Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 29 (2012) 4244 – 4248

Procedia Engineering

www.elsevier.com/locate/procedia

2012 International Workshop on Information and Electronics Engineering (IWIEE)

# Study on Overlapping Community Structure

Cui Li*

*School of Information ,Xi'an University of Finance and Economics, Xi'an, China*

**Abstract**

Discovery and classification of overlap nodes in communities is an important topic in data mining. In this paper, a new method is proposed, it is based on fuzzy bargaining sets. To uncover overlap nodes, we assume that each node is a selfish player who selects communities to join by participation level base on her own utility measurement. We allow each player to select multiple communities, which naturally captures the concept of "overlapping communities". We conduct experiments on this method, and the results show that our algorithm is effective in discovering overlapping communities, the value in practice and academic study is demonstrated.



Keywords: Overlapping communities; fuzzy reactive bargaining set; fuzzy semireactive bargaining set; fuzzy Mas-Colell bargaining set

## 1. Introduction

Identifying the community structure is crucial to understand the structural and functional properties of the networks(M.E.J. Newman,2003).Many methods have been proposed to identify the community structure of complex network. The first kind of method produces a partition, generally, this kind of method is suitable for understanding the entire structure of networks, especially for the networks with a small size. The second kind of method aims to discover the vertex sets with a high density of edges. This kind of method is appropriate to find the cohesive regions in large scale networks (A.Clauset, M.E.J. Newman, C. Moore,2004; J.Duch, A.Arenas,2005).

In this paper, we present a novel overlapping community structures discovery algorithm base on fuzzy bargaining sets.. We model each node as a player trying to optimize her own utility by joining communities with participation level, which we define as a overlapping community formation game. This

* * Corresponding author. Tel.: 13096992918.
*E-mail address*: bxdlicui@gmail.com


doi:10.1016/j.proeng.2012.01.651

game naturally incorporates overlapping communities, because we allow each individual player to select multiple communities by her participation level as long as it could improve her utility.

## 2. Overlapping community formation game

In this section, we formally define overlapping community formation game. In our fuzzy network game, a community is a vector $s\in[0,1]^N$, The i-th coordinate $s_i$ of s is the participation level of player i in the overlapping community s. Instead of $[0,1]^N$ we will also write $F^N$ for the set of community on player set N. We denote by ei the community eN=(1,1, ⋯,1) is called the grand community and e∮ =(0,0, ⋯,0) corresponds to the empty community. We denote the set of all nonempty community by $F_0^N = F^N\backslash\{e∮\}$. For $s\in F^N$, we define the carrier of s by $car(s)=\{i\in N|s_i>0\}$ a proper community if $car(s)\neq N$. The set of proper communities on player set N is denoted by $PF^N$, and the set of nonempty proper communities on player set N by $PF_0^N$. A overlapping community formation fuzzy network game with player set N is a Map $v:F^N\rightarrow R$,with the property v(e∮)=0. The map v assigns to each community a real number, telling what such a community can achieve in cooperation. The set of overlapping community formation games with player set N will be denoted by $FG^N$.

We are given a directed fuzzy network $G_F^u(V,A;s\cdot u)$ with a set of nodes *V*, and a set of arcs $A$, an element in *V* may be called as a player. Each node chooses a collection of communities that it wants to join by her participation level, and where $s_a\cdot u_a$, it denotes the actual revenue from a unit of flow by corresponding participation level through arc $a$, $a\in A, s\in F^A$, $0\le s_a\le 1$. Let $s', t'\in V$ be the single source and sink nodes of $G_F^u$, and assume that each arc in $G_F^u$ has a unit flow capacity. We call the fuzzy network $G_F^u$ *simple* if $f_j=1$ for every $j\in A$ and each arc is owned by the all owners of one community by corresponding participation level. We do not impose any community on the sign of $s\cdot u$. In general, one may expect some of the components $s_a\cdot u_a$ to be negative(reflecting the cost of the flow), while others are positive(to account for the associate revenue).

Suppose that each arc in $G_F^u$ is owned by all owners of community by corresponding participation level. Then, one naturally encounters the problem, resulting from an optimal flow in $G_F^u$,among the arc's owners. Such a problem can be formulated as a overlapping community formation game.Formally, Let $G_F^u(V,A;s\cdot u)$ be a simple fuzzy network. For $car(s)\subseteq car(e^A), s, e^A\in F^A$ (i.e. $[0,1]^A$), denote by $G_s^u$ the subnetwork of $G_F^u$ induced by the arcs in car(s), that is, denote by $G_s^u$ the fuzzy network restricted to arcs whose owners belong to car(s). Define v(s) to be the optimal revenue (maximum with respect to the objective value vector $(s_a\cdot u_a; a\in car(s))$resulting from an optimal s' to t' flow in $G_s^u$. Clearly, the set function v and the vector $e^A$ define overlapping community formation game on $car(e^A)$, denoted by $\Gamma(G_F^u)=(e^A;v)$.

We propose that we shall interpret the natural overlapping communities, as a overlapping community formation game played by players. Every player has her intrinsic utility that associates with which communities she joins by participation level. The sheer goal every player aims to achieve is to maximize her own utility. The formation of overlapping communities is thus the joint result of each player by participation level.

An *imputation* for the overlapping community formation game is a vector $x=(x_1,x_2,\cdots,x_n)$ satisfying

(1) $\sum_{i\in N} x_i = v(e^N)$; (2) $x_i\ge v(e^i)$, $\forall i\in N$

The preimputation set I*(v) of v∈FGN is the set: $$I^*(v)=\left\{x\in R^N \middle| \sum_{i\in N} x_i = v(e^N)\right\}$$

The imputation set of v∈FGN is the set: $$I(v)=\left\{x\in R^N \mid \sum_{i\in N} x_i = v(e^N) \quad and \quad x_i\ge v(e^i), \forall i\in N\right\}$$

The *Aubin core C(v)* of a community formation fuzzy network game $v \in FG^N$ is the set:

$$C(v) = \left\{ x \in I(v) \mid \sum_{i \in N} s_i \cdot x_i \geq v(s) \qquad \forall s \in F^N \right\}$$

That is, each player $i \in N$ is supposed to gain a payoff proportional to her participation level when she partially cooperates. Where for each community *s* the total payoff is not smaller than *v(s)*. so, if player *i* gains $x_i$ for full cooperation, then his gain will be $s_i \bullet x_i$ in case he participates at level $s_i$.This Aubin core can be interpreted into the overlapping community structure.

## 3. Overlapping community structure

The fuzzy bargaining sets of the fuzzy game can be interpreted into the overlapping communities structure(Yaron Azrieli, Ehud Lehrer,2007; Aumann RJ,Maschler M.,1964).

### *3.1. Fuzzy reactive bargaining set*

Let $x$ be an imputation of the overlapping community formation game $v$ .Then, a *fuzzy objection* of player $k$ against player $l$ with respect to $x$ is a pair *(y,q)* ,where $q \in F^N$ with $q_k > 0, q_l = 0$ and $y$ is a vector whose indices are the members of *car(q)* such that

$$q_i \cdot y_i \geq q_i \cdot x_i \text{,for all } i \in car(q) \text{ ; } \sum_{i \in car(q)} q_i \cdot y_i = v(q).$$

where for all $car(s) \subseteq N$ and $u \in R^N, u(s) - \sum_{j \in car(s)} s_j \cdot u_j$. We further say that $k$ has a *justified objection* against $l$ at $x$ ,if for every subset *car(s)* containing $l$ and not $k$ , there is an objection *(y,q)* of $k$ against $l$ and $x$ such that there does not exist a vector $z$ , indexed in *car(s)* ,satisfying

$$s_i \cdot z_i \geq s_i \cdot x_i \text{,for all } i \in car(s) \cap car(q) \text{; } s_i \cdot z_i \geq s_i \cdot x_i \text{, for all } i \in car(s) \setminus car(q) \text{ and } \sum_{i \in car(s)} s_i \cdot z_i = v(s).$$

The fuzzy *reactive bargaining set* for the community is the set, $M_F{}^r(v)$ ,given by

$$M_F{}^r(v) = \left\{ x \in I(v) \middle| \, for\, ordered \;\; pair(k,l) \;\; , k\, does \;\; not\, have \;\; a\, justified \;\; objection \;\; against \;\; l \right\}$$

In the definition of the reactive fuzzy bargaining set,the objecting player is allowed to retain the specifics of his/her objection until the objected player has announced his/her defending community. This way the objector is able to react to the move of his/her opponent,thus it becomes easier to raise justified objections. Intuitively, player i has a justified objection against player j with respect to x ,if he can demonstrate that j can not sustain any community, containing player j and not i , and still maintain her current payoff, $q_j \cdot x_j$. Conversely, player i does not have a justified objection against player j ,with respect to x ,if player j can sustain at least one community, containing j but not i ,and maintain her current payoff, $q \in F^N$, no matter how i forms his community and distributes the resulting proceeds. In the later case, a community which j uses to maintain her current payoff is called protecting community for player j .

### *3.2. Fuzzy semireactive bargaining set*

Let (N,v) be a overlapping community formation game, $x \in R^N$ ,and $k, l \in N$ be distinct players.Define the collection $\Gamma_{kl}(N)$ by $\Gamma_{kl} = \Gamma_{kl}(N) = \{car(s) \subseteq car(e^{N \setminus \{l\}}) | k \in car(s)\}$

*Hence*, $\Gamma_{kl}$ is the set of communities containing $k$ and not containing $l$ . A *fuzzy objection* of $k$ against $l$ at $x$ is a pair *(y,p)* , $car(p) \in \Gamma_{kl}, y \in R^{car(p)}$ ,satisfying

$$\sum_{i \in car(p)} p_i \cdot y_i = v(p). and \;\; p_i \cdot y_i \geq p_i \cdot x_i \text{,for all } i \in car(p),$$

If (y,p) has these properties, then we say that $k$ is able to object against $l$ via community $p$. Note that $k$ is able to object against $l$ via $car(s) \in \Gamma_{kl}$, if and only if the excess e(s,x)=v(s)- $\sum_{i \in car(s)} s_i \cdot x_i$ is strictly positive.

A *fuzzy counter objection* to an fuzzy objection *(y,p)* of $k$ against $l$ at $x$ is a pair *(z,q)*, $car(q) \in \Gamma_{kl}, z \in R^{car(q)}$,satisfying

$$\sum_{i \in car(q)} q_i \cdot z_i = v(q);\ z \geq x_{car(q)} \text{ and } z_{car(p)\cap car(q)} \geq y_{car(p)\cap car(q)}$$

If (z,q) has *these* properties, then we say that $l$ is able to counter (y,p) via community $q$. Note that $l$ is able to counter (y,p) via $car(q) \in \Gamma_{kl}$, if and only if $e(q,x) = \sum_{i \in car(p)\cap car(q)} (p \Lambda q)_i \cdot y_i - \sum_{i \in car(p)\cap car(q)} (p \Lambda q)_i \cdot x_i$.

The *fuzzy semireactive prebargaining set* $M_F^{sr*}(v)$ of a overlapping community formation game *(N,v)* is the set of all preimputations $x \in I^*(v)$ that satisfy the following condition for any pair of distinct players $(k,l) \in N \times N$ for any $car(q) \in \Gamma_{kl}$:

There is $car(q) \in \Gamma_{kl}$ such that any objection of $k$ against $l$ via community $p$ can be countered by $l$ via community $q$.

The *fuzzy semireactive bargaining set* of (N,v) is defined to be the set $M_F^{sr}(v) = M_F^{sr*}(v) \bigcap I(v)$ of *individually* rational elements of the fuzzy semireactive prebargaining set.

### *3.3. Fuzzy Mas-Colell bargaining set*

Let $x$ be an imputation of the overlapping community formation game $v$. Then, an weak fuzzy objection of player $k$ against player $l$ with respect to $x$ is a pair (y,s), where $s \in F^N, s_k > 0, s_l = 0$, and y is a vector whose indices are the members of car(s) such that

$s_i \cdot y_i \geq s_i \cdot x_i$,for all $i \in car(s)$,With at least one strict ($y_i > x_i$); $\sum_{i \in car(s)} s_i \cdot y_i = v(s)$.

Let $x$ be an imputation of the overlapping community formation game $v$ and let (y,s) be an objection of player $k$ against player $l$ with respect to $x$.

Then, (z,t) is a strong fuzzy counter-objection of player $l$ to the objection (y,s) of player $k$ if $t \in F^N, t_k > 0, t_l = 0$, and $z$ is a vector whose indices are the member of car(t) such that

$t_i \cdot z_i \geq t_i \cdot x_i, \forall i \in car(t) \setminus car(s)$; $t_i \cdot (z_i - x_i) \geq s_i \cdot (y_i - x_i), \forall i \in car(t) \bigcap car(s)$ and $\sum_{i \in car(t)} t_i \cdot z_i = v(t)$.

The Mas-Colell fuzzy bargaining set $MC_F(v)$ of the overlapping community formation game (N,v) is defined as the set of imputations at which every objection between any two players can be countered. Formally,

$$MC_F(v) = \left\{ x \in I^*(v) \middle| \begin{array}{l} \text{every weak fuzzy objection} \\ \text{at x can be strongly countered.} \end{array} \right\}$$

**Corollary 1.** In every simple overlapping community formation game,

$$C_A(v) \subseteq M_F{}^r(v) \subseteq M_F{}^{sr}(v) \subseteq MC_F(v)$$

## 4. Overlapping community structure discovery based on fuzzy bargaining sets

We propose that player's choice depends on her current state when he needs to respond to the other players' one. A player can only implement the following three operations.

We provide one instance of experiment which demonstrates our observation is reasonably accurate. Our experiment searches for the node with one name in the co-authorship graph, which in fact represents more than 20 individuals that have published in total more than 100 papers in computer science or relevant areas. We use only a subgraph of the co-authorship graph that contains 16 nodes because processing the whole graph would otherwise be too computationally intensive. The subgraph is obtained by using breadth first search from the given node until 16 nodes are discovered. Our algorithm discovers two communities containing the given name, Fig 1 summarizes the interrelation of co-authors in the two communities.

| **Algorithm** of overlapping community structure discovery based on fuzzy bargaining sets |
|---|
| Step 1: *Join.* Player $v_i$ joins a new community on top of the community he joins by the participation level $s_i$.; <br> Step 2: *leave.* Player $v_i$ leaves a community she is in by participation level 0.; <br> Step 3: *switch.* Player $v_i$ switches from one community to another by changing participation level $s_i$. |

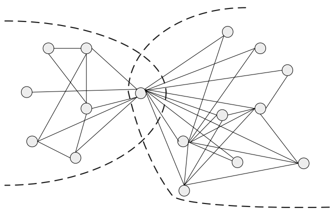

Figure 1. The two communities structure discovered by the fuzzy Mas-Colell bargaining set

## 5. Conclusion

In this paper, we present a method to discover the overlapping community structures in networks. The algorithm is mainly based on fuzzy bargaining sets. Our experiment show that, the resulting community structure naturally incorporate overlapping communities.

In future, we will do further analysis in the overlapping problem and try to find the overlapping community structures from fuzzy game view.


## Acknowledgment

This paper was supported by Teaching Research Project(No.10xcj23) and Scientific Research Fund Project (No.11XCK11) of Xi'an University of Finance and Economics . Furthermore, this work was supported by project 09JK440 and 08JK290, Research of Nature Science of Education Department of Shaanxi Government, and Research Project SJ08F32, Foundation of Nature Science of Shaanxi Province.


## References


[1] M.E.J. Newman. The Structure and Function of Complex Networks. SIAM Rev. 45(2003)167-256.

[2] A.Clauset, M.E.J. Newman, C. Moore. Finding Community Structure in Very Large Networks. Phys. Rev. E79(2004)066111.

[3] J.Duch, A.Arenas. Community Detection in Complex Networks Using Extremal Optimization. Phys. Rev. E72(2005)027104.

[4] Yaron Azrieli, Ehud Lehrer.On Some Families of Cooperative Fuzzy Games. Int J Game Theory,(2007)36:1-15.

[5] Aumann RJ,Maschler M, The Bargaining Set for Cooperative Games.In:Dresher M,Shapley LS,Tucker AW(eds)Advances in game theory.Princeton University Press,Princeton, 1964,pp.443-476.