Decision Support Systems 59 (2014) 1–14

Contents lists available at ScienceDirect

Decision Support Systems

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

A multi-criteria ranking framework for partner selection in scientific collaboration environments

Daniel Schall

Siemens Corporate Technology, Siemensstrasse 90, 1211 Vienna, Austria

ARTICLE INFO

Article history:
Received 11 June 2013
Received in revised form 10 September 2013
Accepted 1 October 2013
Available online 9 October 2013

Keywords:
Scientific communities
Multi-criteria selection
Time-aware authority ranking
Structural holes

ABSTRACT

Scientific collaborations commonly take place in a global and competitive environment. Coalitions and project consortia are formed among universities, companies and research institutes to apply for research grants and to perform jointly collaborative projects. In such a competitive environment, individual institutes may be strategic partners or competitors. Measures to determine partner importance have practical applications such as comparison and rating of competitors, reputation evaluation or performance evaluation of companies and institutes. Many network-centric metrics exist to measure the importance of individuals or companies in social and collaborative networks. Here we present a novel approach for measuring and combing various criteria for partner importance evaluation. The presented approach is cost sensitive, aware of temporal and context-based partner authority, and takes structural information with regard to structural holes into account. Well-established graph models such as the notion of hubs and authorities provide the basis for the presented authority ranking approach and are systematically extended towards a novel unified HITS/PageRank model. The applicability of the proposed approach and the effects of parameter selection are extensively studied using real data from the European Union's research program.

1. Introduction

Scientific collaboration in an international environment takes place among partners such as organizations, universities or research institutes to jointly perform projects. The main motivation for organizations and individual research groups to collaborate is to enable knowledge and resource sharing to effectively perform research projects. Scientific collaboration can be defined as *interaction taking place within a social context among two or more scientists that facilitates the sharing of meaning and completion of tasks with respect to a mutually shared, superordinate goal* [33].

However, the success of research and innovation is based on the right balance between cooperation and competition. Hence, formation of coalitions and consortia is influenced by partner reputation [14], institutional constraints, and mechanism of self-organization [35]. Scientific collaboration can be analyzed at the level of researchers through co-authorship and citation networks [11,17,26] or at the level of organizations or research institutions [23]. The former has been widely studied by existing research while the latter lacks a principled approach for selecting and aggregating ranking criteria that may be influenced by context. Generally, scientific collaboration and endorsement can be analyzed according to three different methods [24]: (i) qualitative methods such as using a questionnaire-based approach, (ii) bibliometric methods including publication and citation counting or co-citation analysis, and (iii) complex network methods including network centrality metrics such as PageRank [28] or Hyperlink Induced Topic Search (HITS) [21]. Here we focus on the analysis of scientific collaboration at the organizational or institutional level. We apply complex network methods to automate the analysis of partner importance in scientific collaboration. In this work, importance is a concept that is governed by multiple factors including average cost of a partner, temporal trend and context of partner authority, and partner importance with regards to effective size of the partner's social network. Effective size in the context of structural holes and social networks means low redundancy among social contacts thereby yielding control benefits of individuals. Here we apply a similar principle but focus on the organizational level rather than individuals in social networks.

In our previous work [32] we introduced an approach for measuring contextual importance in scientific collaboration networks. In this work, we build upon our previous work [32] but significantly expand the concepts. Here we provide the following novel key contributions:

- We introduce a personalized partner authority model that is able to capture context-dependent and time-aware partner reputation.
- We introduce a model to measure structural importance of organizations embedded in scientific collaboration networks. The idea of our structural importance metric is drawn from the notion of structural holes as established in a sociological research context.
- To support partner selection using multiple-criteria, the factors contributing to a partner importance are aggregated through a systematic approach to a single partner importance ranking score. Here we apply

E-mail address: daniel.schall@siemens.com.

0167-9236/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.dss.2013.10.001

analytic hierarchy process (AHP) to derive the partner importance score.
- We present experimental results by providing a comprehensive study on the influence of different parameters using real data from the EUs Seventh Framework Programme (FP7) for research in Information and Communication Technology (ICT).

This work is structured as follows. Section 2 gives an overview of related work and literature in the context of network formation and network analysis. Section 3 introduces basic concepts and definitions used throughout this work. In Section 4 our personalized partner authority model is introduced. Section 5 introduces the structural importance model and Section 6 details the analytic hierarchy process to compute the final partner importance scores. In Section 7 the evaluation results are presented followed by the conclusion and outlook to future work in Section 8.

2. Literature overview

We structure related work into two basic areas: *network formation* in the context of collaborative environments and *network analysis* methods with particular emphasis on authority ranking. From a technique point of view, many approaches found in both network formation and network analysis methods for authority ranking are based on graph theory and algorithms. In this section, we review literature in both areas as they will provide the foundation for our work.

2.1. Network formation

The rapid advancement of ICT-enabled infrastructure has fundamentally changed how businesses and companies operate. Global markets and the requirement for rapid innovation demand for alliances between individual companies [7]. It is widely agreed that knowledge of the structure of interaction among individuals or organizations is important for a proper understanding of a number of important questions such as the spread of new ideas and technologies and competitive strategies in dynamic markets [15]. Work by [34] investigated the evolutionary dynamics of network formation by analyzing how organizational units create new linkages for resource exchange. The potential gains from bridging different parts of a network were important in the early work of Granovetter [16] and are central to the notion of structural holes developed by Burt [5,6]. The theory is based on the hypothesis that individuals can benefit from serving as intermediaries between others who are not directly connected. A formal approach to strategic formation based on advanced game-theoretic broker incentive techniques was presented in [22]. In [2] group formation in social networks is studied.

2.2. Network analysis

We propose a model for importance that is based on well-established techniques such as the notion of hubs and authorities [21] and PageRank [28]. PageRank can be personalized [28] to estimate node importance with regard to certain topics [18–20]. After the seminal work of [28] and the far-reaching work of [19], related research (see also [4]) addressed, for example, efficient computation of personalized PageRank [9,13] and a generalization of personalized PageRank towards bipartite graphs [10]. In [3], the authors proposed time-aware authority ranking by considering temporal properties of scientific publication activity. Our previous work addressed PageRank personalization techniques for expertise ranking in a social network context [30,31].

In this work, we propose a new framework which utilizes both information from structural holes and authority importance scores to discover valuable collaboration partners. Here we propose a unified HITS/PageRank model that is able to measure network importance at the individual as well as the organizational or institutional level with respect to a certain context. In contrast to existing rankings such as the Shanghai academic ranking,[1] our approach is able to capture importance at a fine grained contextual level. Our approach is able to utilize various additional ranking parameters including desirable partner properties (e.g., high topic-sensitive authority) and low undesirable partner properties (e.g., partner costs). At the core of this framework are link-based algorithms such as HITS and extensions towards personalized, time-aware PageRank, structural metrics to measure the brokerage potential of a given network node, and an analytic hierarchy process (AHP) algorithm [29] to aggregate these metrics into a single ranking score.

The proposed model is tested with data from the ICT research projects having received grants under the EU's FP7 program. The data as described in [25] and covers a period from 2007 to 2011.

3. Definitions and solution framework

3.1. Basic definitions

We start with a definition of basic concepts that are used throughout this work. Let us consider a simple collaboration scenario in a scientific community where individual partners (e.g., organizations, research institutes, and universities) collaborate in the context of research projects. Fig. 1 depicts a set of organizations $\{o_1, o_2, o_3\}$ and a set of research projects $\{p_1, p_2, p_3\}$. Each project is associated with a certain topic that determines the *context* of the performed collaboration (for example, 'services' or 'internet'). Organizations are involved in projects by having certain roles. Roles include project coordinator and project partner. In addition to the involvement relation, a weighted edge is created from the project to the organization to depict the degree of involvement. For example, o_1 is involved in projects p_1 and p_2 with weights w_{11} and w_{21} respectively. In our work, the weight will be based on the funding an organization receives in the context of a project. More funding typically means that an organization is able to allocate more (human) resources to the project and thereby perform more work. Finally, based on joint projects performed by organizations we model collaboration relations among them. Since o_1 and o_2 have been involved in the joint projects p_1 and p_2, a collaboration relation between o_1 and o_2 is established as a dashed line. Similarly, o_2 and o_3 have been involved in the joint projects p_2 and p_3 and therefore a collaboration relation between o_2 and o_3 is established. Also, a collaboration relation between o_1 and o_3 exists because they jointly worked on p_2. A collaboration relation is a mutual (undirected) edge. The applications of the presented concepts will be illustrated in the next section.

3.2. Solution framework

As already outlined before, our solution approach to support multi-criteria partner selection in scientific communities utilizes heavily graph-based models. Graph-based models are widely used in complex- and social-network analysis. Fig. 2 shows the solution framework as a layered view.

3.2.1. Data management

The layer underneath the top-layer shows the data management that is responsible for retrieval of project relevant data, managing the needed graph structures to perform analysis and ranking, and persistence management of analysis and ranking results. From the top-layer (Offline analysis) point of view, the data management can be accessed via the Data Manipulation Handler in a CRUD (Create-Read-Update-Delete) manner. The Data Provider offers read access to graph structures and offline mining and ranking results. The Project Database contains information such as organizations, projects, project involvements,

[1] www.shanghairanking.com.