Information Sciences 473 (2019) 31–43

Contents lists available at ScienceDirect

Information Sciences

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

Direction recovery in undirected social networks based on community structure and popularity

Yi-Ming Wen, Ling Huang, Chang-Dong Wang*, Kun-Yu Lin

School of Data and Computer Science, Sun Yat-sen University, Guangzhou Higher Education Mega Center, Panyu District, Guangzhou, Guangdong 510006, China

ARTICLE INFO

Article history:
Received 5 June 2018
Revised 10 September 2018
Accepted 17 September 2018
Available online 18 September 2018

Keywords:
Direction recovery
Popularity
Community detection
Behavioural similarity

ABSTRACT

Directionality is a significant property for analyzing and understanding social networks. Unfortunately, the potential directionality is hidden in undirected social networks. The previous studies on recovering directionality in undirected social networks are mainly based on the microscopic patterns discovered in the existing directed social networks. In this paper, we attempt to recover the directionality by considering the macroscopic community structure. To this end, a variant of the existing modularity model, called behavioural modularity, is designed for discovering community membership of nodes. Assuming that members in the same community have higher behavioural similarity, we introduce the concept of the intra-community popularity, and then estimate the directionality of undirected ties based on the community structure and the intra-community popularity. Accordingly, we propose a novel Community and Popularity based Direction Recovery (CPDR) approach to recover the directionality of undirected social networks. Experimental results conducted on three real-world social networks have confirmed the effectiveness of the proposed approach.

1. Introduction

With the rapid development of modern information technology, a large number of networks are generated, such as e-mail networks [21], file sharing networks [16], voting networks [5,6], etc. In these networks, directionality is a significant property for many network analysis tasks, such as link prediction [2,9], community detection [4,10,17], product recommendation [12] and consensus analysis [7,8], etc. However, the directionality is hidden in undirected networks. Our task is to recover the directions of undirected social networks.

The previous studies about direction recovery are mainly based on microscopic patterns [14,23,24]. In [23,24], by analyzing the existing directed networks, Zhang et al. discovered four patterns, based on which a direction recovery approach was developed by minimizing the global inconsistency to these patterns. However, these patterns do not always exist in all networks, which results in degenerated performance. In this paper, to recover the directions of networks, instead of microscopic patterns, we focus on macroscopic community structure.

Modularity is a popular model for measuring the community structure in both directed [4] and undirected networks [13]. In the previous study of modularity in directed networks [4], directionality is used to measure the connection strength

* Corresponding author.
E-mail addresses: yi-Ming.wen@outlook.com (Y.-M. Wen), huanglinghl@hotmail.com (L. Huang), changdongwang@hotmail.com (C.-D. Wang), kunyulin14@outlook.com (K.-Y. Lin).

https://doi.org/10.1016/j.ins.2018.09.036
0020-0255/

between nodes. However, more information is hidden in directionality, e.g. behaviour of individuals. Nodes belonging to the same community are of higher behavioural similarity. For instance, in a citation network, physicists are in the same community. They may cite a number of significant research papers about physics, leading to higher similarity in citation behaviours. In undirected networks, a tie between two individuals indicates a high possibility to be in the same community. However, it does not work in directed networks, since in a directed relationship, two individuals have different behaviours, one as the source while the other as the target. Therefore, we need to measure modularity by taking into account behavioural similarity.

In a community, some individuals are regarded popular when a large number of other individuals in this community propose a relationship to them. For example, in the communication networks of classes, the teachers or those who are helpful in solving problems are popular since many other individuals send them messages to ask questions. Chances are high that a relationship is proposed towards those with high popularity. The concept of popularity has been studied in community detection [19,20] and link analysis [20].

From the perspective of macroscopic community structure, we propose a novel Community and Popularity based Direction Recovery (CPDR) approach to recover the directions in undirected social networks. In particular, a variant of the existing modularity model, called behavioural modularity, is designed for recovering community membership of nodes, which is suitable for the task of recovering directions. Moreover, we introduce the concept of the intra-community popularity and then estimate the direction probability based on the resulting community membership and the intra-community popularity of nodes. Finally, the directions are recovered based on the direction probability. Experiments are conducted on the three real-world datasets and the results have confirmed the effectiveness of the proposed method.

The main contributions of this paper are summarized as follows.

- We propose an approach to estimate the direction probability based on the community membership and the intra-community popularity of nodes.
- We introduce a new method to measure the modularity in directed networks by taking into account the behavioural similarity of nodes.
- We design a mathematical trick to convert the optimization problem with constraints into an optimization with a weaker constraint. And we provide a theoretical proof of the validity of the conversion.

The rest of this paper is organized as follows. In Section 2, we will briefly review the related work. In Section 3, we will introduce in detail the proposed model, including the behavioural modularity and the estimation of direction probability. In Section 4, we will introduce the mathematical tricks designed to solve the optimization problem. In Section 5, a clear framework of our algorithm will be presented and the time complexity will be analyzed. In Section 6, experiments will be conducted on the three real-world datasets to confirm the effectiveness of the proposed method. Finally, the conclusions and future work will be presented in Section 7.

The main results in this paper were first presented in [18].

2. Related work

Direction recovery was first studied in [23], which was later extended in [24]. The early work [23] mainly focused on microscopic patterns. By analyzing the existing directed networks, four patterns were discovered, namely degree consistency pattern, triad status consistency pattern, similarity consistency pattern and collaborative consistency pattern. Accordingly, some algorithms were proposed to recover the directions of undirected social networks by minimizing the inconsistency to these patterns. Besides, the supervision information was utilized for direction recovery under the semi-supervision and self-supervision frameworks. In the extended work [24], different from the original node-centric approach, a new tie-centric approach was developed, which exploited the same patterns, but iterated from the perspective of each tie. Furthermore, Peng et al. [14] proposed to combine the node attributes and the microscopic patterns, aiming to achieve better direction recovery performance by leveraging more information. Despite success, the aforementioned methods mostly focus on microscopic patterns. However, these patterns do not always exist in all networks, which results in poor performance. In this paper, to recover the directions of networks, instead of microscopic patterns, we focus on macroscopic community structure.

Modularity has been widely used for community detection [3,4,11,13]. Some variants of modularity have been developed for different types of networks, e.g. directed networks [4] and undirected networks [13]. The main idea of modularity is to calculate the total strength of connection between nodes within the same community. By maximizing modularity, similar community membership can be assigned to nodes with higher strength of connection. However, in directed networks, the strength of connection between two nodes does not indicate a higher similarity in their community membership, since the two nodes of an edge have different behaviours. The direction of an edge can be viewed as the behaviour of two nodes. Therefore, to take into consideration the behavioural similarity of nodes, we propose the idea of behavioural modularity. Through maximizing such modularity, nodes with higher similarity can be assigned with similar community membership, which facilitates our work in direction recovery.

Some efforts have been made in utilizing popularity and community structure to measure the link probability of edges [19,20]. In Popularity Conditional Link Model (PCL) [19], given the community membership of node i, denoted as C_i, its probability pointing to node j is measured by calculating the ratio of node j's popularity within C_i to the total popularity of neighbour of node i within C_i. Similarly, a new formula was proposed in Popularity Productivity Link Model (PPL) [20] based