Physica A 471 (2017) 20–32

Contents lists available at ScienceDirect

Physica A

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

Modularity maximization using completely positive programming

Sakineh Yazdanparast *, Timothy C. Havens

Department of Electrical and Computer Engineering, Michigan Technological University, Houghton, MI, 49931, USA

A R T I C L E   I N F O

Article history:
Received 21 January 2016
Available online 23 November 2016

Keywords:
Community detection
Modularity maximization
Completely positive programming
Quadratic programming
Linear programming

A B S T R A C T

Community detection is one of the most prominent problems of social network analysis. In this paper, a novel method for *Modularity Maximization* (MM) for community detection is presented which exploits the *Alternating Direction Augmented Lagrangian* (ADAL) method for maximizing a generalized form of Newman's modularity function. We first transform Newman's modularity function into a quadratic program and then use *Completely Positive Programming* (CPP) to map the quadratic program to a linear program, which provides the globally optimal maximum modularity partition. In order to solve the proposed CPP problem, a closed form solution using the ADAL merged with a rank minimization approach is proposed. The performance of the proposed method is evaluated on several real-world data sets used for benchmarks community detection. Simulation results shows the proposed technique provides outstanding results in terms of modularity value for crisp partitions.

1. Introduction

Social network analysis has recently attracted a lot of attention. Online users in social networks such as Facebook and Twitter provide a veritable treasure trove of social network data, facilitating significant applications, such as product recommendation systems for on-line retail sites, political election prediction based on discussions of certain topics on Twitter, and so on. One important problem in social network analysis is community detection. A social network community is commonly defined as a group of nodes, i.e., users that have denser connections among the group's members than with the rest of the network. The process of finding network communities is called *community detection*.

Numerous approaches have been proposed for community detection. Fortunato [1] divided them into three different categories: traditional methods [2,3], divisive algorithms [4,5], and modularity based methods [6–20]. Traditional methods, such as graph partitioning and clustering, divide or merge clusters based on the similarity between nodes, whereas divisive algorithms are based on removing edges connecting nodes with low similarity. Modularity based methods work by optimizing an objective function such as the modularity introduced by Newman and Girvan [21]. Most of the recent works on community detection are based on modularity maximization proposed by Newman [7–12,18]. Other works have proposed other objective functions and solved them using extremal optimization [13], genetic algorithm [14,15], simulated annealing [16], *expectation–maximization* (EM) [17] and convex optimization [19].

Finding the partition with maximum modularity is very difficult (NP-hard [22]) due to its nonconvexity, and usually yields a sub-optimal partition, e.g., see Fast Unfolding Algorithm [15,23]. Some approaches, such as spectral optimization [3],

* Corresponding author.
E-mail address: syazdanp@mtu.edu (S. Yazdanparast).

http://dx.doi.org/10.1016/j.physa.2016.11.108

greedy methods [7,24,8], extremal optimization [13], and simulated annealing [16] have used searching to obtain solutions for crisp entries of the cover matrix. Although some approaches such as greedy methods, extremal optimization, simulated annealing, and spectral optimization have been used searching the global solution, but the proposed result is crisp. In most of social networks, many of users do not belong to a specific community which cause overlap among communities. To deal with this problem, the crisp overlapping and fuzzy overlapping community structures were proposed in [25]. Crisp overlapping communities let a node belong to more than one community; however, its membership still is binary. In fuzzy overlapping communities, memberships in communities are on the interval [0, 1], and the sum of the memberships for each node is 1. Several works have addressed fuzzy community detection [14,18,20].

In this paper, we present a novel model for reformulating the crisp modularity maximization problem. First, the objective function is converted to a linear programming problem with completely positive and rank-1 constraints. Then ADAL is applied to solve the CPP, followed by a rank minimization procedure to impose the rank-1 constraint on the final crisp solution. This contribution not only results in a highly-effective algorithm for modularity maximization, but also offers new insight on how to effectively solve the CPP problem with minimum rank. Burer [26] proposed reforming the standard quadratic problem with positive constraints to the linear programming with CPP constraints. However, it does not apply the rank-1 constraint which reconstructs the desired solution (the variable vector in the quadratic problem) from the final solution (the variable matrix in the CPP problem). To the best of our knowledge, our paper is the first work to address the modularity maximization problem by reforming its quadratic form to the linear programming problem. However, the proposed approach for solving the quadratic problem with positive constraints can be applied to any other standard problem. The main contributions of this paper can be summarized as follows:

1. Reformulation of the modularity maximization problem to a linear program with completely positive and rank-1 constraints;
2. Use of the ADAL method to solve the CPP;
3. Application of rank minimization algorithm to the ADAL method to minimize the rank of the ADAL output without contravening its optimality;
4. Application of the proposed method to several benchmark networks to investigate the optimality of the obtained cover matrices; and
5. Discussion on the limitations and scalability of the proposed algorithm for large scale networks.

The rest of the paper is organized as follows. Section 2 introduces the mathematical model of the community detection problem. The proposed algorithm for community detection is presented in Section 3. Section 4 presents experiments, analysis, and discussions; Section 5 concludes the paper. Table 1 contains a selected list of notations and symbols used in this paper.

2. Problem definition

2.1. Generalized modularity function

Every social network can be represented by a graph $G = (V, E, \mathbf{W})$, where V is a set of n vertices, E is a set of edges, and $\mathbf{W}$ is an $n \times n$ edge weight(or adjacency) matrix, where w_{ij} in $\mathbf{W}$ denotes the weight of the edge connecting node i and node j. Community detection for a network is the process of finding a $c \times n$ partition matrix (or in graph theory, a cover matrix) $\mathbf{U}$, where each element u_{ki} in $\mathbf{U}$, $k = [c]$; $i = [n]$, is the membership of the ith node in the kth community. There are three main types of partitions [27]:

$$M_{pc\times n} = \left\{ \mathbf{U} \in R^{c\times n};\, 0 \le u_{ki} \le 1, \forall\, k, i;\, \sum_{k=1}^{c} u_{ki} \le c, \forall\, i;\, \sum_{i=1}^{n} u_{ki} < n, \forall\, k \right\}; \tag{1a}$$

$$M_{fc\times n} = \left\{ \mathbf{U} \in M_{pc\times n};\, \sum_{k=1}^{c} u_{ki} = 1, \forall\, i \right\}; \tag{1b}$$

$$M_{hc\times n} = \left\{ \mathbf{U} \in M_{fc\times n};\, u_{ki} \in \{0, 1\} \right\}; \tag{1c}$$

where $M_{pc\times n}$ is the set of probabilistic; $M_{fc\times n}$ is the set of fuzzy, and $M_{hc\times n}$ is the set of crisp partitions. Much work has been done on crisp community detection, i.e., searching for the best $\mathbf{U} \in M_{hc\times n}$. Other works have focused on fuzzy community detection, i.e., searching for the best $\mathbf{U} \in M_{fc\times n}$. In this work, we propose a generalized community detection algorithm for finding partitions in $M_{hc\times n}$, deriving the crisp partition by hardening with the maximum membership rule. Recently, modularity based methods have been very popular among social network researchers. For the community detection problem, modularity works as the objective function to evaluate the goodness of a given community represented by a partition $\mathbf{U}$. Modularity was originally introduced by Newman and Girvan [28] as a way to evaluate crisp communities in networks. It is defined as

$$\mathcal{Q} = \frac{1}{\|\mathbf{W}\|} \sum_{k=1}^{c} \sum_{i=1,j=1}^{n} \left(w_{ij} - \frac{m_i m_j}{\|\mathbf{W}\|} \right) \delta(i, j), \tag{2}$$