Part II. Springer-Verlag.
[13] Fan Zhang, Wenjie Zhang, Ying Zhang, Lu Qin, Xuemin Lin: (2017) OLAK: An Efficient Algorithm to Prevent Unraveling in Social Networks. PVLDB 10(6): 649-660.
[14] Fan Zhang, Ying Zhang, Lu Qin, Wenjie Zhang, Xuemin Lin: (2017) Finding Critical Users for Social Network Engagement: The Collapsed k-Core Problem. AAAI : 245-251.
[15] Fan Zhang, Ying Zhang, Lu Qin, Wenjie Zhang, Xuemin Lin: (2018) Efficiently Reinforcing Social Networks over User Engagement and Tie Strength. ICDE : 557-568.
[16] Discovering Strong Communities with User Engagement and Tie Strength. DASFAA (1) 2018: 425-441
[17] D.Song and D.A.Meyer, (2015) “Recommending positive links in signed social networks by optimizing a generalized AUC,” in AAAI.
[18] D.Song and D.A.Meyer,and D.Tao, (2015) "Efficient latent link recommendation in signed networks" in KDD .
[19] J.Tang,C.C. Aggarwal,and H.Liu, (2016) “Recommendations in signed social networks,” in WWW .
[20] Ou-Yang L, Dai D Q, Zhang X F. (2015) Detecting protein complexes from signed protein-protein interaction networks[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 12(6):1-1.
[21] Bohua Yang, Dong Wen, Lu Qin, Ying Zhang, Lijun Chang, Rong-Hua Li: (2019) Index-Based Optimal Algorithm for Computing K-Cores in Large Uncertain Graphs. ICDE: 64-75
[22] Chen Zhang, Ying Zhang, Wenjie Zhang, Lu Qin, Jianye Yang: (2019) Efficient Maximal Spatial Clique Enumeration. ICDE: 878-889
[23] Boge Liu, Long Yuan, Xuemin Lin, Lu Qin, Wenjie Zhang, Jingren Zhou: (2019) Efficient (a, &bgr;)-core Computation: an Index-based Approach. WWW: 1130-1141