Information Processing and Management 000 (2016) 1–20

Contents lists available at ScienceDirect

Information Processing and Management

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

Measuring the social influences of scientist groups based on multiple types of collaboration relations

Jiuchuan Jiang[a,*], Peng Shi[b], Bo An[a], Jianyong Yu[b], Chongjun Wang[c]

[a] School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798, Singapore
[b] School of Computer Science and Engineering, Southeast University, Nanjing 211189, China
[c] National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210046, China

A R T I C L E I N F O

Article history:
Received 14 April 2015
Revised 14 April 2016
Accepted 10 June 2016
Available online xxx

Keywords:
Scientist groups
Scientific collaboration
Multiple types of relations
Social influence

A B S T R A C T

Scientists often collaborate with each other and may produce social influences through their collaboration on scientific activities. While the subject of ranking scientists has received significant attention in previous studies, these studies often examine the social influences of individual scientists and are based on the assumption that all collaborations between scientists are of the same type. However, these two limitations do not match real scientific collaborations. Currently, scientists are often associated with groups in which the scientists always study related research topics or have close collaboration relationships. Moreover, current scientists often collaborate through multiple relationship types, and different types of relationships may have different effects on the social influences of the scientists. To solve these two problems, this paper presents a model that measures the social influences of scientist groups based on multiple types of collaboration relationships. The model addresses two general group types (hierarchical and nonhierarchical) and two general collaboration situations: one is that the multiplex collaboration relationships are independent, and the other is that the multiplex collaboration relationships are correlated. The model mainly adopts the linear weighted sum methodology, which can make the time complexities of key algorithms low. Finally, we create some case studies and make experiments to demonstrate the effectiveness of our model.

1. Introduction

Collaboration is a critical aspect of scientific research (Hara et al., 2003; Jiang, 2008), which indicates that scientists may collaborate with each other to conduct certain scientific activities (Abbasi, Hossain, Uddin, & Rasmussen, 2011; Giuliani et al., 2010; Hou, Kretschmer, & Liu, 2008; Newman, 2004). The basic reasons for collaboration among scientists are that no scientists are versatile and each scientist may only have certain resources or abilities for a scientific activity. A scientist can profit from collaboration with other scientists, as well as benefit other scientists, which can improve the output of the overall scientific community (Jiang, 2008). In reality there are many types of collaboration among scientists, such as collaborate to conduct a large project, co-author a paper, and collaborate to hold a conference.

* Corresponding author.
E-mail addresses: jiangjiuchuan@163.com (J. Jiang), pengshi@seu.edu.cn (P. Shi), boan@ntu.edu.sg (B. An), yujyong@126.com (J. Yu), chjwang@nju.edu.cn (C. Wang).

http://dx.doi.org/10.1016/j.ipm.2016.06.003
0306-4573/

Generally, scientific collaboration relationships among scientists form networks (Abbasi et al., 2011; Hou et al., 2008; Newman, 2001b; Zhao et al., 2014). Many data mining techniques (Huang et al., 2014) or graph theory methods (Zhang et al., 2011) can be used to analyze the scientific collaboration networks. Each scientist occupies a social position against other scientists in the scientific collaboration network, which can be shaped and described not only by the scientist's own resources but also by the scientist's relationship with other scientists (Jiang and Huang, 2012; Jiang et al., 2011). Measuring the social influences of scientists, i.e., the ranking of scientists, is an important problem for the effective management of science and technology. Although there are many factors that may be related to the social influence of a scientist (Root-Bernstein et al., 1995), the number of citations of scientific articles has been a widely used metric in previous studies to evaluate the influence of a scientist (Zhou et al., 2012). One famous metric to measure the influence of a scientist is h-index, where a scientist with T articles has an h-index equal to h if h of his/her articles received at least h citations each, and the other T-h articles each received no more than h citations (Egghe & Rousseau, 2006).

Those existing works that measure the influences of scientists always follow two assumptions: First, the ranking metrics were designed mainly to evaluate the influences of individual scientists; second, all scientific collaboration relations among scientists are of the same type, i.e., *the connections among scientists are simplex*. However, in many circumstances these two assumptions do not match the peculiarities of real scientific collaborations, which we introduce in brief below.

In a large scientific collaboration network, scientists may be organized into groups. Scientists adjust to the problems in their research fields by forming social groups of various types based upon shared communication and shared interpretations of the situation (Crane, 1969). A scientist group is comprised of scientists who work on similar research problems. A group in a paper co-authorship network might represent related scientists collaborating on papers with some related or similar topics (Newman, 2004); a group in a research project collaboration network might represent a set of scientists that work on the same research project (Jiang, 2008). A typical example of scientist groups is the SIGA (Special Interest Groups) in ACM (Association for Computing Machinery), which offers a wealth of conferences, publications, and activities to provide opportunities for sharing technical expertise and first-hand knowledge of the latest development trends in a special area of computing. Due to the group organization, the scientists often partake in group decision-making and group activities when they work on scientific collaborations. Therefore, a group of scientists may produce the social influence of the entire group and not only the individual scientists in the scientific collaboration network.

There are multiple types of collaboration relationships among scientists. Two scientists may coauthor a paper and work on a cooperative project at the same time, or two scientists may cooperate to hold a conference and simultaneously apply for research funding, etc. Real scientific collaboration networks often appear as multiplex networks, in which the scientists are connected by different types of collaboration relationships. Due to the multiple types of relationships, the social influences of scientist groups may have varying meanings in reality. For example, a group whose papers are seldom cited by others may be influential because such group has many practically valuable patents; a group in which some members are not active in publication but other members are active in conducting many large engineering projects may also have higher social influence. Measurements of the social influences of scientists should consider the effects of the group and the multiple types of collaboration relationships.

To solve the above two problems, this paper presents a model that measures the social influences of scientist groups based on multiple types of collaboration relationships. The model addresses two general group forms (hierarchical and non-hierarchical) and includes two measuring methods: one measures the social influences of groups by considering the multiplex types of collaboration relationships *independently*, and the other measures the social influences of groups by *correlating* multiplex types of collaboration relationships. *In comparison with related work, our study makes the following contributions: 1) we are concerned about the social influences of scientist groups, which addresses the limitation in previous studies that only considered the social influences of individual scientists; 2) we measure the social influences of scientist groups by analyzing the multiple types of collaboration relationships among scientists, which addresses the limitation in previous studies that only measured the social influence from one type of collaboration.*

The rest of this paper is organized as follows: In Section 2, we introduce the related work on the subject; in Section 3, we give an overview of our methodology; in Section 4, we model the scientist groups with multiple types of collaboration relationships; in Section 5, we present the model for measuring the social influences of scientist groups; in Section 6, we make case studies to demonstrate our model; in Section 7, we make experiments to test the effectiveness of our model; finally, we discuss and conclude our paper in Section 8.

2. Related work

Our research is related to the scientific collaboration among scientists and the social influences of scientists. Generally, related work can be categorized as follows:

2.1. Scientific collaboration among scientists

The basic reason for collaborations in scientific research is that an individual scientist can seldom provide all of the expertise and resources necessary to address complex research problems. Scientific collaborations may be different from other varieties of collaboration because they are shaped by social norms of the practice, the structure of knowledge, and the technological infrastructure of the scientific discipline (Hara et al., 2003). In previous studies on collaboration among