Social Networks 43 (2015) 177–189

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Fuzzy duocentric community detection model in social networks

Samira Malek Mohamadi Golsefid [a], Mohammad Hossein Fazel Zarandi [a,*], Susan Bastani [b]

[a] Department of Industrial Engineering, Amirkabir University of Technology (Polytechnic of Tehran), P.O. Box 15875-4413, Tehran, Iran
[b] Department of Sociology, Alzahra University, 19938-93973 Tehran, Iran

ARTICLE INFO

Keywords:
Duocentric networks
Community detection
Overlapping community
Center-based clustering
Type-2 fuzzy clustering
Dual center clustering

ABSTRACT

The main goal of this paper is to present a clustering model to identify duocentric communities in the complex networks. A duocentric community is built around two central nodes which are as close as possible to other nodes, while the central nodes are connected enough to each other to shape the center of the community. To detect such communities, we develop a new objective function based clustering model. The network's nodes are assigned to the duocentric communities by the type-2 fuzzy numbers which indicate the degrees of belonging to the communities by upper and lower membership values. Generated interval type-2 fuzzy membership values by our proposed model are able to determine how much each node belongs to the both central nodes and how it is shared among communities. Also, the compatible verification index with the proposed model is introduced to evaluate and compare the results of the proposed model with the existing approach in the literature. Finally, the performance of the proposed algorithm is validated by detecting duocentric communities in three artificial networks and two real social networks.

1. Introduction

A network is a set of nodes (or objects) and edges (or relations) which describe the relationships between the nodes (Kadushin, 2004; Schaeffer, 2007). A group of nodes which probably share common properties and/or play similar roles within the network is called a community or cluster (Estrad et al., 2012; Fortunato, 2010). The task of grouping nodes with edges which are connected to one another but have no connection to outside the group is referred to as graph clustering or community detection (Fortunato, 2010; Schaeffer, 2007). Networks' nodes may be shared among different communities and form overlapping communities. Discovering and detecting overlapping communities which exist in the most real social networks is an important topic in social network analysis (Leskovec et al., 2008; Li, 2012).

The fuzzy clustering model is one of the methods in the literature which were used for detecting overlapping community. The fuzzy objective function-based clustering method comprises a family of local graph clustering methods that can be formulated as the problem of minimizing an objective function. These methods detect the overlapping communities and assign the nodes to communities with different degrees of belonging. The node's degree of belonging to the communities are measured based on the distance to the center of the communities.

The structure of the detected communities could be considered as egocentric networks which center of the community is an ego and the other community members are the alters connected to the ego. However, in some cases, there are two egos play the main role in the network. This structure of the community which forms around a pair of central nodes is called duocenteric community which was introduced by Coromina et al. (2008) in 2007. Detecting the communities in these cases are required to find two central nodes rather than one central node. Consequently, the degree of belonging to such communities should be determined based on the distances to both central nodes rather than a single node.

There is not any community detection model in the literature for detecting duocentric communities and defining the degree of belonging to such communities. The goal of this paper is to present a new model for detecting duocentric communities in the complex networks to detect overlapping communities with two central nodes. In addition, we defined the degree of belonging to such communities by type-2 fuzzy membership value which indicates the degree of belonging to both central nodes as well as degree of sharing nodes among the communities.

This paper is organized as follow: in the next section, the basic concepts of the network structure and the community detection models are addressed. Section 3 describes the proposed fuzzy

* Corresponding author. Tel.: +98 2164545378.
E-mail addresses: samira.malek@gmail.com (S.M.M. Golsefid), zarandi@aut.ac.ir (M.H. Fazel Zarandi), bastani_susan@yahoo.com (S. Bastani).

http://dx.doi.org/10.1016/j.socnet.2015.04.009

duocentric community detection model and a verification index to evaluate and compare the results of the proposed model with existing approach. Section 4 describes the experimental results and finally, the last section concludes the paper with further research suggestions.

2. Background

This section describes some basic concepts of the network structure and reviews the community detection approaches. It also reviews approaches which applied fuzzy logic theory for detecting overlapping communities.

2.1. Network structure

The relations between the nodes represent the different network concepts and, consequently, network structures. In the literature, the network structure mainly divided into the complete and egocentric network. A complete (or whole) network is built upon every node from a population and any relation is considered for every node composing the network (Martino and Spoto, 2006). Since all relations between the set of networks' nodes are considered, analysis of complete network may face challenges due to limitations of either accessing data of the whole network or computational complexity (Almquist, 2012). This impracticality of using a whole network could be solved by considering network as an egocentric network. In the anthropology literature, egocentric network (also known as a local or personal network) is used to represent the social network between an ego (also known as a focal person or central node) and the alters (other people or non-central nodes) who have some predefined social relationships with ego (Almquist, 2012; Everett and Stephen, 2005; Passarella et al., 2012; Wasserman and Faust, 1994). The ego is the only center or hub in its egocentric network, and is connected to all its alters or non-central nodes (Passarella et al., 2012). If k nodes are considered in a network, egocentric network defines a specific class of networks that considers a network which has one point as ego or center and it is connected to each $k-1$ nodes. However, in a complete network all relations between $k-1$ nodes are also considered. While an egocentric network has the fewest possible links, a complete network has the maximum number of relations in the network (Freeman, 1982).

The egocentric network is one of the most important representations of human social networks (Passarella et al., 2012) and is of interest for a number of reasons (Newman, 2003). Egocentric network models do not describe all social links between alters which would be required to model the complete social network formed by all egos and their alters (Passarella et al., 2012). Therefore, the collecting of data compared with whole networks is easier. Information on the alters, including how they are connected, is obtained entirely from the ego (Everett and Stephen, 2005). In addition, in some cases, egocentric network mapped from a large population can be used to make statistically significant conclusions about the whole population (Everett and Stephen, 2005; Freeman, 1982) and it has been widely explained and studied in the literature of social network analysis (Burt, 1992; Coleman, 1990; Everett and Stephen, 2005; Granovetter, 1973; Knoke and Kuklinski, 1982; Newman, 2003; Scott, 2000; Vehovar et al., 2008; Wasserman and Faust, 1994; Wellman and Berkowitz, 1988).

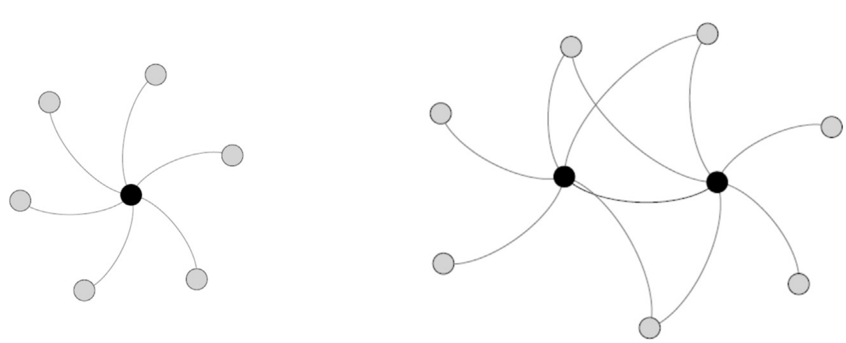
The egocentric network is used in network analysis when a single node play the main role in the network. However, in some cases, instead of one node two nodes play the main role in the network. Coromina et al. (2008) have defined the duocentric networks for the case in which the center of the network is a pair of relevant nodes. Detecting communities in such networks requires identifying the two main important actors within a network. Networks which are built around a husband and wife, buyer and seller, exporter and importer, a PhD student and his/her supervisor are good examples of such networks. In a duocentric network, the relations of two central nodes with other nodes in the network are considered, while the relations amongst non-central nodes are neglected (Coromina et al., 2008). An example of an egocentric and a duocentric community are shown in Fig. 1.

A duocentric network is a compromise between an egocentric network and a complete network that can be used when there is a pair of relevant central actors in the network. Similar to egocentric network, duocentric network models do not considered all social relations between alters which would be required to model the complete social network formed by all duocenters and their alters. Although the number of considered connections in duocentric networks are doubled compared to egocentric networks, it is still significantly less than complete networks.

Duocentric communities are made up of social units (like persons, groups or organizations) and social relations (like marriage or friendship) in which nodes are connected to at least one of the two central egos. Duocentric networks, however, are special in that they are built around a pair of particular designated social units. To uncover duocentric network, first, a pair of nodes is detected as the center of network. Then, other non-central nodes are determined by their relations to duocenters.

Defining a duocentric network by only one center, can result in misunderstanding of the whole network. Fig. 2(a) shows a sample of a duocentric network (central nodes are shown in blue). Defining the network by only one center causes neglecting the gray nodes, as shown in Fig. 2(b) and (c); whereas the gray nodes would be also belong to the networkif the two nodes are considered as center, as shown in Fig. 2(d).

In this paper, we focus on the network with duocentric structure and develop a new model to detect communities (or sub-networks)



(a) (b)

Fig. 1. An example of (a) egocentric and (b) duocentric network.

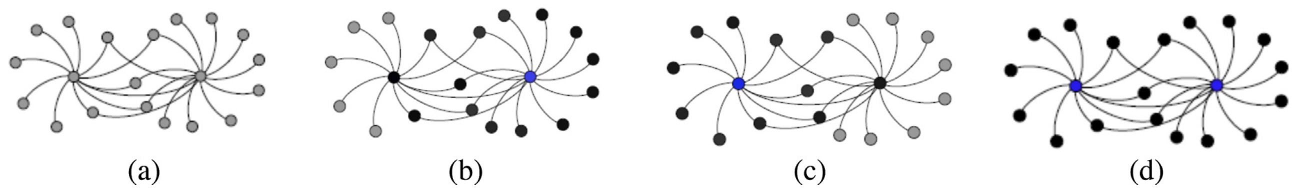


Fig. 2. (a) A sample network when its center was considered, (b) right central node as single center, (c) left central node as single center, and (d) as two nodes. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

with such structure. In the following subsection, the existing approaches to community detection will be reviewed.

2.2. Community detection models

A group of nodes, which probably share common properties and/or play similar roles within the network, form a community (or a cluster) (Estrad et al., 2012; Fortunato, 2010); relations within the community are greater than relations between the community's members with other nodes which are not members (Kadushin, 2004). If data is represented as a network whose nodes are objects and links represent connections among objects, then a community or cluster can be defined as a connected component like a group of nodes that are connected to one another, but which have little connection to nodes outside the group. Generally, the task of grouping nodes with edges within each community and relatively few edges between the communities is called community detection or graph clustering (Fortunato, 2010; Schaeffer, 2007). The graph clustering methods in the literature are divided into two main groups. The first group includes global methods such as hierarchical clustering, divisive clustering and agglomerative clustering, while the second group is comprised of local clustering methods, such as local search and fitness function methods. In a global clustering, each node of the input network is assigned to a cluster in the output of the method, whereas in a local clustering the cluster assignments are only done for a certain subset of nodes, commonly only one node (Schaeffer, 2007). Center-based graph clustering algorithms are a class of local methods and they create a one-level partitioning of data objects and attempt to minimize the distance between nodes labeled in a particular community and a point designated as the center of that community. The center-based graph clustering objective function is as follows:

$$J = \sum_{i=1}^{c}\sum_{k=1}^{g} D_{ik} \tag{1}$$

where c is the number of communities, D_{ik} is the distance between $n_{i'}$ and n_k, and $n_{i'}$ is the center of community i. D_{ik} is defined as $D_{ik} = \sum_{j=1}^{g} \left| a_{i'j} - a_{jk} \right|$, a is an entry of adjacency matrix A with size of $g \times g$ (g is the number of network nodes) for one mode network. The entry in the adjacency matrix, a_{ij}, records which pairs of nodes are adjacent. In the adjacency matrix, if nodes n_i and n_j are adjacent, then $a_{ij} = 1$, and if nodes n_i and n_j are not adjacent, then $a_{ij} = 0$ (Wasserman and Faust, 1994). In this paper, we are focusing on networks whose links are not directed and are neither signed nor valued. If a link between two nodes is present, it goes both from n_i to n_j and from n_j to n_i, thus, $a_{ij} = 1$ and $a_{ji} = 1$. In other words, the adjacency matrix for a non-directional relation network is symmetric, thus $a_{ij} = a_{ji} = 1, \forall\, i, j$ (Wasserman and Faust, 1994). In this approach, the node n_k is assigned to the community with the minimum distance to the center of the community, $n_k \in G_i : \min_{i=1,\ldots,g_i} D_{ik}$.

In many real networks, nodes may be shared among different communities rather just being part of a specific community. Next subsection reviews the overlapping community detection models.

2.3. Overlapping community detection models

Communities are not always well separated in the real networks and the nodes may be shared among different communities and formed overlapping communities. Fig. 3(a) shows two well-separated communities. However, there is overlapping between two communities in Fig. 3(b) and two nodes are shared between two communities. Discovering and detecting overlapping communities which exist in the most real social networks is an important topic in social network analysis (Leskovec et al., 2008; Li, 2012) and the number of studies in this area, either heuristic or local search procedures, have recently increased (Fortunato, 2010; Fu et al., 2013; Santo, 2010; Wanga et al., 2009; Xie et al., 2011). The intuition behind overlapping community detection models is based on the fact that the real complex networks are not usually divided into sharp sub-networks or communities; instead, nodes may naturally belong to more than one community (Malliaros and Vazirgiannis, 2013). One of the approaches for detecting overlapping communities is fuzzy logic theory. In recent studies, methods have been developed based on fuzzy relations and theory (Dave and Krishnapuram, 1997; Li, 2012; Schaeffer, 2007; Sun et al., 2011; Yan

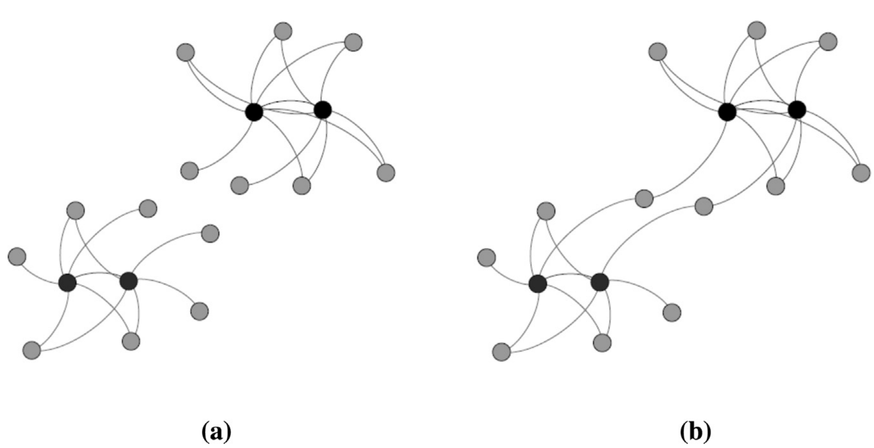


Fig. 3. A sample of (a) duocentric communities and (b) overlapping duocentric communities.

and Hsiao, 1995) and fuzzy c-means (FCM) clustering model (Zhang et al., 2007).

Despite the fact that fuzzy clustering was developed and applied widely to general clustering tasks, little research can be found on fuzzy clustering in graph clustering (Schaeffer, 2007). In general, the past decade has been quiet concerning the application of fuzzy graph clustering in this area (Achlioptas et al., 2005; Wanga et al., 2013). Although some methods for discovering fuzzy overlapping communities have been presented recently, there is still space for improving their performance and universality (Schaeffer, 2007; Wanga et al., 2013). Fuzzy center-based clustering methods comprise a family of local graph clustering methods that can be formulated as the problem of minimizing an objective function. These methods assign the nodes to communities with different degree of belonging values and form overlapping communities. The most popular fuzzy center-based clustering model is fuzzy c-means (FCM) clustering and is mostly used in combination with other techniques for community detection (Jiang et al., 2009; Liu, 2010; Zhang et al., 2007). FCM clustering is the most well-known fuzzy clustering algorithm proposed by Dunn (1974) and extended by Bezdek (1981). It allows an object to belong to several clusters with different membership values and defines the degrees of belonging to the clusters by the type-1 fuzzy number as a crisp value over the interval $[0, 1]$ (Gath and Geva, 1989). FCM partitions the data set into c clusters by minimizing the following evaluation function:

$$J = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^{m} D_{ik} \tag{2}$$

where $1 \le m < \infty$ is the fuzzifier parameter, $D_{ik} = \sum_{j=1}^{g} \left| a_{i'j} - a_{jk} \right|$ and $u_{ik} \in [0, 1]$ is the type-1 fuzzy membership degree of pattern $k = 1, ..., n$ in cluster $i = 1, ..., c$. To reduce the effect of outliers, Krishnapuram and Keller relaxed the condition of sum of the membership values to all clusters for each point is equal to 1 and introduced the Possiblistic c-means (PCM) (Krishnapuram and Keller, 1993). Golsefid et al. (2015) have extended PCM model for detecting overlapping communities as follows:

$$J = \sum_{i=1}^{c}\sum_{k=1}^{g} u_{ik}^{m} D_{ik} + \sum_{i=1}^{c} \Delta_i \sum_{j=1}^{g} (1 - u_{ik})^{m} \tag{3}$$

where, Δ_i is the density of cluster i. Updating formulas for fuzzy membership values and cluster center are:

$$u_{ik} = \left(1 + \left(\frac{D_{ik}}{\Delta_i} \right)^{1/(m-1)} \right)^{-1} \tag{4}$$

$$n_{i'} = \min_{i' \in [1, g_i]} \sum_{j=1}^{g} a_{i'j} - \sum_{k=1}^{g} u_{ik}^{m} \sum_{j=1}^{g} a_{jk} / \sum_{k=1}^{g} u_{ik}^{m} \tag{5}$$

The structure of the community in all mentioned community detection models is considered as egocentric community and it forms around a single node. However, communities could be formed around two central nodes instead of a single node as we discussed earlier in Section 2.1. The existing models cannot detect duocenteric communities. In this study, a new clustering method is defined which is able to detect communities with two central nodes. In the following section, a new approach to detect duocentric community will be introduced.

3. Proposed model

In this section, we introduce the proposed "Fuzzy Duocentric Community Detection Model" to detect overlapping duocentric communities in the complex networks. First, we define the center of community as duocenter with two central nodes. The membership values of nodes are determined by using the type-2 fuzzy numbers which indicates: (1) how much each node belongs to the both central nodes and (2) how it is shared among communities. Also, the compatible verification index with the proposed model is introduced to evaluate and compare the results of the proposed model with the existing approach in the literature.

3.1. Defining the duocenter of duocentric community

The main characteristic of a duocentric community is building community around a pair of central nodes. In this case, as shown in Fig. 1(b), we attempt to find a pair of central nodes which both nodes have high similarity with other non-central nodes and are also relevant and similar enough to each other to build a community. The similarity between central nodes and non-central nodes are measured based on the degree of central node. The degree of central node $n_{i'}$ is equal to the number of connected nodes to $n_{i'}$:

$$P(n_{i'}) = \sum_{k=1}^{g_i} a_{i'k} \tag{6}$$

where $a_{i'k}$ indicates the relation between node k and central node $n_{i'}$. Similarly, the degree of central node $n_{i''}$ is equal to:

$$P(n_{i''}) = \sum_{k=1}^{g_i} a_{i''k} \tag{7}$$

where $a_{i''k}$ indicates the relation between node k and central node $n_{i''}$. The relation between two central nodes $n_{i'}$ and $n_{i''}$ is defined as:

$$P(n_{i'} \cap n_{i''}) = \sum_{k=1}^{g_i} \Phi_{i'i''} \tag{8}$$

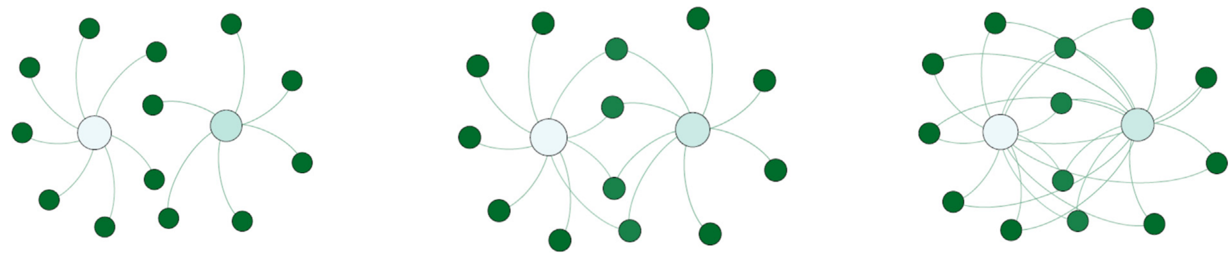
where, if $a_{i'k} = 1$ and $a_{i''k} = 1$, then $\Phi_{i'i''} = a_{i'k} \cdot a_{i''k} = 1$, and otherwise $\Phi_{i'i''} = 0$. Node $n_{i'}$ and $n_{i''}$ can build duocentric communities if they link to non-central nodes of community as much as possible, $\max(P(n_{i'}))$ and $\max(P(n_{i''}))$, and also have at least a minimum number of connections to the shared nodes, $P(n_{i'} \cap n_{i''}) > \Phi^{\min}$.

For example, consider a community with 12 nodes as shown in Fig. 4. By assuming $\Phi^{\min} = 3$, the communities in Fig. 4(b) and (c) are duocentric communities. However, Fig. 2(a) is not a duocentric community since $P(n_{i'} \cap n_{i''}) = 0 < \Phi^{\min}$. The value of $\Phi^{\min}$ determines the acceptable minimum number of shared nodes that should exist between the two central nodes. $P(n_{i'} \cap n_{i''}) \to 0$ led us to build two separated egocentric communities instead of one duocentric community. $\Phi^{\min}$ determined the minimum desired dependency of $n_{i'}$ and $n_{i''}$ which should be higher than the average degree of the network and less than $g_i - 2$; g_i is the number of a community's nodes.

3.2. Proposed fuzzy duocentric community detection model

In our proposed model, since a community is built around a pair of nodes, both central nodes are considered in defining the structure of the model and assigning nodes to the communities.

Suppose that $G(N, \ell)$ is a network consisting of: a set of nodes, $N = \{n_1, n_2, ..., n_g\}$ and a set of lines, $\ell = \{l_{1.}, l_{2.}, ..., l_{L.}\}$. The adjacency matrix of size $g \times g$ is denoted as A and entries in this adjacency matrix, a_{ij}, equal to 1 if nodes n_i and n_j are adjacent, and equal to



(a) (b) (c)

Fig. 4. An example of duocentric community with (a) $P(n_{i'} \cap n_{i''}) = 0$, (b) $P(n_{i'} \cap n_{i''}) = 4$, and (c) $P(n_{i'} \cap n_{i''}) = 13$.

0 if nodes n_i and n_j are not adjacent. The duocentric community detection objective function from (1) becomes:

$$J = \sum_{i=1}^{c}\sum_{k=1}^{g}(D_{ik}^1 + D_{ik}^2) \tag{9}$$

where $D_{ik}^1 = \sum_{j=1}^{g} w_j \left|a_{i'j} - a_{kj}\right|$ is the distance of node n_k to $n_{i'}$ and $D_{ik}^2 = \sum_{j=1}^{g} w_j \left|a_{i''j} - a_{kj}\right|$ is the distance of node n_k to $n_{i''}$. We also consider w_j in measuring distance which it indicates how much node j connects to other nodes and it is calculated as $w_j = \sum_{j'1}^{g} a_{jj'}/g$. $N^{c^{1,2}} = \{n_1^{c^{1,2}}, \ldots, n_c^{c^{1,2}}\}$ is a set of duocenters where $n_i^{c^1} = n_{i'}$, $n_i^{c^2} = n_{i''}$ and i', i'' are the index numbers of duocentric nodes of the community i.

A duocentric community is formed around two central nodes; therefore, the degree of belonging should be defined based on both central nodes. In this case, the type-1 fuzzy membership values are not able to determine the degree of belonging. However, the type-2 fuzzy number is able to express the nodes' degree of belonging to the duocentric community base on the interval values.

As shown in Fig. 5, the type-1 fuzzy membership values are between zero and one, whereas the type-2 fuzzy membership values are considered as type-1 fuzzy membership values themselves, $\tilde{u}_{ik} = [\underline{u}_{ik}, \bar{u}_{ik}]$. This interval type-2 fuzzy set is described as follows (Mendel and John, 2002):

$$\tilde{U} = \int_{x\in X}\int_{u\in J_x} 1/(x,u) = \int_{x\in X}\left[\int_{u\in J_x} 1/u\right]/x \tag{10}$$

where x is the primary variable, J_x, an interval in [0, 1], is the primary membership of x, u is the secondary variable, and $\int_{u\in J_x} 1/u$ is the secondary membership function (MF) at x (Mendel, 2007).

In the proposed model, we used the membership values as type-2 fuzzy numbers to measure how much each node belongs to the both central nodes and how it is shared among communities. The proposed model was defined based on a dual center fuzzy clustering model (Golsefid and Fazel Zarandi, 2015) and we adjusted it for detecting overlapping communities in the complex networks.

Now consider the problem of minimizing J by putting $\tilde{U}$ as degree of belonging to duocentric community in (9):

$$J(\tilde{U}, N^c) = \sum_{i=1}^{c}\sum_{k=1}^{g}(\underline{u}_{ik}^m D_{ik}^{\max} + \bar{u}_{ik}^m D_{ik}^{\min}) \tag{11}$$

where $m \in [1, \infty)$ is a fuzzy weighting exponent. Here, $\tilde{u}_{ik}$ denotes the kth column of $\tilde{U} \in m_f$, that is, $\underline{U}_k = (\underline{u}_{1k}, \ldots, \underline{u}_{ck})$ as lower membership values and $\bar{U}_k = (\bar{u}_{1k}, \ldots, \bar{u}_{ck})$ as upper membership values, $1 \le k \le g$. $D_{ik}^{\max}$ and $D_{ik}^{\min}$ are the minimum and maximum distance between node k and two central nodes of cluster i. Since it is not specified which distance from n_k to which central node $n_{i'}$ or $n_{i''}$ is maximum or minimum, we define $D_{ik}^{\max}$ and $D_{ik}^{\min}$ as follows:

$$\begin{aligned} D_{ik}^{\max} &= 0.5(D_{ik}^1 + D_{ik}^2 + \left|D_{ik}^1 - D_{ik}^2\right|) \\ D_{ik}^{\min} &= 0.5(D_{ik}^1 + D_{ik}^2 - \left|D_{ik}^1 - D_{ik}^2\right|) \end{aligned} \tag{12}$$

where, $D_{ik}^1 = \sum_{j=1}^{g} w_j \left|a_{i'j} - a_{kj}\right|$, $D_{ik}^2 = \sum_{j=1}^{g} w_j \left|a_{i''j} - a_{kj}\right|$ and $w_j = \sum_{j'=1}^{g} a_{jj'}/n$. Therefore, $D_{ik}^{\max} \ge D_{ik}^{\min}$, and consequently $\underline{u}_{ik} \le \bar{u}_{ik}, \forall i, k$.

Eq. (11) is minimization function. For not assigning membership values to 0 in minimization function J, Δ is considered as the density of community. The proposed objective function is formulated as:

$$\begin{aligned} J(\tilde{U}, N^c, \Delta) = &\sum_{i=1}^{c}\sum_{k=1}^{g}(\underline{u}_{ik}^m D_{ik}^{\max} + \bar{u}_{ik}^m D_{ik}^{\min}) + \sum_{i=1}^{c}\Delta_i\sum_{j=1}^{g}((1-\underline{u}_{ik})^m \\ &+ (1-\bar{u}_{ik})^m) \end{aligned} \tag{13}$$

where Δ_i is the density of cluster i, which its value will be discussed later. The first part minimizes the distance from the duocenters as much as possible, and the second part forces $\tilde{u}_{ik}$ to be as large as possible, thus avoiding the trivial solution.

Theorem. *If $D_{ik}^1 > 0$ and $D_{ik}^2 > 0$ for all i and k, and $G(N, \ell)$ is a network consisting of: a set of nodes, $N = \{n_1, n_2, \ldots, n_g\}$ and a set of lines, $\ell = \{l_{1.}, l_{2.}, \ldots, l_{L.}\}$. Then $\tilde{U}$ may be a global minimum for $J(\tilde{U}, N^c, \Delta)$ only if updating formulas for type-2 fuzzy membership values are:*

$$\begin{aligned} \underline{u}_{ik} &= \frac{1}{1 + (D_{ik}^{\max}/\Delta_i)^{1/(m-1)}} \\ \bar{u}_{ik} &= \frac{1}{1 + (D_{ik}^{\min}/\Delta_i)^{1/(m-1)}} \end{aligned} \tag{14}$$

and the duocenters of cluster i are defined as:

$$\begin{aligned} n_{i'}, n_{i''} = \arg_{i',i''}\min &\left\{\sum_{j=1}^{g} w_j(a_{i'j} + a_{i''j}) - \sum_{k=1}^{g}\underline{u}_{ik}^m\sum_{j=1}^{g} w_j a_{kj}\Big/\sum_{k=1}^{g}\underline{u}_{ik}^m\right. \\ &\left. - \sum_{k=1}^{g}\bar{u}_{ik}^m\sum_{j=1}^{g} w_j/a_{kj}\Big/\sum_{k=1}^{g}\bar{u}_{ik}^m\right\} \end{aligned} \tag{15}$$

where $\left\{n_{i'}, n_{i''}\right\} \in g^D$, $g^D = \{n_{i'}, n_{i''} | P(n_{i'} \cap n_{i''}) > \Phi^{\min}\}$.

The proof of theorem will be presented in Appendix A.

Based on our experiments, we suggest that the values of Δ_i could be considered as a constant positive number or estimated as the proportion of common relations between duocenters to all the links presented in the network and calculated as:

$$\Delta_i = \frac{P(n_{i'} \cap n_{i''})}{g} \tag{16}$$

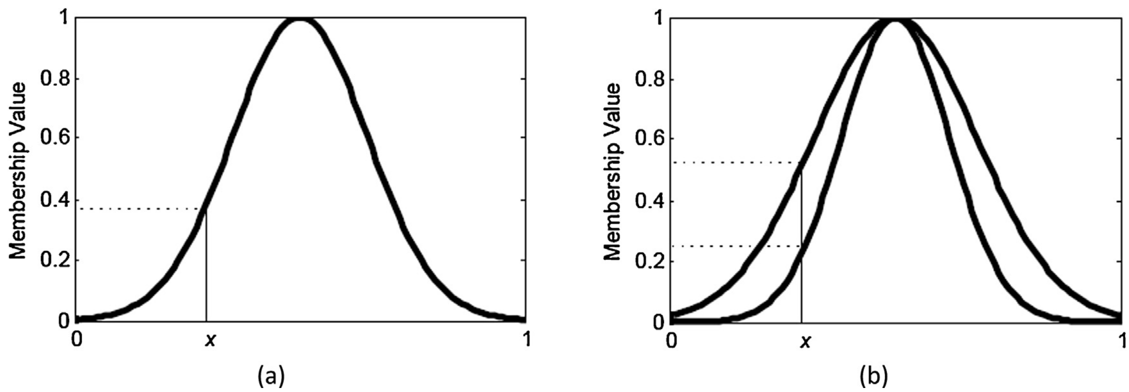


Fig. 5. (a) Type-1 and (b) type-2 fuzzy membership values.

The Δ_i goes from 0, if there isno relation between duocenters, and goes to 1, if all graph's nodes have relation with both central nodes. The higher value of Δ_i shows the greater number of connections from duocenters to other non-central nodes and more dense clusters.

Fig. 6 illustrates the new fuzzy duocentric community detection algorithm. After calculating the degree of each node and the distance matrix, all possible pairs of nodes which $P(n_{i'} \cap n_{i''}) > \Phi^{\min}$ are put in set g^D. Therefore, instead of searching all nodes for determining the duocenters, only members of this set are evaluated for redetermining the duocenters in each algorithm iteration.

The necessary condition of the convergence of the proposed algorithm in Fig. 6 is met when:

$$\lim_{t\to\infty} \left\| \tilde{U}^{(t)} - \tilde{U}^{(t-1)} \right\| = 0 \tag{17}$$

The proof of this condition and the proposed algorithm convergence will be presented in Appendix B.

3.3. Proposed verification index

In this subsection, the compatible verification index with the proposed model is defined to establish sufficient confidence in the proposed model. The proposed model is verified by evaluating the technical correctness of model equations (Coyle and Exelby, 2000). The objective function of the proposed fuzzy duocentric community detection model in (13) minimizes the distance between the community nodes and two central nodes. The proposed verification index checks that the structure of the proposed clustering model optimizes the distance between the non-central and two central nodes. Therefore, the model equations are verified, if these distances become minimum. The minimum distance from n_k to cluster i from (14) is equal to following equation:

$$D_{ik}^{\min} = \Delta_i \left(\frac{1 - \bar{u}_{ik}}{\bar{u}_{ik}} \right)^{m-1} \tag{18}$$

So, if n_k is close to either $n_{i'}$ or $n_{i''}$ the objective function minimizes the distance between the non-central nodes and central nodes. Therefore, the verification index is defined as following:

$$I_{GFT2}^{Verification} = \frac{1}{nc} \sum_{k=1}^{g} \sum_{i=1}^{c} \left(\Delta_i \left(\frac{1 - \bar{u}_{ik}}{\bar{u}_{ik}} \right)^{m-1} \right) \tag{19}$$

where $I_{GFT2}^{Verification}$ measures the average closer distance between non-central nodes and central nodes based on upper type-2 fuzzy membership values and density of communities. Similarly, for type-1 fuzzy clustering (19) becomes as following:

$$I_{GFT1}^{Verification} = \frac{1}{nc} \sum_{k=1}^{n} \sum_{i=1}^{c} \left(\Delta_i \left(\frac{1 - u_{ik}}{u_{ik}} \right)^{m-1} \right) \tag{20}$$

4. Experimental results

In this section, the performance of the proposed fuzzy duocentric community detection model isshown by several examples.

Calculate degree of each node, $w_j = \sum_{j'=1}^{g} a_{jj'}/g$, and distance matrix $[D_{ik}] = \sum_{j=1}^{g} w_j \mid a_{ij} - a_{kj} \mid$

Determine members of set g^D using (8) where $P(n_{i'} \cap n_{i''}) > \phi^{min}$

Select c pairs of node as initial duocenters from g^D

Estimate Δ_i using (16) Repeat:

- Calculate $D^{min(t)}$ and $D^{max(t)}$ using (12)
- Calculate belonging degree of each node to each community, $\tilde{U}^{(t)}$, using (14)
- Redetermine the duocenters of each community, $N^{c(t)}$, using (15)

Until $\left\| \tilde{U}^{(t)} - \tilde{U}^{(t-1)} \right\| < \varepsilon$, $t > 1$

Fig. 6. The proposed fuzzy duocentric community detection algorithm.

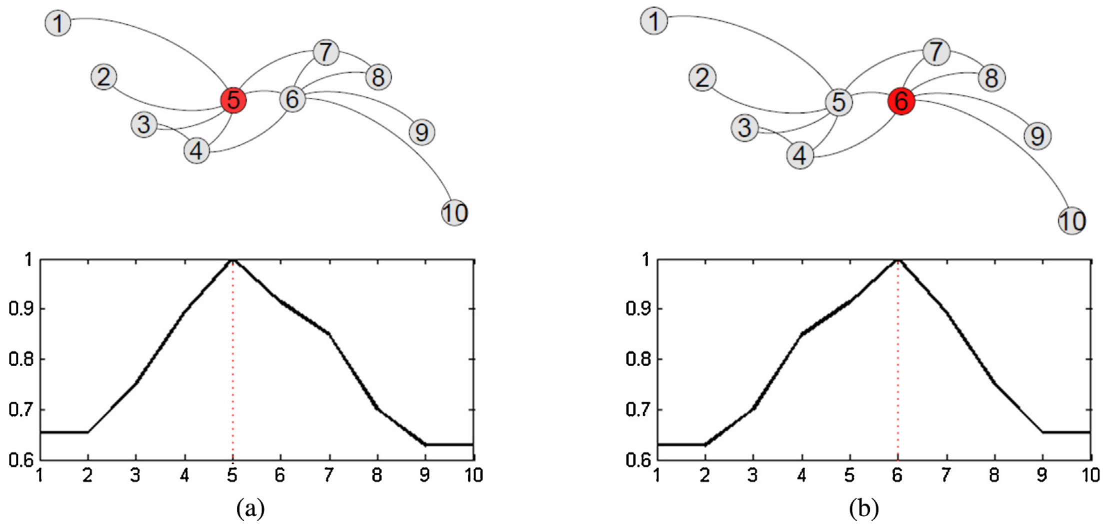


Fig. 7. Example 1: type-1 fuzzy membership values of a egocentric community when its center is (a) n_5 node (case 1) and (b) n_6 node (case 2).

Additionally, the results of the proposed algorithm are compared with the fuzzy egocenteric community detection algorithm such as PCM. The results were illustrated by using Gephi software (Bastian et al., 2009).

Example 1 is a visible and understandable simple data set which indicates the difference between type-1 and type-2 fuzzy membership values for egocentric and duocentric community. Example 2 consists of two generated communities to indicate how node's degree of belonging to the community is depended on its situation and connections in the community. The proposed model is also applied on Example 3, an artificial network, to evaluate the performance of the proposed model and compare it to the PCM model. Examples 4 and 5 are a part of the facebook network and a co-authorship network, respectively.

In all examples, the density of the communities and fuzzifier parameter were considered as $\Delta = 1$ and $m = 1.5$. The maximum possible value of the distance between central nodes and non-central nodes is $d^{\max} = 1 \Rightarrow \underline{u} = 1/1 + (1/1)^{1/(m-1)}$ and the minimum possible distance is $d^{\min} = 0 \Rightarrow \bar{u} = 1/1 + (0/1)^{1/(m-1)}$. Therefore, the type-2 fuzzy membership values are between 0.5 and 1, $\tilde{u}_{..} = [0.5, 1]$.

Example 1. This simple example indicates the difference between type-1 and type-2 fuzzy numbers in defining the degree of belonging to a community. The network consists of 10 nodes as shown in Fig. 7. Nodes n_5 and n_6 could be considered as central node since they had more connections compare to the other nodes. Fig. 7 shows the network, when it is considered as egocentric network.

In egocentric community which a community's center is a single node, the degree of belonging is determined as type-1 fuzzy membership values. The nodes' degree of belonging to the community when n_5 and n_6 are the center of the community are shown in Fig. 7(a) and (b), respectively.

The structure of this community also could be considered as the duocentric community. In this case, both nodes n_5 and n_6 form the center of the community as shown in Fig. 8. The nodes' degree of belonging to this duocentric community are determined based on type-2 fuzzy membership values.

The type-1 and type-2 membership values are shown in Table 1. The degree of belonging to the egocentric community were calculated as type-1 fuzzy membership value using (4). The degree of belonging to the duocentric community are calculated as type-2 fuzzy membership value using (14). Consider n_4 as an example, the type-1 fuzzy membership value of n_4 is $u_4 = 0.8930$ when the center of egocentric community is n_5 and it is $u_4 = 0.8482$ when center is n_6. However, the degree of belonging to the duocentric community is equal to $\tilde{u}_4 = [0.8482, 0.8930]$.

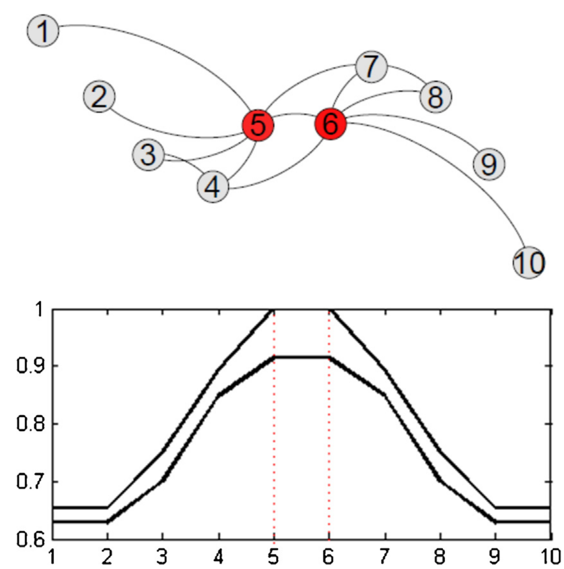


Fig. 8. Example 1: type-2 fuzzy membership values of duocentric community when its center is both n_5 and n_6 nodes.

Table 1
Example1: type-1 and type-2 fuzzy membership values resulting from PCM and the proposed model.

Node	PCM		Proposed model	
	u ; $n^c = n_5$	u ; $n^c = n_6$	$\underline{u}$	$\bar{u}$
n_1	0.6519	0.6283	0.6283	0.6519
n_2	0.6519	0.6283	0.6283	0.6519
n_3	0.7503	0.7005	0.7005	0.7503
n_4	0.8930	0.8482	0.8482	0.8930
n_5	1.0000	0.9135	0.9135	1.0000
n_6	0.9135	1.0000	0.9135	1.0000
n_7	0.8482	0.8930	0.8482	0.8930
n_8	0.7005	0.7503	0.7005	0.7503
n_9	0.6283	0.6519	0.6283	0.6519
n_{10}	0.6283	0.6519	0.6283	0.6519

Example 2. Example 2 contains 21 nodes as shown in Fig. 9. There are two duocentric communities in this sample network. The

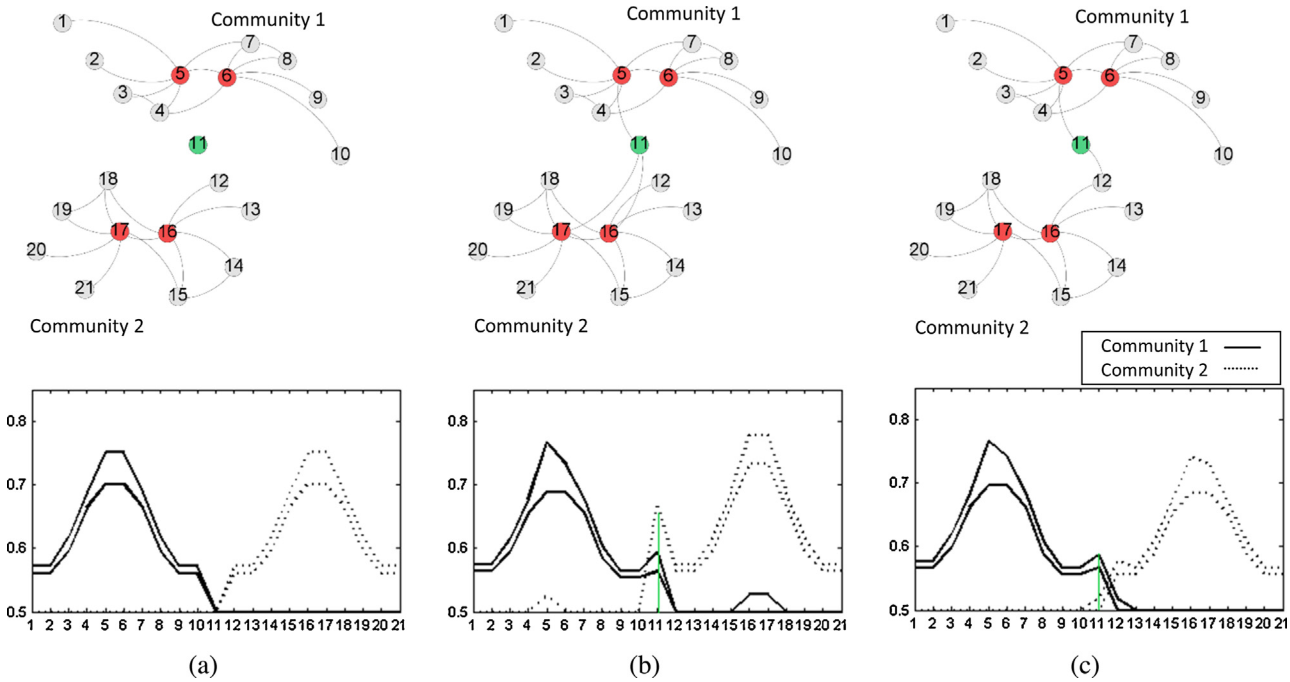


Fig. 9. Example 2: type-2 fuzzy membership values in three different situations of n_{11} (a) case 1, (b) case 2, and (c) case 3.

degrees of belonging to the communities were calculated by using (14). Node n_{11} is considered in three different situations in the network. In first case, n_{11} do not have any connections to the both communities as shown in Fig. 9(a). Since n_{11} do not connect to the neither central nodes of community 1 nor community 2, the type-2 fuzzy membership value of n_{11} to both communities are the same and equal to $\tilde{u}_{1,11} = \tilde{u}_{2,11} = [0.5, 0.5]$.

In second case, n_{11} is connected to the one of the central node of community 1, n_5, and both central nodes of community 2, n_{16} and n_{17} as shown in Fig. 9(b). The degree of belonging n_{11} to community 1 is $\tilde{u}_{1,11} = [0.5640, 0.5935]$ and to community 2 is $\tilde{u}_{2,11} = [0.6668, 0.6668]$. In third case, node n_{11} connects to the central node of community 1, while it connects to the one of the non-central node of community 2. It still connects to community 2 but the connection is not as strong as its connection to community 1. The type-2 fuzzy membership value to community 1 is equal to $\tilde{u}_{1,11} = [0.5664, 0.5867]$. However, its degree of belonging to community 2 is equal to $\tilde{u}_{2,11} = [0.5000, 0.5182]$ as shown in Fig. 9(c). Type-2 membership values are able to determine the degree of belonging to duocenters and alsothe overlapping among communities. Table 2 indicates the generated type-2 fuzzy membership for these three cases by using (14).

In addition, this network was segmented by PCM using (3) and its results were compared to the results of the proposed model based on the proposed verification index. The verification index for the proposed model using (19) is equal to $I_{GFT2}^{Verification} = 0.4134$. However, the value of verification index for PCM using (20) is equal to $I_{GFT1}^{Verification} = 0.4351$. Therefore, the performance of the proposed model is better than PCM since the value of verification for the proposed model is lower than PCM model.

Example 3. This generated network consists of 75 nodes and 124 links as shown in Fig. 10(a). There are three duocentric communities in this network and the performance of PCM and the proposed fuzzy duocentric community detection model were evaluated to detect these three communities.

First, this sample was segmented based on PCM model using (3) and its result is shown in Fig. 10(b). The nodes' type-1 fuzzy membership values were calculated using (4) and illustrated in Fig. 11(a). Nodes were assigned to the community which had the highesttype-1 fuzzy membership valuecompared to other communities. The member of community 1, 2 and 3 are shown respectively by blue, green and red. The center of the communities were determined based on (5) as $n_1^c = n_{54}$, $n_2^c = n_{29}$ and $n_3^c = n_{41}$. The communities' centers are shown by darker colors and doubled line in Fig. 10(b).

Then the network was segmented by the proposed fuzzy duocentric community detection model using (13). The nodes' type-2 fuzzy membership values were calculated using (14) and illustrated in Fig. 11(b). Nodes were assigned to the community which both lower and upper membership values were higher than other communities. The members of community 1, 2 and 3 are shown respectively by blue, green and red. The duocenters of the communities are determined based on (15) as $n_1^c = \{n_1, n_{54}\}$, $n_2^c = \{n_{29}, n_{41}\}$ and $n_3^c = \{n_{61}, n_{66}\}$. The communities' centers are shown by darker colors and doubled line in Fig. 10(c).

The results of segmentation by PCM algorithm and the proposed model were compared based on the proposed verification index. The verification index for the proposed model using (19) is equal to $I_{GFT2}^{Verification} = 0.3113$. However, the value of verification index for PCM using (20) is equal to $I_{GFT1}^{Verification} = 0.3339$. The value of $I_{GFT2}^{Verification}$ is lower than $I_{GFT1}^{Verification}$. Therefore, the performance of the proposed model is better than PCM. The network in Fig. 10(a) would benefit from identifying duocenters as shown in Fig. 10(c) rather than detecting egocentric communities as shown in Fig. 10(b).

Example 4. The forth example is a real ego-facebook network which indicates a friendship network of a person on facebook. This sample consists of 110 nodes and 453 links as shown in Fig. 12(a). The number of connected nodes is 78. The network was segmented based on PCM model using (3) and its result is shown in Fig. 12(b). The nodes' type-1 fuzzy membership values were calculated using (4) and illustrated in Fig. 13(a). Nodes were assigned to the community which had highestmembership value compared to

Table 2
Example 2: type-2 fuzzy membership values resulting from fuzzy duocentric community detection model.

Node	Case 1				Case 2				Case 3			
	$\underline{u}_{1\cdot}$	$\bar{u}_{1\cdot}$	$\underline{u}_{2\cdot}$	$\bar{u}_{2\cdot}$	$\underline{u}_{1\cdot}$	$\bar{u}_{1\cdot}$	$\underline{u}_{2\cdot}$	$\bar{u}_{2\cdot}$	$\underline{u}_{1\cdot}$	$\bar{u}_{1\cdot}$	$\underline{u}_{2\cdot}$	$\bar{u}_{2\cdot}$
n_1	0.5610	0.5718	0.5000	0.5000	0.5640	0.5737	0.5000	0.5000	0.5664	0.5765	0.5000	0.5000
n_2	0.5610	0.5718	0.5000	0.5000	0.5640	0.5737	0.5000	0.5000	0.5664	0.5765	0.5000	0.5000
n_3	0.5939	0.6166	0.5000	0.5000	0.5935	0.6139	0.5000	0.5000	0.5971	0.6183	0.5000	0.5000
n_4	0.6639	0.6882	0.5000	0.5000	0.6560	0.6777	0.5000	0.5000	0.6622	0.6847	0.5000	0.5000
n_5	0.7005	0.7503	0.5000	0.5000	0.6886	0.7666	0.5265	0.5265	0.6961	0.7654	0.5000	0.5000
n_6	0.7005	0.7503	0.5000	0.5000	0.6886	0.7331	0.5000	0.5000	0.6961	0.7422	0.5000	0.5000
n_7	0.6639	0.6882	0.5000	0.5000	0.6560	0.6777	0.5000	0.5000	0.6622	0.6847	0.5000	0.5000
n_8	0.5939	0.6166	0.5000	0.5000	0.5835	0.6036	0.5000	0.5000	0.5867	0.6076	0.5000	0.5000
n_9	0.5610	0.5718	0.5000	0.5000	0.5544	0.5640	0.5000	0.5000	0.5564	0.5664	0.5000	0.5000
n_{10}	0.5610	0.5718	0.5000	0.5000	0.5544	0.5640	0.5000	0.5000	0.5564	0.5664	0.5000	0.5000
n_{11}	0.5000	0.5000	0.5000	0.5000	0.5640	0.5935	0.6668	0.6668	0.5664	0.5867	0.5000	0.5182
n_{12}	0.5000	0.5000	0.5610	0.5718	0.5000	0.5000	0.5640	0.5737	0.5000	0.5182	0.5564	0.5765
n_{13}	0.5000	0.5000	0.5610	0.5718	0.5000	0.5000	0.5640	0.5737	0.5000	0.5000	0.5564	0.5664
n_{14}	0.5000	0.5000	0.5939	0.6166	0.5000	0.5000	0.5935	0.6139	0.5000	0.5000	0.5867	0.6076
n_{15}	0.5000	0.5000	0.6639	0.6882	0.5000	0.5000	0.6668	0.6886	0.5000	0.5000	0.6510	0.6734
n_{16}	0.5000	0.5000	0.7005	0.7503	0.5000	0.5265	0.7331	0.7778	0.5000	0.5000	0.6847	0.7422
n_{17}	0.5000	0.5000	0.7005	0.7503	0.5000	0.5265	0.7331	0.7778	0.5000	0.5000	0.6847	0.7307
n_{18}	0.5000	0.5000	0.6639	0.6882	0.5000	0.5000	0.6668	0.6886	0.5000	0.5000	0.6510	0.6734
n_{19}	0.5000	0.5000	0.5939	0.6166	0.5000	0.5000	0.5935	0.6139	0.5000	0.5000	0.5867	0.6076
n_{20}	0.5000	0.5000	0.5610	0.5718	0.5000	0.5000	0.5640	0.5737	0.5000	0.5000	0.5564	0.5664
n_{21}	0.5000	0.5000	0.5610	0.5718	0.5000	0.5000	0.5640	0.5737	0.5000	0.5000	0.5564	0.5664

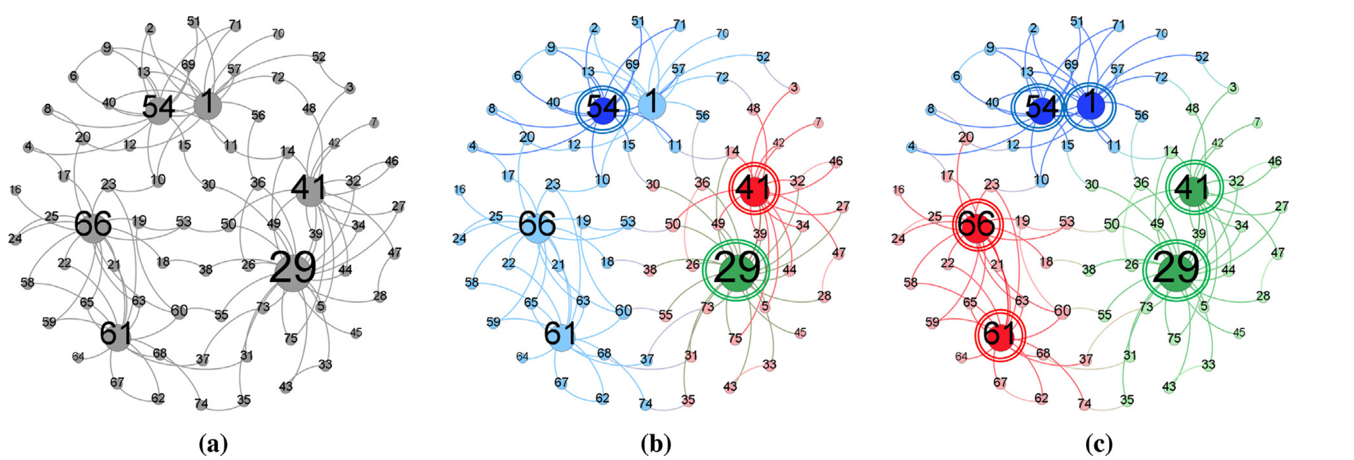


Fig. 10. Example 3: (a) an artificial network consists of three duocentric communities, (b) communities were detected by the PCM, and (c) communities were detected by the proposed model. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

other communities. The member of community 1 and 2 are shown respectively in blue and red. The center of the communities are determined based on (5) as $n_1^c = n_{40}$, $n_2^c = n_{12}$. The communities' centers are shown by darker color and doubled line in Fig. 12(b).

Then the network is segmented by the proposed fuzzy duocentric community detection model using (13). The member of community 1 and 2 are shown respectively in blue and red. The duocenters of the communities were determined based on (15)

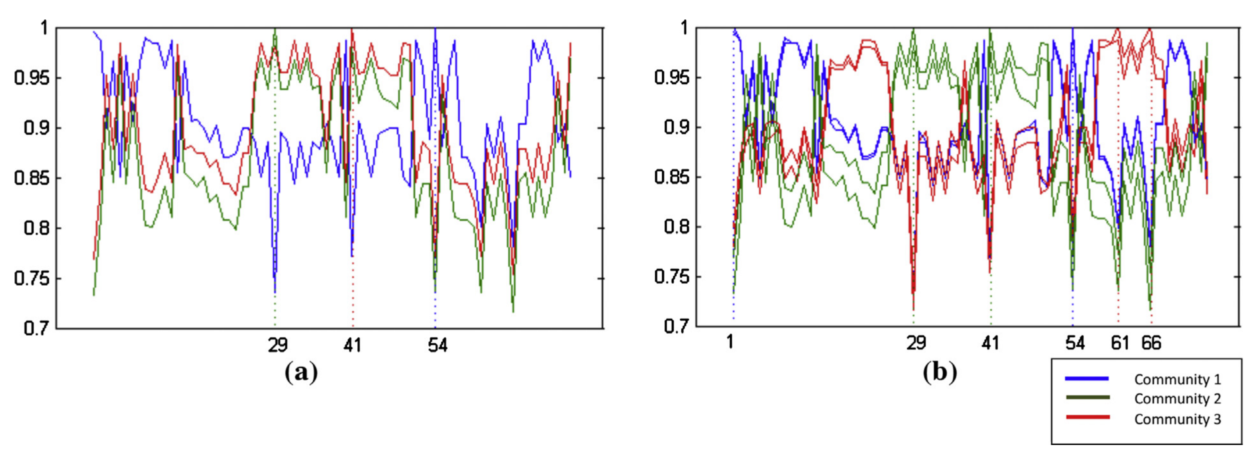


Fig. 11. Example 3: (a) type-1 fuzzy membership value of the communities detected by the PCM and (b) type-2 fuzzy membership values of the communities detected by the proposed fuzzy duocentric community detection model. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

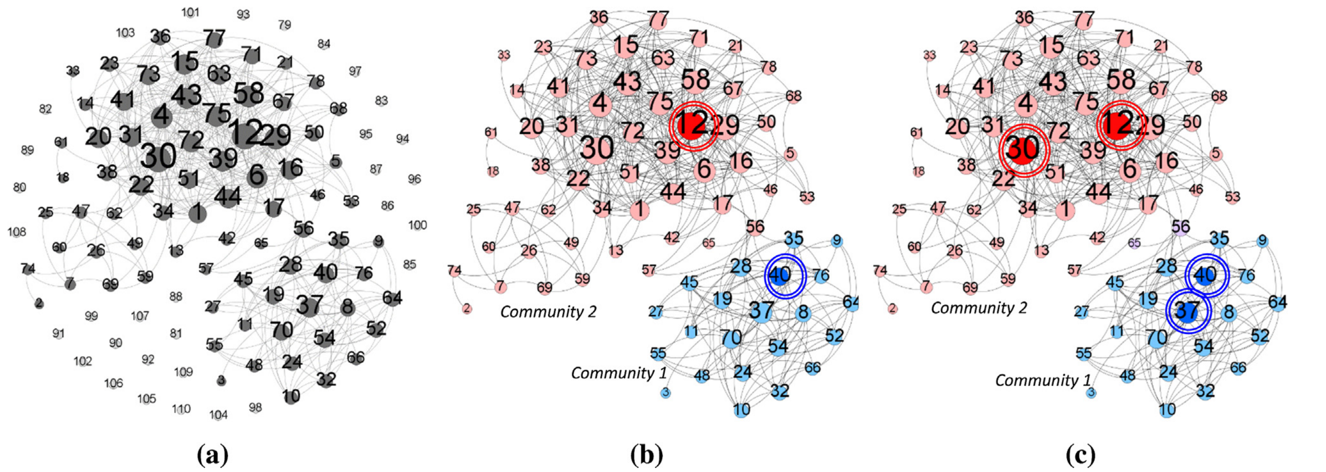


Fig. 12. Example 4: (a) an age-facebook network, (b) communities were detected by the PCM, and (c) communities were detected by the proposed fuzzy duocentric community detection model. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

as $n_1^c = \{n_{40}, n_{37}\}$, $n_2^c = \{n_{12}, n_{30}\}$. The communities' duocenters are shown by darker colors and doubled line in Fig. 12(c). The nodes' type-2 fuzzy membership values were calculated using (14) and illustrated in Fig. 13(c). Nodes are usually assigned to thecommunity which hasthe maximum membership values. Here, membership values are interval numbers and in some cases both lower and upper membership values are not higher than others. For example, the membership values of node n_{56} are $\tilde{u}_{1,56} = [0.5503, 0.5503]$ and $\tilde{u}_{2,56} = [0.5313, 0.5613]$, therefore, $\underline{u}_{1,56} = \underline{u}_{2,56}$ and $\bar{u}_{1,56} < \bar{u}_{2,56}$. Similarly, the membership values of node n_{65} were $\tilde{u}_{1,65} = [0.5061, 0.5061]$ and $\tilde{u}_{2,65} = [0.5000, 0.5061]$. These situations show how the type-2 fuzzy membership values can describe the nodes' sharing among the communities and detect the overlapping communities in uncertain situation which existing approaches are more limited in this respect.

The results of segmentation by PCM algorithm and the proposed model were compared based on the proposed verification index. The verification index for the proposed model using (19) is equal to $I_{GFT2}^{Verification} = 0.4194$. However, the value of verification index for PCM using (20) is equal to $I_{GFT1}^{Verification} = 0.4312$. Therefore, the performance of the proposed model is better than PCM since the value of verification for the proposed model is lower than PCM model.

Example 5. The last example is a real co-authorship network of scientists who were working on network theory and experiments, as compiled by Newman (2006). Scientific collaboration is a complex phenomenon that improve the sharing of competence and production of new scientific knowledge. Social network analysis is often used to describe the collaboration patterns defined by co-authorship relationship. This network includes 1589 nodes and 2792 links, as shown in Fig. 14(a). To detect communities, first we consider the largest connected component of this network. This connected network consists of 379 nodes and 914 links, as shown in black color in Fig. 14(a).

The network was segmented based on the PCM model using (3) into 10 communities. As shown in Fig. 14(b), the center of the communities were determined based on (5) as $n^c = \{n_{65}, n_4, n_5, n_{16}, n_{15}, n_{25}, n_{70}, n_{45}, n_{51}, n_2\}$. The nodes' type-1 fuzzy membership values were calculated using (4) and illustrated in Fig. 15.

The network was also segmented by the proposed fuzzy duocentric community detection model using (13). The duocenters of the communities using (15) are $n^c = \{\{n_{22}, n_{24}\}, \{n_4, n_{16}\}, \{n_5, n_{15}\}, \{n_{70}, n_{303}\}, \{n_{231}, n_{236}\}, \{n_{52}, n_{170}\}, \{n_{118}, n_{214}\}, \{n_{112}, n_{131}\}, \{n_{85}, n_{86}\}, \{n_{32}, n_{33}\}\}$ as shown in Fig. 14(c). The nodes' type-2 fuzzy membership values were determined by using (14) and illustrated in Fig. 15.

The verification index values, $I_{GFT2}^{Verification}$ and $I_{GFT1}^{Verification}$, were calculated when number of communities being between 2 and 10 as shown in Fig. 16. The maximum number of communities was considered as 10 communities since the number of members of set g^D is equal to 10. when $\Phi^{\min} = 5$ and the number of duocentric communities could not be more than the number of existing duocentrsin the network. In addition, the optimum number of communities in this example is 10 communities. Since, the values of

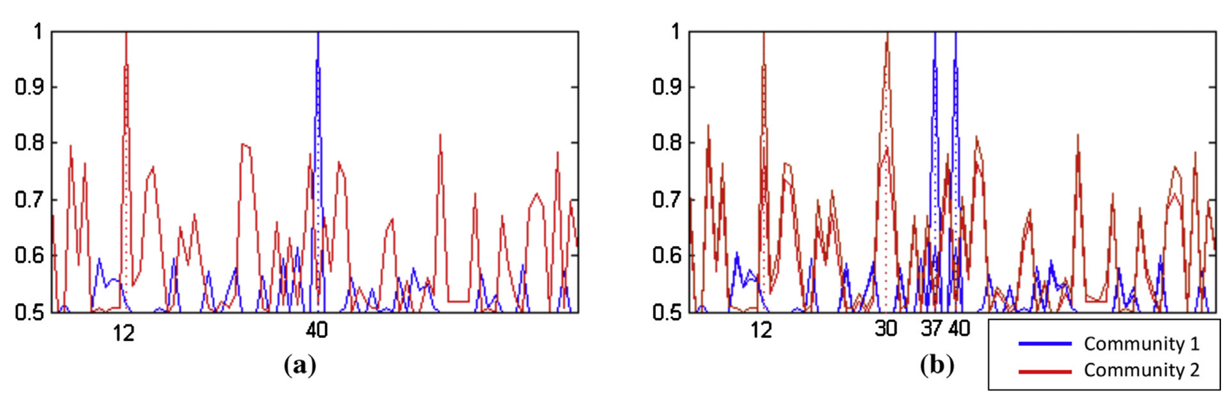


Fig. 13. Example 4: membership values or community (a) communities were detected by the PCM and (b) communities were detected by proposed fuzzy duocentric community detection model. (For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

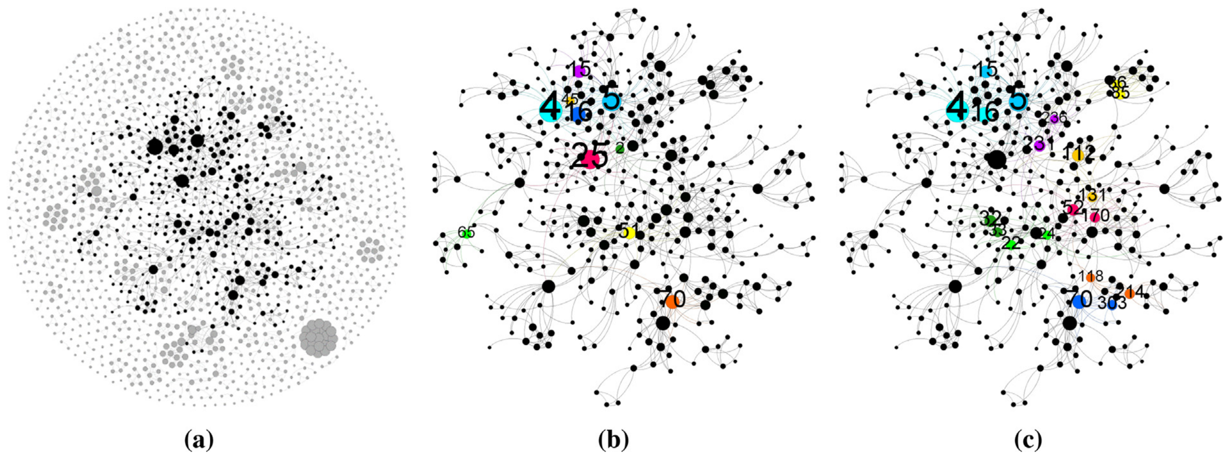


Fig. 14. Example 5: (a) a real co-authorship network of scientists working on network theory and experiment, (b) centers were detected by the PCM, and (c) duocenters were detected by the proposed model.

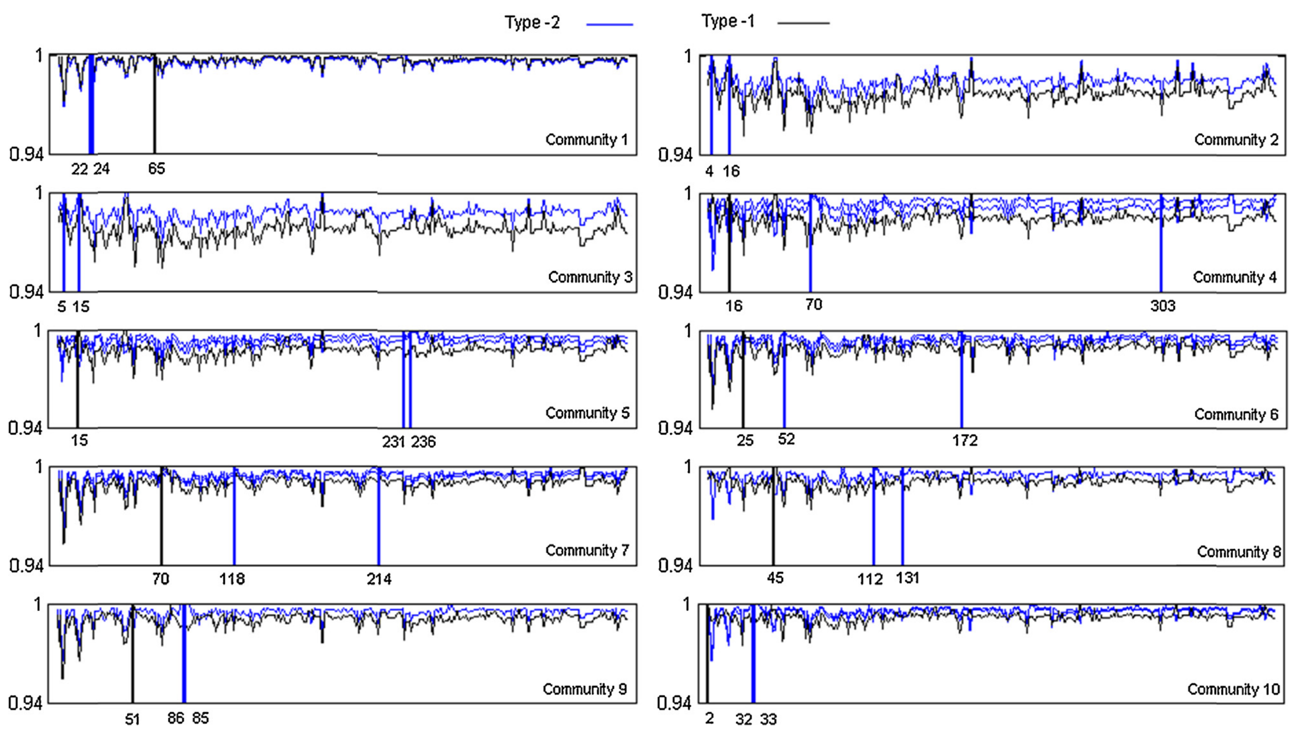


Fig. 15. Example 5: the type-1 and type-2 fuzzy membership values.

verification index between 2 and 10 is minimum at this point. When the number of communities is equal to 10, the verification index for the proposed model using (19) is equal to $I_{GFT2}^{Verification} = 0.0925$. However, the value of verification index for PCM using (20) is equal to $I_{GFT1}^{Verification} = 0.1079$. Therefore, the performance of the proposed model is better than PCM since the value of verification for the proposed model is lower than PCM model. As shown in Fig. 16, the trend of verification index for the proposed model is generally lower than the PCM model which shows its better performance.

The verification index values for all examples are shown in Table 3. As results show, the values of $I_{GFT2}^{Verification}$ are lower than $I_{GFT1}^{Verification}$ which indicates the better perfomance of segmentation by the proposed model rather than PCM model.

Table 3
Examples 1–5: verification index values for the results of the PCM and the proposed fuzzy duocentric community detection model.

Example	No. of nodes	No. of edges	No. of connected nodes	No. of connected edges	$\Phi^{\min}$	No. of communities	$I_{GFT1}^{Verification}$	$I_{GFT2}^{Verification}$
Example 1	10	13	10	13	4	1	0.4308	0.3538
Example 2	21	27	20	27	4	2	0.4351	0.4134
Example 3	75	124	75	124	5	3	0.3339	0.3113
Example 4	86	453	78	453	10	2	0.4312	0.4194
Example 5	1589	2792	379	914	5	10	0.1079	0.0925

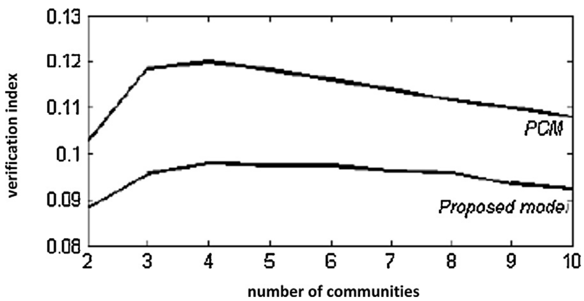


Fig. 16. Example 5: verification index values for the results of the PCM and the proposed fuzzy duocentric community detection model.

5. Conclusion

In this paper, a graph clustering model was defined for detecting overlapping duocentric communities in the complex networks. Our proposed model was developed based on an extension of a center-based objective function clustering model to dual center clusters. In the case that there are two central actors in a community, it is difficult if not impossible to detect the community around only one node. Our proposed duocentric community detection model is able to identify the communities with two active actors and also determine the degree of belonging to such communities.

The nodes' membership values to the communities were defined as type-2 fuzzy numbers, which indicate the degree of belonging to both central nodes as upper and lower membership values. Moreover, in this case in which communities do not have sharp boundaries, interval type-2 fuzzy membership values are able to describe how nodes are shared between the communities and formed overlapping communities.

The compatible verification index was proposed to evaluate the performance of the proposed model and PCM. Results indicated that the proposed fuzzy duocentric community detection model was able to detect duocentric communities especially when they were not well-separated from each other. The interval fuzzy type-2 membership values were determined the strength of belonging to both central nodes and distinguish between nodes which were close to both central nodes, close to only one of them, or not close to non of them. Our proposed model is able to describe this kind of degree of belonging that previous approaches are unable to.

Appendix A.

Proof of Theorem. All elements $\tilde{u}_{ik}$ of $\tilde{U}$, $\forall i, k$, are independent. Therefore, minimizing $J(\tilde{U}, N^c, \Delta)$ with respect to $\tilde{U}$ is equivalent to minimizing $J(\tilde{u}_{ik}, n_i^{c1,2}, \Delta_i)$ with respect to $\tilde{u}_{ik}$. To find the first-order necessary conditions for optimality, the gradients of $J(\tilde{u}_{ik}, n_i^{c1,2}, \Delta_i)$ with respect to $\tilde{u}_{ik}$, $\underline{u}_{ik}$ and $\bar{u}_{ik}$, are set to zero:

$$\begin{aligned}\frac{\partial J}{\partial \underline{u}_{ik}} &= m\underline{u}_{ik}^{m-1} - m\Delta_i(1-\underline{u}_{ik})^{m-1} = 0 \Rightarrow \underline{u}_{ik} = \frac{1}{1+(D_{ik}^{\max}/\Delta_i)^{1/(m-1)}} \\ \frac{\partial J}{\partial \bar{u}_{ik}} &= m\bar{u}_{ik}^{m-1} - m\Delta_i(1-\bar{u}_{ik})^{m-1} = 0 \Rightarrow \bar{u}_{ik} = \frac{1}{1+(D_{ik}^{\min}/\Delta_i)^{1/(m-1)}}\end{aligned} \tag{A.1}$$

To find the optimal community center node, if $D_{ik}^1 > D_{ik}^2$ we have:

$$\begin{aligned}\frac{\partial J}{\partial n_{i'}} = 0 \Rightarrow \frac{\partial}{\partial n_{i'}}\left[\sum_{k=1}^{g}\underline{u}_{ik}^m\sum_{j=1}^{g}w_j\left|a_{i'j}-a_{kj}\right| = \sum_{k=1}^{g}\underline{u}_{ik}^m\sum_{j=1}^{g}w_j a_{i'j} - \sum_{k=1}^{g}\underline{u}_{ik}^m\sum_{j=1}^{g}w_j a_{kj}|\right] = 0 \\ \frac{\partial J}{\partial n_{i''}} = 0 \Rightarrow \frac{\partial}{\partial n_{i''}}\left[\sum_{k=1}^{g}\bar{u}_{ik}^m\sum_{j=1}^{g}w_j\left|a_{i''j}-a_{kj}\right| = \sum_{k=1}^{g}\bar{u}_{ik}^m\sum_{j=1}^{g}w_j a_{i''j} - \sum_{k=1}^{g}\bar{u}_{ik}^m\sum_{j=1}^{g}w_j a_{kj}|\right] = 0\end{aligned} \tag{A.2}$$

(A.2) is solved together if $n_{i'}$ and $n_{i''}$ are the duocenters of community i and has the minimum distance to other community members with respect to their membership values, $\tilde{u}_{ik}$. Thus, we have:

$$\begin{aligned}\sum_{j=1}^{g}w_j a_{i'j} = \left(\sum_{k=1}^{g}\underline{u}_{ik}^m\sum_{j=1}^{g}w_j a_{kj}/\sum_{k=1}^{g}\underline{u}_{ik}^m\right) \Rightarrow \min\left\{\sum_{j=1}^{g}w_j a_{i'j} - \left(\sum_{k=1}^{g}\underline{u}_{ik}^m\sum_{j=1}^{g}w_j a_{kj}/\sum_{k=1}^{g}\underline{u}_{ik}^m\right)\right\} \\ \sum_{j=1}^{g}w_j a_{i''j} = \left(\sum_{k=1}^{g}\bar{u}_{ik}^m\sum_{j=1}^{g}w_j a_{kj}/\sum_{k=1}^{g}\bar{u}_{ik}^m\right) \Rightarrow \min\left\{\sum_{j=1}^{g}w_j a_{i''j} = \left(\sum_{k=1}^{g}\bar{u}_{ik}^m\sum_{j=1}^{g}w_j a_{kj}/\sum_{k=1}^{g}\bar{u}_{ik}^m\right)\right\}\end{aligned} \tag{A.3}$$

from (A.3), the duocenters of community i are defined as:

$$\begin{aligned}n_{i'}, n_{i''} &= \arg_{i',i''}\min\left\{\sum_{j=1}^{g}w_j(a_{i'j}+a_{i''j}) - \sum_{k=1}^{g}\underline{u}_{ik}^m\sum_{j=1}^{g}w_j a_{kj}/\sum_{k=1}^{g}\underline{u}_{ik}^m - \sum_{k=1}^{g}\bar{u}_{ik}^m\sum_{j=1}^{g}w_j a_{kj}/\sum_{k=1}^{g}\bar{u}_{ik}^m\right\} \\ &P(n_{i'} \cap n_{i''}) > \Phi^{\min}\end{aligned} \tag{A.4}$$

If $D^2 \geq D^1$, the result is the same.

Appendix B.

The necessary condition of the convergence of the proposed algorithm in Fig. 6 is met when:

$$\lim_{t\to\infty}\left\|\tilde{U}^{(t)} - \tilde{U}^{(t-1)}\right\| = 0 \tag{B.1}$$

The iterative formula for $\tilde{u}_{ik}$ follows from the classical gradient descent method, with J_m as the error function to be minimized, namely:

$$\tilde{u}_{ik}^{(t)} = \tilde{u}_{ik}^{(t-1)} - \omega^{(t)}\frac{\partial J_m(\tilde{u}_{ik}, n_i^{c1,2}, \Delta_i)^{(t-1)}}{\partial \tilde{u}_{ik}} \tag{B.2}$$

where, $\omega^{(t)}$ is a positive learning rate parameter and $\frac{\partial J_m(\tilde{u}_{ik}, n_i^{c1,2}, \Delta_i)}{\partial \tilde{u}_{ik}}^{(t-1)}$ is the gradient of J_m with respect to $\tilde{u}_{ik}$ at $(t-1)$ iteration. Rewriting (B.2) for $\tilde{U}$ gives:

$$\tilde{U}^{(t)} - \tilde{U}^{(t-1)} = -\omega^{(t)} \frac{\partial J_m(\tilde{U}, N^c, \Delta)}{\partial \tilde{U}}^{(t-1)} \tag{B.3}$$

Now putting (B.3) in (B.1):

$$\lim_{t\to\infty} \left\| \tilde{U}^{(t)} - \tilde{U}^{(t-1)} \right\| = \lim_{t\to\infty} \left(\left\| \varpi^{(t)} \right\| \left\| \frac{\partial J_m(\tilde{U}, N^c, \Delta)}{\partial \tilde{U}}^{(t-1)} \right\| \right) \tag{B.4}$$

By considering $\omega^{(t)} = \frac{\varpi^{(t)}}{\left\| \frac{\partial J_m(\tilde{U}, N^c, \Delta)}{\partial \tilde{U}}^{(t-1)} \right\|}$, (B.4) becomes:

$$\lim_{t\to\infty} \left\| \tilde{U}^{(t)} - \tilde{U}^{(t-1)} \right\| = \lim_{t\to\infty} \left\| \varpi^{(t)} \right\| = 0 \tag{B.5}$$

where, $\varpi^{(t)} = \varpi_0^{(t)}/t$ in which $\varpi_0^{(t)}$ is a constant value and since then $\varpi^{(t)} \to 0$ when $t \to \infty$. Therefore, (B.5) is proofed and, consequently, the proposed algorithm is convergent.

References

Achlioptas, D., Clauset, A., Kempe, D., Moore, C., 2005. On the bias of traceroute sampling (or: why almost every network looks like it has a power law). In: Gabow, H.N., Fagin, R. (Eds.), Proceedings of the Thirty Seventh Annual ACM Symposium on Theory of Computing, STOC. ACM Press, New York, NY, USA.

Almquist, Z.W., 2012. Random errors in egocentric networks. Soc. Netw. 34, 493–505.

Bastian, M., Heymann, S., Jacomy, M., 2009. Gephi: an open source software for exploring and manipulating networks. In: International AAAI Conference on Weblogs and Social Media (2009), pp. 361–362.

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function. Plenum Press, New York.

Burt, R.S., 1992. Structural Holes: The Social Structure of Competition. Harvard University Press, Cambridge, MA.

Coleman, J.S., 1990. Foundations of Social Theory. Harvard University Press, Cambridge, MA.

Coromina, L., Guia, J., Coenders, G., Ferligoj, A., 2008. Duocentered networks. Soc. Netw. 30, 49–59.

Coyle, G., Exelby, D., 2000. The validation of commercial system dynamics models. Syst. Dyn. Rev. 16 (1), 27–41.

Dave, R.N., Krishnapuram, R., 1997. Robust clustering methods: a unified view. IEEE Trans. Fuzzy Syst. 5 (2), 270–293.

Dunn, J.C., 1974. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. J. Cybern. 3, 32–57.

Estrad, E., Hatanoe, N., Benzi, M., 2012. The physics of communicability in complex networks. Phys. Rep. 514, 89–119.

Everett, M., Borgatti, S.P., 2005. Ego network betweenness. Soc. Netw. 27, 31–38.

Fortunato, S., 2010. Community detection in graphs. Phys. Rep. 486, 75–174.

Freeman, L.C., 1982. Centered graphs and the structure of ego networks. Math. Soc. Sci. 3, 291–304.

Fu, X., Li, L., Wang, C., 2013. Detection of community overlap according to belief propagation and conflict. Physica A 392, 941–952.

Gath, I., Geva, A.B., 1989. Unsupervised optimal fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell. 11 (7), 773–780.

Golsefid, S.M.M., Fazel Zarandi, M.H., Bastani, S., 2015. Center-Based Overlapping Community Detection Model in Social Networks. Int. J. Intell. Syst. (in press).

Golsefid, S.M.M., Fazel Zarandi, M.H., 2015. Dual-centers type-2 fuzzy clustering framework and its verification and validation indices. Appl. Soft Comput. (in press).

Granovetter, M., 1973. The strength of weak ties. Am. J. Sociol. 78, 1360–1380.

Jiang, J.Q., Dress, A.W.M., Yanga, G., 2009. A spectral clustering-based framework for detecting community structures in complex networks. Appl. Math. Lett. 22, 1479–1482.

Kadushin, C., 2004. Introduction to Social Network, Theory Basic Network Concepts.

Knoke, D., Kuklinski, J.H., 1982. Network Analysis. Sage, Newbury Park, CA.

Krishnapuram, R., Keller, J.M., 1993. A possibilistic approach to clustering. Trans. Fuzzy Syst. 1 (2), 98–110.

Leskovec, J., Lang, K.J., Dasgupta, A., Mahoney, M.W., 2008. Community structure in large networks: natural cluster sizes and the absence of large well-defined clusters. Internet Math. 6, 29–123.

Li, C., 2012. Study on overlapping community structure. Procedia Eng. 29, 4244–4248.

Liu, J., 2010. Detecting the fuzzy clusters of complex networks. Pattern Recognit. 43, 1334–1345.

Malliaros, F.D., Vazirgiannis, M., 2013. Clustering and community detection in directed networks: a survey. Phys. Rep., http://dx.doi.org/10.1016/j.phystrep2013.08.002

Martino, F., Spoto, A., 2006. Social network analysis: a brief theoretical review and further perspectives in the study of information technology. Psychnol. J. 4 (1), 53–86.

Mendel, J., 2007. Advances in type-2 fuzzy sets and systems. Inf. Sci. 177 (1), 84–110.

Mendel, J., John, B., 2002. Type-2 fuzzy sets made simple? IEEE Trans. Fuzzy Syst. 10 (2), 117–127.

Newman, M.E.J., 2006. Finding community structure in networks using theeigenvectors of matrices. Phys. Rev. E 74, 036104.

Newman, M.E.J., 2003. Ego-centered networks and the ripple effect. Soc. Netw. 25, 83–95.

Passarella, A., Dunbar, R.I.M., Conti, M., Pezzoni, F., 2012. Ego network models for Future Internet social networking environments. Comput. Commun. 35, 2201–2217.

Santo, F., 2010. Community detection in graphs. Phys. Rep. 486, 75–174.

Schaeffer, S.E., 2007. Graph clustering. Comput. Sci. Rev. 1, 27–64.

Scott, J., 2000. Social Network Analysis: A Handbook. Sage, Newbury Park, CA.

Sun, P.G., Gao, L., Han, S.S., 2011. Identification of overlapping and non-overlapping community structure by fuzzy clustering in complex networks. Inf. Sci. 181, 1060–1071.

Vehovar, V., Manfreda, K.L., Koren, G., Hlebec, V., 2008. Measuring ego-centered social networks on the web:Questionnaire design issues. Soc. Netw. 30, 213–222.

Wanga, W., Liu, D., Liu, X., Pana, L., 2013. Fuzzy overlapping community detection based on local random walk and multidimensional scaling. Physica A, http://dx.doi.org/10.1016/j.physa.2013.08.028

Wanga, X., Jiao, L., Wu, J., 2009. Adjusting from disjoint to overlapping community detection of complex networks. Physica A 388, 5045–5056.

Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge Unversity Press.

Wellman, B., Berkowitz, S.D., 1988. Social Structures: A Network Approach. Social Science.

Xie, J., Kelley, S., Szymanski, B.K., 2011. Overlapping Community Detection in Networks: The State of the Art and Comparative Study. arXiv:1110.5813v1113.

Yan, J.T., Hsiao, P.Y.,1995. A new fuzzy clustering based approach for twoway circuit partitioning. In: Proceedings of the Eight International Conference on VLSI Design. IEEE, New York, NY, USA.

Zhang, S., Wang, R.S., Zhang, X.S., 2007. Identification of overlapping community structure in complex networks using fuzzy c-means clustering. Physica A 374, 483–490.