Available online at www.sciencedirect.com

ScienceDirect

CAAI Transactions on Intelligence Technology 1 (2016) 114–123
http://www.journals.elsevier.com/caai-transactions-on-intelligence-technology/

Original article

An efficient shortest path approach for social networks based on community structure☆

Maoguo Gong [a,*], Guanjun Li [a], Zhao Wang [a], Lijia Ma [a], Dayong Tian [b]

[a] *Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Xidian University, Xi'an 710071, China*

[b] *Center for Quantum Computation and Intelligent Systems, University of Technology, Sydney, Broadway, NSW 2007, Australia*

Available online 2 June 2016

Abstract

Finding the shortest path (SP) in a large-scale network analysis between any two nodes is a tough but very significant task. The SP can help us to analyze the information spreading performance and research the latent relationship in the weighted social network, and so on. As the size of the social network increases, the traditional SP algorithms have poor performance and there is not a suitable algorithm for weighted social network. Some features of the network analysis are beneficial to solve this problem, and community structure ignored by the traditional methods is one of the most important features. In this paper, we propose a shortest path algorithm based on community detection (SPCD) by integrating community detection algorithm with traditional search methods. SPCD constructs a community graph by using community structure to narrow the searching scope. The algorithm presented improves the time efficiency and maintains the accuracy scale of the SP. Experimental results on five real-world networks demonstrate the effectiveness of the proposed methods for the SP problem.
Copyright © 2016, Chongqing University of Technology. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Shortest path; Community structure; Weighted social network

1. Introduction

Recently, studies on social network have attracted a lot of attention from sociology, informatics and computer science [1,2]. A social network is a social structure made up of a set of social actors (such as individuals or organizations) and a set of the dyadic ties between these actors. A huge amount of data and resources make it critical to analyze social network related problems. A series of research on social network have been done in many aspects of sociology, such as social influence, social groupings, inequality, disease propagation and communication of information [3]. The SP problem which calculates the distance or relationship between the nodes in the social network is an extremely important problem of social network analysis. It can be used to study the behavior of information spreading, especially fastest information spreading. The recommendation system is also based on the SP problem. For example, when we analyze the relationship between scientists, the scientists collaboration network is established and analyzed utilizing the SP problem. Analyzing the scientist collaboration network, the work and paper reviewing efficiency can be improved. The SP problem serves as an elementary aspect when analyze the network structure, for example betweenness and average distance. With the increase of network scale, the previous algorithms may not suitable for the existing network, so we need effective methods for large-scale network. In this paper, we focus on the SP problem between any two nodes in the weighted social network. The SP problem concerns with finding the shortest path from a

☆ This work was supported in part by the National Natural Science Foundation of China under Grants 61273317, and 61422209, the National Program for Support of Top-notch Young Professionals of China, and the Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20130203110011.

* Corresponding author.
E-mail address: gong@ieee.org (M. Gong).

Peer review under responsibility of Chongqing University of Technology.

http://dx.doi.org/10.1016/j.trit.2016.03.011
2468-2322/Copyright © 2016, Chongqing University of Technology. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

specific origin to a specified destination in a given network while minimizing the total cost associated with the path. SP is a fundamental problem in social network. In a graph, finding the path with the minimum cost from a source node s to a destination node d is called the point-to-point (P2P) problem, but a common variant fixes a single node as the source node and finds shortest paths from the source to all other nodes in the graph. In addition to P2P problem, other shortest path problem, such as single-destination, all-pairs and so on, could be converted to single-source shortest path. The single-source problem with non-negative arc lengths has been studied most extensively [4]. For a general network, traditional dijkstra's algorithm could be used for solving all the non-negative, weighted or non-weighted networks without any domain-specific information. Though it is a single-source algorithm, we can transform it to a P2P algorithm by terminating at destination node. But it has high time complexity of $O(n^2)$, so it might not work well on the SP in social network, especially in large-scale social network. The time complexity of modified standard breadth-first is $O(n)$, but it is suited to non-weighted network [5]. With the increase of network scale, authors used domain-specific information about the network to deal with it. In traffic networks, researchers have adopted the natural hierarchies to speed up a shortest path algorithm significantly [6–10], and more algorithm are given in [11]. It is obvious that these algorithms could not be applied to social network. From the same point of view in traffic networks, the feature of social network can be utilized to the SP problem. Community feature is one of the most important characteristics of social network.

In recent years, community detection has gained a lot of attention in the field of network analysis. Community structure is similar to the small world effect and the right-skewed degree distributions which are an important distinctive property in a complex network [12]. Qualitatively, a community is defined as a subset of a graph in which the interconnections of nodes are denser than those observed with the rest of the network [13,14]. For the general case of a weighted graph, many approaches mainly focused on various criteria including modularity [15], structural density [16] and partition density [17]. Blondel et al. introduced a fast greedy approach (BGLL) to optimize the modularity [18]. Besides, parameter-free hierarchical network clustering algorithm (SHRINK) proposed by Huang et al. combines the advantages of density-based clustering and modularity optimization methods [19]. In the field of social networks, Xie et al. proposed a general speaker-listener algorithm named SLPA based on label propagation [2]. Evolutionary computation is also an important and influential method [20,20,21]. Community detection clusters the edges into two classes: the shorter ones and the longer ones. In that case, we can pay our attention to the large edges and find out the communities including the SP, and then searching the SP in these communities. As far as we know, there is no method using community information to solve P2P problem in weighted social network. In this paper, we propose a new shortest path algorithm based on community detection (SPCD). Using community information, SPCD can narrow down the search space and decrease computing time for a path while sacrificing accuracy within a certain level. It can strike a balance between accuracy and time efficiency while being faced with different problems.

The paper is organized as follows. Section2 describes the proposed SPCD algorithm in detail. Section3 is experiment in five datasets. This subsection analyses accuracy and efficiency. Moreover the influence of community path number and community detection method to SPCD is discussed. We conclude the paper is Section4 and point out the future work.

2. Algorithm description

In this section, the proposed algorithm is described in detail. First, the community definition and detection is introduced, and then the method of constructing community graph is described. The k shortest community paths are given in subsection2.3. Finally, the shortest path in sub-graph is given.

2.1. Community definition and detection

Adopted symbols are listed in Table 1. Communities are groups of vertices which share common properties and/or play similar roles within the graph. Qualitatively, connections between the nodes in a community are denser and closer than connections with the rest of the network. For SP in weighted social network, we define the community in such a way that the connections in the community have a lower cost but connections between the communities have higher cost. In this paper, we use distance to measure the relationship of any two nodes in the network. The smaller the distance, the more highly related the any two nodes. Thus the target of the SP problem is to find the minimal distance between any two nodes in the network. As for the other applications which need to find the maximum the distance, they can be easily transformed into this minimization problem by reverting the evaluation function.

A demonstration of community in a network is shown in Fig. 1. The weight of edge between j and l is 20, which is

Table 1
Notation.

Symbols	Definition
$G=(V,E)$	Target network, where $V=\{v_1,v_2,\ldots,v_n\}$ is the set of vertices and E is the set of edges
$e_{i,j}\in E(i,j\in V)$	The non-negative edge between vertice i and j
s,d	Source node and destination node
C	Community information of network
$c_i(i\in V)$	Vertice i's community
$\overline{c}$	The set of nodes in community c
$\overleftrightarrow{E_{i,j}}$	The set of edges between community i and j
$SG=(SV,SE)$	Sub-graph,SV is the set of vertices and SE is the set of edges
P_{ij}	The path from i to j
SP	The shortest path
$\|SP\|$	The weighted sum of SP
$KSCP$	The shortest path of community graph
SV'	Resort the SV
SP'	The shortest path in sub-graph

Download English Version:

https://daneshyari.com/en/article/378345

Download Persian Version:

https://daneshyari.com/article/378345

Daneshyari.com