Neural Networks 58 (2014) 4–13

Contents lists available at ScienceDirect

Neural Networks

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

2014 Special Issue

Discrete particle swarm optimization for identifying community structures in signed social networks

Qing Cai, Maoguo Gong*, Bo Shen, Lijia Ma, Licheng Jiao

Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Xidian University, Xi'an, Shaanxi Province 710071, China

ARTICLE INFO

Article history:
Available online 13 May 2014

Keywords:
Signed social network
Community detection
Particle swarm optimization
Evolutionary algorithm

ABSTRACT

Modern science of networks has facilitated us with enormous convenience to the understanding of complex systems. Community structure is believed to be one of the notable features of complex networks representing real complicated systems. Very often, uncovering community structures in networks can be regarded as an optimization problem, thus, many evolutionary algorithms based approaches have been put forward. Particle swarm optimization (PSO) is an artificial intelligent algorithm originated from social behavior such as birds flocking and fish schooling. PSO has been proved to be an effective optimization technique. However, PSO was originally designed for continuous optimization which confounds its applications to discrete contexts. In this paper, a novel discrete PSO algorithm is suggested for identifying community structures in signed networks. In the suggested method, particles' status has been redesigned in discrete form so as to make PSO proper for discrete scenarios, and particles' updating rules have been reformulated by making use of the topology of the signed network. Extensive experiments compared with three state-of-the-art approaches on both synthetic and real-world signed networks demonstrate that the proposed method is effective and promising.

1. Introduction

The modern science of social networks is an active domain within the new interdisciplinary science of complex systems. In reality, many intricate systems can be represented as social networks, such as the complex collaboration networks (Newman, 2001), the world-wide-web (Albert, Jeong, & Barabasi, 1999; Broder et al., 2000), etc. Social networks are characterized by big data volume, dynamics and heterogeneous, especially for the Internet based social networks. Most of the social network data are natural language based. Data mining from social networks based on natural language is challenging (Cook & Holder, 2006; Kleinberg, 2007). Recent years, opinion mining and sentiment analysis become two of the most studied tasks of natural language processing, and have gathered momentum on both theoretical and empirical studies (Cambria, Schuller, Xia, & Havasi, 2013; Gangemi, Presutti, & Reforgiato Fortunato(2010), 2014; Poria et al., 2013).

One effective way to discover knowledge from a social network is to first model the network as a graph that is composed of a set of vertices and edges, where nodes represent the objects and links represent the interactions amongst them, and then apply certain techniques to analyze the properties of the graph based network. Network has many salient properties and amongst which the community structure is believed to be an eminent feature of networks (Girvan & Newman, 2002). In the academic domain, communities, also called clusters or modules, are groups of vertices which probably share common properties and/or play similar roles within the graph. Community detection is such a tool that helps to identify community structures in networks. Network community detection is of great significance. For example, mining cybercriminal networks from online social networks can facilitate cybercrime forensics so as to reduce the financial loss (Lau, Xia, & Ye, 2014). A recent survey on network community structure mining can be found in Fortunato (2010).

Mostly, the detection of community structures in networks can be considered either as a clustering problem or an optimization problem (Newman, 2004), thus, the choice of an appropriate evaluation function affects the ultima detection performance. For this purpose, Girvan and Newman (2002) had put forward the concept of modularity as a criterion to stop the division of a network into sub-networks in their divisive hierarchical clustering algorithm based on the iterative removal of edges with high betweenness. Based on modularity, other algorithms also appear in

* Corresponding author. Tel.: +86 029 88202661; fax: +86 029 88201023.
E-mail address: gong@ieee.org (M. Gong).

http://dx.doi.org/10.1016/j.neunet.2014.04.006
0893-6080/

the literature. Guimerò, Sales-Pardo, and Anaral (2004) employed simulated annealing for modularity optimization. Extremal optimization was used for modularity optimization by Duch and Arenas (2005). Spectral optimization technique has also been utilized to optimize modularity by replacing the Laplacian matrix with the modularity matrix (Newman, 2006). Besides modularity based methods, spectral clustering methods (Mitrovic & Tadic, 2009), dynamic approaches such as spin models (Reichardt & Bornholdt, 2004), random walk (Hughes, 1995) and synchronization (Boccaletti, Ivanchenko, Latora, Pluchino, & Rapisarda, 2007), methods based on statistical inference such as block modeling (Reichardt & White, 2007) and information theory (Ziv, Middendorf, & Wiggins, 2005), have all found their niche in this area.

Evolutionary algorithms (EAs), because of their inherent global searching abilities, hold an important position in the computational intelligence domain. EAs have been proved to be effective tools for solving optimization problems. Recently, some scholars have successfully applied either single or multiobjective evolutionary algorithms to discover community structures in networks. Pizzuti (2008, 2012) proposed a single objective genetic algorithm and a multi-objective genetic algorithm, respectively. In Gong, Fu, Jiao, and Du (2011) and Ma, Gong, Liu, Cai, and Jiao (2014) we had proposed a Memetic algorithm for community detection, and in Gong, Cai, Chen, and Ma (2014) we had suggested a multiobjective particle swarm optimization based approach.

In the field of social science, social networks with both positive and negative links are called signed networks (Doreian & Mrvar, 1996). In a signed network, the positive links denote "positive relationships" such as "friendship, common interests" and the negative links may denote "negative relationships" such as "hostility, different interests". To probe community structures in signed networks will shed light on how real society operates. Yang et al. have proposed a Markov random walk based algorithm called FEC to mine community structures in signed networks in Yang, Cheung, and Liu (2007). Doreian has designed an evaluation index to measure the quality of the partition of a signed network in Doreian (2008) and Traag and Bruggeman have applied the Potts Model (Wu, 1982) to solve the signed network community detection problem in Traag and Bruggeman (2009). Modularity is a very popular metric for community detection. Gómez, Jensen, and Arenas (2009) presented a reformulation of modularity that allows the analysis of signed networks. However, to maximize modularity is proved to be NP-hard (Brandes et al., 2006).

Although, in recent years many EAs based approaches have been developed to disclose communities in social networks, seldom of them have paid attention to signed networks. In this paper, we have newly suggested a particle swarm optimization (PSO) based algorithm to discover communities in signed social networks. PSO (Kennedy & Eberhart, 1995), originated from social animals' behavior, is well known by its fast convergence and has been proved to be one of the most popular optimization techniques. Due to its effectiveness and extremely easy implementation, PSO is gathering attention and it has found nationwide applications in diverse domains (Kiranyaz, Ince, Yildirim, & Gabbouj, 2009; Sharafi & ELMekkawy, 2014; Xu, Venayagamoorthy, & Wunsch, 2007). PSO works with a swarm of particles. Each particle adjusts its velocity by learning from its neighbors. This process is carried on simultaneously. Each particle can be seen as independent agents evolving in parallel, with some synchronizations. Thus, PSO can be regarded as an implicit parallel and distributed computational optimization algorithm, which makes it capable to handle large scale global optimization problems.

PSO is originally designed for continuous optimization which confounds its applications. In this paper, we redefine the particles' velocity and position and the main arithmetic operators between them in discrete form, consequently, a discrete PSO algorithm designed for identifying community structures in signed networks is proposed for the first time. Our method makes full use of networks' prior knowledge such as node degree information and linkage correlations. In order to speed up the algorithm convergence, a novel particle swarm initialization mechanism proposed in our previous work in Gong, Cai, Li, and Ma (2012) is adopted in this paper. Extensive experiments on both synthetic and real-world signed networks prove that the proposed algorithm is more efficient and much faster than several state-of-the-art approaches.

The rest of this paper is organized as follows. Section 2 gives the related background. In Section 3, the proposed method is presented in detail. Section 4 shows the experimental studies of the proposed method, and the conclusions are finally summarized in Section 5.

2. Related background

2.1. Community definition

The task for community detection is to separate the whole network into small parts which are called communities. In the literature, communities are regarded as subgraphs which have dense intra-links and sparse inter-links. Radicchi, Castellano, Cecconi, Loreto, and Parisi (2004) gave a community definition based on the node degree, but the community of a signed network is defined not only by the density of links but also by the signs of links. In Gong et al. (2014) we have suggested a signed community definition.

Given a signed network modeled as $G = (V, PL, NL)$, where V is the set of nodes and PL and NL are the set of positive and negative links, respectively. Let A be the adjacency matrix of G and l_{ij} be the link between node i and j. Then the element of A is defined as:

$$\begin{cases} A_{ij} = 1 & \text{if } l_{ij} \in PL \\ A_{ij} = -1 & \text{if } l_{ij} \in NL \\ A_{ij} = 0 & \text{if } \nexists l_{ij}. \end{cases} \tag{1}$$

Given that $S \subset G$ is a subgraph where node i belongs to. Let $(d_i^+)^{\text{in}} = \sum_{j \in S, l_{ij} \in PL} A_{ij}$ and $(d_i^-)^{\text{in}} = \sum_{j \in S, l_{ij} \in NL} |A_{ij}|$ be the positive and negative internal degrees of node i, respectively. Then S is a signed community in a strong sense if

$$\forall i \in S, \quad (d_i^+)^{\text{in}} > (d_i^-)^{\text{in}}. \tag{2}$$

Let $(d_i^-)^{out} = \sum_{j \notin S, l_{ij} \in NL} |A_{ij}|$ and $(d_i^+)^{out} = \sum_{j \notin S, l_{ij} \in PL} A_{ij}$ be the negative and positive external degrees of node i, respectively. Then S is a signed community in a weak sense if

$$\begin{cases} \sum_{i \in S} (d_i^+)^{\text{in}} > \sum_{i \in S} (d_i^+)^{out} \\ \sum_{i \in S} (d_i^-)^{out} > \sum_{i \in S} (d_i^-)^{\text{in}}. \end{cases} \tag{3}$$

Thus, in a strong sense, a node has more positive links than negative links within the community; in a weak sense, the positive links within a community are dense while the negative links between different communities are also dense.

The above definitions only give the conditions that a signed community should satisfy. In order to give a quantitative standard, Gómez et al. (2009) presented a reformulation of modularity that allows the analysis of signed networks. The signed modularity (SQ) is formulized as:

$$SQ = \frac{1}{2w^+ + 2w^-} \sum_{i,j} \left(w_{ij} - \left(\frac{w_i^+ w_j^+}{2w^+} - \frac{w_i^- w_j^-}{2w^-} \right) \right) \delta(i,j) \tag{4}$$

where w_{ij} is the weight of the signed adjacency matrix, $w_i^+ (w_i^-)$ denotes the sum of all positive (negative) weights of node i. If node i and j are in the same group, $\delta(i,j) = 1$, otherwise, 0. Normally by assumption we take it that the larger the value of SQ, the better the community structure is.