Physics Letters A 382 (2018) 872–879

Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla

A game theoretic algorithm to detect overlapping community structure in networks

Xu Zhou [a,b], Xiaohui Zhao [b], Yanheng Liu [c,d,*], Geng Sun [c]

[a] Center for Computer Fundamental Education, Jilin University, Changchun, Jilin 130012, China
[b] College of Communication Engineering, Jilin University, Changchun, Jilin 130012, China
[c] College of Computer Science and Technology, Jilin University, Changchun, Jilin 130012, China
[d] Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, Jilin 130012, China

ARTICLE INFO

Article history:
Received 20 October 2017
Received in revised form 24 January 2018
Accepted 25 January 2018
Available online 2 February 2018
Communicated by C.R. Doering

Keywords:
Game theory
Utility function
Community detection

ABSTRACT

Community detection can be used as an important technique for product and personalized service recommendation. A game theory based approach to detect overlapping community structure is introduced in this paper. The process of the community formation is converted into a game, when all agents (nodes) cannot improve their own utility, the game process will be terminated. The utility function is composed of a gain and a loss function and we present a new gain function in this paper. In addition, different from choosing action randomly among join, quit and switch for each agent to get new label, two new strategies for each agent to update its label are designed during the game, and the strategies are also evaluated and compared for each agent in order to find its best result. The overlapping community structure is naturally presented when the stop criterion is satisfied. The experimental results demonstrate that the proposed algorithm outperforms other similar algorithms for detecting overlapping communities in networks.

1. Introduction

A community is a group of nodes in a network that are densely connected with regard to the rest of the network and it is the most meaningful characteristic in networks [1]. Nodes in the community probably share common properties and play similar roles within the networks. In biological networks, the sets of biological components(proteins and metabolites) which are related to a particular function is considered as communities. In social networks people with similar interest could be organized as a community. Detecting communities in networks becomes more attractive than ever. Generally speaking, it is found that community detection can be categorized into disjoint and overlapping community detection according to the number of community to which the node belongs [2]. In reality overlapping community structure is ubiquitous, for example, a person can belong to different groups like family,friends,colleagues at the same time. Community detection can help us understand the hidden social structure of networks, and the detected overlapping nodes play an important role in the communication across the members of different modules, it can be helpful for understanding how the modules affect each other [3]. Therefore, the study of discovering overlapping communities is a critical research issue with wide applications. For e-commerce, identifying communities could often help to achieve more accurate targeting and better marketing results. It can be also used for link prediction, market recommendation and anomaly behavior analysis.

Game theory provides a systematic framework to model the competition and coordination among the players. It is provided as a satisfactory tool in various fields such as economics, biology, computer science and military. Methods of detecting communities by making use of game theory are discussed in Ref. [4]. In general, the process of community formation can be modeled as a game. Node in the network could be recognized as the rational selfish player of the game. There are a kind of coordination and competition among the players. Chen presented the first work that was a randomized algorithm to model the dynamics of community formation by using a game-theoretic approach [5]. In his framework, firstly, to initialize each node to a singleton community. Then randomly pick a node and perform the best action among (leave, join or change his membership with a community), when no node can improve itself, an equilibrium state is achieved and the game will stop. Specifically, the equilibrium state means that all the players have reached their maximize utility, that is to say, no player of the game can improve his utility by unilaterally deviating from current

* Corresponding author at: College of Computer Science and Technology, Jilin University, Changchun, Jilin 130012, China.
E-mail address: yhliuj@163.com (Y. Liu).

https://doi.org/10.1016/j.physleta.2018.01.036
0375-9601/

strategy. The strategy space at this state is the final communities of the given network.

Owing to the importance of identifying communities and availability of promising solutions in game theory, we try to detect overlapping communities based on game theoretic model in this paper. The proposed method is extension of Chen's work [5]. Each node is taken as an agent that wants to maximize its utility during the process of game. The utility function is composed of a gain and a loss function. When all the agents play their best action, their utility values will reach to a satisfying limit, then the game gets to an equilibrium state, finally the overlapping community structure is formed at this state. To get best action for each agent,we assume that each agent does the decision based on the effects from its neighbors. If the similarity between agent and one of the neighbors is maximum, the agent will more rely on the neighbor. The difference between ours and Chen's work is that we take two strategies instead of action set for getting new label of each agent. Here we give some core contributions of this paper as belows: (1) A game theory based overlapping community detection algorithm is proposed. (2)A gain function with similarity is presented. (3)New methods for each agent to update its label are designed during the game (4)there is no restrictions on the size or the number of clusters to be detected.

The reminder of the paper is organized as follows. The related work is reviewed and discussed in Section 2. Section 3 presents some definition and framework of the proposed game theory based algorithm. The experimental results and corresponding analysis are given in Section 4. Finally the conclusion of the study is stated in Section 5.

2. Related work

The problem of overlapping community detection is widely studied by researchers coming from many areas such as physics and computer science. Many methods have been proposed for identifying overlapping community structures in network. It can be roughly categorized into 5 categories consisting of clique percolation methods, local expansion and optimization methods, link partition methods, fuzzy clustering methods and agent-based methods [6].

With the development of research, game theory based approach that is in the class of agent-based algorithms has been applied to solve problem of community detection. Game theory is an abstract mathematical framework, and it puts emphasis on decision making of the scenarios, in which one player's decision can influence the decision of other players [7]. Some methods based on game theory have been investigated to detect disjoint communities [8,9]. Chen presented the first work that was a randomized algorithm to model the dynamics of community formation by using a game-theoretic approach [5]. They used the non-cooperative game theory to achieve the overlapping community results. Followed by Chen's work, some methods are presented to design different utility functions. For example, Alvari proposed a new framework based on game theory and made use of structural equivalence to detect overlapping community structures [10]. The approach incorporates Pearson correlation and neighborhood relation as similarity separately into game theory. Some methods are proposed to design different operators for nodes to improve accuracy. Crampes proposed a potential game so as to reassign the nodes of clusters to overlapping communities, and the method had two phases including of partitioning and reassigning the nodes [11]. But the drawback of this method is that the number of the clusters should be known first. A distributed model is proposed to find overlapping community structures in Ref. [12]. It is a local approach and it does well on small datasets. An iterative algorithm consisting of two phases is presented in Ref. [13]. The method proposed two complex communal decision based operators such as suggestion and eviction to help the model converge to a promising solution. A non-cooperative game is proposed such that the equilibrium conditions of the game correspond to the clusters in the graph [14]. Two convergent algorithms to compute the equilibrium is introduced and one of the algorithms can be parallelized. In addition, Jiang proposed a non-cooperative game theory with four actions to discover community structures in dynamic networks [15]. The interactions among nodes of network is modeled by Szezepanski in the form of an influence game [16]. They made a new similarity metrics to measure the similarity between nodes in the network. Apart from non-cooperative game based community detection methods, some methods using cooperative game are proposed. Basu proposed a maximal quasi-clique according to hedonic coalition game model [17]. Each player in this framework has three types of actions and the goal of this framework is to simulate the self-organizing property of the community. Shapley value in cooperative games is adopted to detect communities in networks in Ref. [18], and an iterative formula of computing the Shapley value is also proposed to increase efficiency. Zhou developed a coalitional game to solve problem of community detection in multirelational social networks [19]. They used a greedy agglomerative algorithm to merge the coalitions with fewer players based on the utility of coalition to be enhanced. Zhou proposed a coalitional form game framework to reveal the stable communities when the game reaches equilibrium state [20]. The algorithm can deal with identifying communities in directed network.

3. Algorithm

The model of overlapping community discovery in networks based on game theory is described in this section. Our objective is to design a game and try to divide the graph into some clusters at the end of the game. We do not explicitly place any restrictions on the size or the number of clusters to be detected. Before the proposed method is presented in detail, some basic definitions about the game theory are necessary to be introduced first.

3.1. Technical preliminaries

Consider an undirected graph $G = (V, E)$. Let E denote the set of edges and V represent the set of nodes in the graph. Detecting overlapping communities is to find a set of clusters $(C_1, C_2, ..., C_p)$, where the graph is composed of these clusters, but there are common nodes among some of these clusters. It can be represented as belows. $\bigcup_{1\le i\le p} C_i = G$ and $\exists i, j \in p, C_i \cap C_j \neq \emptyset$. The process of community formation can be effectively modeled as a game. To be noticed that, a game comprises a number of agents, a set of strategies for each agent and a utility that quantifies the payoff of each agent. Node in the network could be recognized as the rational selfish agent of the game. Each node will select a collection of communities that it would like to participate.

Definition 1 *(Strategy space).* The strategy of agent i noted as $s_i \subset [k]$ are the communities that it wants to participate. In fact it is the list of community labels that agent i belongs to. When $s_i = \emptyset$, it means that agent i do not select any community to belong to.

Definition 2 *(All strategy profile).* Strategy profile of all agents is defined as $S = (s_i, s_2, ..., s_n)$, which is a vector of community labels for all agents and s_i represents the strategy of agent i.

Definition 3 *(Utility function).* The utility of the ith agent $(i \in V)$ comprises a gain function $g_i(\cdot)$ and a loss function $l_i(\cdot)$. Every agent has a utility function $(u_1, u_2, ..., u_n)$, and $u(\cdot) = \sum_{i\in[n]} u_i(\cdot)$.