Expert Systems with Applications 42 (2015) 6913–6934

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Investigating community structure in perspective of ego network

Anupam Biswas *, Bhaskar Biswas

Department of Computer Science and Engineering, Indian Institute of Technology (BHU), Varanasi, India

ARTICLE INFO

Article history:
Available online 14 May 2015

Keywords:
Data mining
Community detection
Social network
Ego network
Graph clustering

ABSTRACT

Complex relationships within the data are modeled as information network in various application areas of data mining. Identification of connected groups of nodes associated with similar information is one of the common interests behind such modeling. These connected groups are referred as community structure. This work investigates community structure by incorporating a sociological property known as ego network. Ego network facilitates personalized view of relationships. We introduce the notion of *mutual interest* in the relationship by utilizing such personalized view and re-define community structure. Different from classical way of defining communities through dense connectivity, proposed definition incorporates two properties: *Reachability* and *Isolability*. Reachability measures the ability of any node to reach out members of community, while Isolability accounts the ability of any community to isolate itself from rest of the network. Exploiting new definition of community structure, we propose an algorithm for identifying communities. Experimental results on a variety of real world data and synthetic data show communities identified with proposed algorithm is highly inclined towards accuracy in comparison to other state-of-the-art approaches.

1. Introduction

Modern data mining is often faced with problems turning out from complex relationships present within the data. Graphical interpretation of data as information network or relationship network has gained attention in several domains. Application areas include text clustering (Wei, Lu, Chang, Zhou, & Bao, 2015), social network analysis (Fan & Yeung, 2015; Moradi, Olovsson, & Tsigas, 2014; Nikolaev, Razib, & Kucheriya, 2015), image processing (Boykov & Funka-Lea, 2006; Grady, 2006; Li & Wu, 2015), video processing (Yeung, Yeo, & Liu, 1998), biological network (Wang, Li, Chen, & Pan, 2011) and neural network (Liu & juan Ban, 2015), world wide web (Jeevan, Prashanth, Punith, & Hegde, 2011; Mukhopadhyay & Sing, 2004), supply chain network (Li & Daie, 2014; Xue, Wei, & Zeng, 2011) and many others. Growing size and complexity of relationship network in various application domain seek more effective, and accurate analysis to study such complex data system. One predominant aspect of such analysis is community detection, which is known to present in most of the real world network.

From the expert systems view point, community detection problem can be treated as unsupervised learning of connectivity patterns in the network data. Discovery of such connectivity patterns has a significant role in understanding the relationships among various entities of the network and their functions in a complex network. Communities are the groups of entities of the network those are similar and connected with each other. Similarity among entities of network is defined in terms of connectivity. Generally, dense connectivity represents high similarity. Thus, community detection problem is nothing but a clustering problem (Liu & juan Ban, 2015) in network. Often, community detection problem is referred as graph clustering. In graphical representation, densely connected regions of the network are considered as probable communities as shown in the Fig. 1.

Certainly such informal visual observations can be created only on networks of smaller size to identify communities. Nowadays, data become very huge and to deal with such big network data needs formal automated system. Numerous algorithms have been developed to reach this goal. Though many shows good quality communities, most of these algorithms detect inaccurate and biased communities (Almeida, Guedes, Meira, & Zaki, 2011; Ebrahimi & Abadeh, 2012; Moradi, Olovsson, & Tsigas, 2012; Orman, Labatut, & Cherifi, 2012). Communities are defined in multiple ways specific to any particular domain, but there is no generalized definition for communities in the network other than the

* Corresponding author.
E-mail addresses: abanumail@gmail.com (A. Biswas), bhaskar@iitbhu.ac.in (B. Biswas).

http://dx.doi.org/10.1016/j.eswa.2015.05.009
0957-4174/

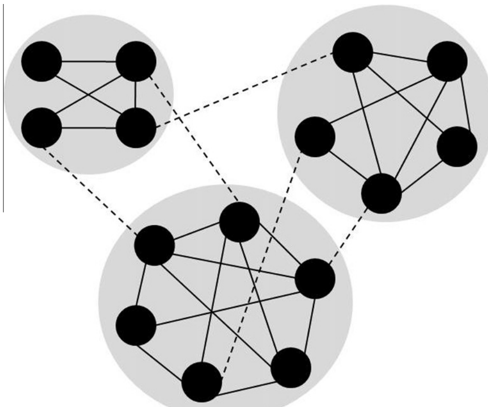

Fig. 1. In many network data the nodes form communities (shaded) within which the connections are more dense (solid lines) than among communities (dashed lines).

dense connectivity (Khorasgani, Chen, & Zaïane, 2010). Most of the community detection algorithms utilize this property to identify communities in the network.Communities detected with such a precept which holds no clear correlation with entities of the real world network. There has a consensus process for forming communities in real world network, they are not formed instantaneously. The shaping of communities involves interaction of entities and togetherness among entities. Hence, involvement of entities also have to be considered for defining communities in real sense.

- In this paper, we investigate entity or node level properties of real world network from the perspective of the social ego network (Arnaboldi, Conti, Passarella, & Pezzoni, 2012; Freeman, 1982; McAuley & Leskovec, 2014; Passarella, Dunbar, Conti, & Pezzoni, 2012; Wang & Gao, 2013). With thorough study of communities and nodes involved in the communities we have redefined community structure (Section 2) considering node level property along with community level property.
- We have proposed an Ego Network Based Community detection (ENBC) algorithm (Section 4) for exploring communities in network data by harnessing both node and community level properties. Time complexity of the ENBC algorithm is $O(n^2)$ (Section 10).
- We have validated our algorithm with several popularly used evaluation metrics (Section 5) and performed comparative analysis with six state-of-the-art community detection algorithms. Experimentation is carried out with seven real world network data sets as well as two synthetic network data sets (Section 6). We have analyzed results (Section 9) in three aspects (Section 7) with different experimental setups (Section 8).
- Experimental results indicate that our algorithm is more adept at finding the true communities than other competitor algorithms (Section 9). Moreover, our algorithm does not require prior knowledge about number of communities and has flexibility to control quality and accuracy of communities (Section 9.4). The conclusion drawn from the results (Section 11) that our algorithm is highly inclined towards the accuracy of communities.

2. Community structure

In this section we investigate community structures in terms of sociological facts and phenomena which lead to form communities in social networks. Humans are society loving creature, they cannot live alone too much time. They eagerly interact with their neighbors and collectively engaged with the society forming some kind of social bonds such as relationships or ties. Such mutual relationships among different persons connect each other, forming network of social relationships. Constructive interaction yields stronger relationship, which apparently shape different groups or communities within the network. An interaction can only be constructive, when both interacting persons have certain mutual consent. Hence, indirectly personal interest matters a lot in forming communities. Ego network is the well known social phenomenon which deals with such personal interest and the relationship. Virtual network built around any arbitrary person as shown in the Fig. 2, accompanying those persons with whom it has direct relationship is called ego network. Arbitrary person with respect to whom the network is drawn referred as ego and persons connected with ego are designated as alters. Connections between ego and alters are called ties.

Definition 2.1 (*Level 1.0 ego network*). If $G(V,E)$ is a network graph, then the ego network ψ or $\psi^{1.0}$ is defined as sub-graph $g(u,V_u,E_u)$ with respect to ego node u such that if $\exists v \in V$ then $v \in V_u$ iff $\exists (u,v) \in E$ and $\forall (x,y) \in E_u$ satisfies following conditions:

1. $x = u$
2. $y \in V_u$

Definition 2.2 (*Level 1.5 ego network*). If $G(V,E)$ is a network graph, then the ego network $\psi^{1.5}$ is defined as sub-graph $g(u,V_u,E_u)$ with respect to ego node u such that if $\exists v \in V$ then $v \in V_u$ iff $\exists (u,v) \in E$ and $\forall (x,y) \in E_u$ satisfies following conditions:

1. $x = u \vee x \in V_u$
2. $y \in V_u$

Definition 2.3 (*Level 2.0 ego network*). If $G(V,E)$ is a network graph, then the ego network $\psi^{2.0}$ is defined as sub-graph $g(u,V_u,E_u)$ with respect to ego node u such that if $\exists v \in V$ then $v \in V_u$ iff $\exists (u,v) \in E$ and satisfies following conditions:

1. $\forall (x,y) \in E_u$ implies $(x = u) \vee (x \in V_u)$ and $y \in V_u$
2. $V'_u = \bigcup_{\forall y \in V_u} (V_u, V_y)$ and $E'_u = \bigcup_{\forall y \in V_u} (E_u, E_y)$, where $g(V_y, E_y)$ is $\psi^{1.0}$
3. $V_u = V'_u$ and $E_u = E'_u$

An ego network comprises several relationships involving ego and different alters. Considering the personal interest factor, such relationship can be viewed as two way engagement of personal interest from both ego and alter's side as shown in the Fig. 3. The natural social community has to be started evolving somewhere within the network. Suppose, a community has started to evolve from any arbitrary person (ego). Definitely, $\psi^{1.0}$ alters are the next probable candidates for the community which has only one member at present i.e. the ego. From the set of $\psi^{1.0}$ alters of ego, only those alters will qualify for the entrance to the community which involve stronger relationship with the ego. As mentioned earlier, stronger relationship claims mutual consent from both ego and alter's side. Without consent of alter, the relationship between ego and alter cannot be stronger. This indicates the backward interest of alters are more