Information Sciences 321 (2015) 1–13

Contents lists available at ScienceDirect

Information Sciences

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

Visualizing network communities with a semi-definite programming method

Wenye Li *

School of Public Administration, Macao Polytechnic Institute, Rua de Luís Gonzaga Gomes, Macao

ARTICLE INFO

Article history:
Received 12 April 2014
Received in revised form 27 March 2015
Accepted 23 May 2015
Available online 29 May 2015

Keywords:
Graph layout
Network modularity
Semi-definite programming

ABSTRACT

The existence of community structures is commonly believed in complex networked systems and has gained significant research attention in recent years. The automatic detection of network communities poses a non-trivial challenge due to the inherent computational requirements. In this paper, we investigate the problem from a different perspective and propose a novel model to visualize networks with the objective of exposing their community structures based on the idea of modularity maximization. The model is relaxed by a simple convex positive semi-definite program, which can be optimized efficiently. Compared with other visualization approaches, through empirical evaluation our method is able to highlight network communities and the adversary vertices therein effectively. Thereby, it provides a useful tool in the family of community detection algorithms and in the family of graph layout methods.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Many complex systems of scientific and engineering interest can be formulated as networks. Networks provide an effective way of expressing entities and their interactions through vertices and edges. Prominent examples of networked systems include the World Wide Web, social networks, biological networks and communication networks. With rapidly increasing needs from the real world, complex network analysis has gained significant research attention recently, in diverse areas such as computing sciences and information technologies [9,28,26].

Challenges are raised by network analysis, among which a significant one is to detect the community structures therein. Communities are commonly believed to exist in real networks. Vertices tend to fall in groups where connections within the same group are dense, whereas connections are sparse between vertices from different groups. The ability to find and analyze such groups has proved useful in understanding the formation and dynamics of networks and is invaluable with wide applications [9,28,15,31].

To divide vertices into groups, an influential measure, called modularity, is often used to quantify the quality of a given partition. Extensive study has shown that larger modularity scores are correlated with better vertex partitions [12]. Thus maximizing the modularity score provides a principled way to identify network communities [24,23,8,34,42,38,14,5]. At the same time, the mathematical optimization of the measure is challenging in both theory and practice; it involves expensive computation and becomes prohibitive for medium to large-sized networks [3].

* Tel.: +853 8599 6434.
E-mail address: wyli@ipm.edu.mo

http://dx.doi.org/10.1016/j.ins.2015.05.037
0020-0255/© 2015 Elsevier Inc. All rights reserved.

On the other hand, even if an automatically generated partition is available, human experts' effort is still inevitable in verifying the result. Bearing this in mind, our work starts from a different perspective. Instead of completely automating the process of community detection, we wish to visualize networks with the objective of highlighting the underlying community structures in a low-dimensional space. That is, compared to previous community detection approaches, our work provides a semi-automatic method. With the visualization, human experts can easily inspect the communities and make decisions of their own.

Our visualization method is based on the idea of modularity maximization. Computationally it can be relaxed as a simple positive semi-definite program, which provides an efficient and effective convex solution. Through empirical investigation, we observed key differences between our own and previous visualization approaches, and justified the usefulness of the proposed method.

The paper is organized as follows. We first introduce the background and related work briefly. Then, we present the proposed model and its computational relaxation in detail. The empirical evaluation results are given next, followed by conclusions and a discussion.

2. Background and related work

2.1. Network partition and modularity optimization

Network partition refers to a specific form of clustering that divides the vertices of a given network into groups according to the pattern of its edges. We commonly divide the vertices so that the groups formed are tightly knit with many edges inside groups.

Among network partition methods, the technique of "modularity maximization" is routinely used in practice [12,28]. It is performed by quantifying the quality of a given division of a network into communities by a measure of "modularity". Good divisions, which have high modularity values, are those with dense edge connections between the vertices within a community but sparse connections between vertices in different communities.

For simplicity, let us restrict the discussion to an undirected network $G = (V, W)$ where $V = \{v_1, v_2, \ldots, v_n\}$ is a set of vertices and $W = \{w_{ij}\}_{i,j=1}^{n}$ is an adjacency matrix with each w_{ij} giving the number of edges between two vertices v_i and v_j (or, edge weight). We further use $m_i = \sum_j w_{ij}$ as the degree of v_i and $m = \frac{1}{2}\sum_i m_i$ as the total edge number.

For a candidate partition of all vertices into n_c groups, its modularity is defined to be the portion of the edge connections within the same group minus the expected portion if the connections were distributed randomly. Under a random graph model with given expected degrees for the vertices, the expected number of edges between two vertices v_i and v_j is $\frac{m_i m_j}{2m}$ [7]. Thus the observed number minus the expected number is $w_{ij} - \frac{m_i m_j}{2m}$. Summing over all pairs of vertices within the same group, the modularity, denoted by Q, is given by

$$Q = \frac{1}{2m}\sum_{ij}\left[w_{ij} - \frac{m_i m_j}{2m}\right]\delta(c_i, c_j) \tag{1}$$

where c_i is an integer within $1, \ldots, n_c$ indicating the group of vertex v_i, and δ is the Kronecker delta function.

The value of Q is strictly less than 1. It is positive if there are more edges between vertices within the same group than one would expect by chance. It is negative otherwise. Given a larger than expected portion of connections, we infer the presence of an inherent community structure. Therefore, looking for divisions with high modularity values provides a precise way to detect network communities [28].

Unfortunately seeking a partition that has the highest modularity score is usually difficult. It can be proved to be computationally NP-hard [3], and the complexity needed by the optimal solution grows exponentially with the problem size. Thus approximated solutions have to be sought if tractability is to be ensured [12,17,28]. The quality and efficiency of such approximations have posed a significant challenge in practice.

2.2. Positive semi-definite programming

Semi-definite programming (SDP) is a relatively new subfield of convex optimization, and dramatic development has been made in its theory and practice since the 1990s [27,37].

SDP is concerned with optimization problems over symmetric positive semi-definite matrix variables with a linear objective function and linear constraints. Denote by $\mathbf{S}^n$ the space of all $n \times n$ real symmetric matrices, equipped with an inner product $\langle S_1, S_2\rangle = tr\left(S_1^T S_2\right)$ where $S_1, S_2 \in \mathbf{S}^n$ and tr denotes the trace of a square matrix. A matrix $S \in \mathbf{S}^n$ is positive semi-definite if all its eigenvalues are nonnegative; we write $S \succeq 0$. Given $\{A, A_1, \ldots, A_m\} \subset \mathbf{S}^n$ and scalars $a_1, \ldots, a_m$, SDP maximizes a linear objective of the type

$$\max_S tr\left(A^T S\right) \tag{2}$$

subject to

Download English Version:

https://daneshyari.com/en/article/391959

Download Persian Version:

https://daneshyari.com/article/391959

Daneshyari.com