



# 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 combining 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.

© 2013 Elsevier B.V. All rights reserved.

## 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](mailto:daniel.schall@siemens.com).

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 EU's 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,<sup>1</sup> 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,

<sup>1</sup> [www.shanghairanking.com](http://www.shanghairanking.com).

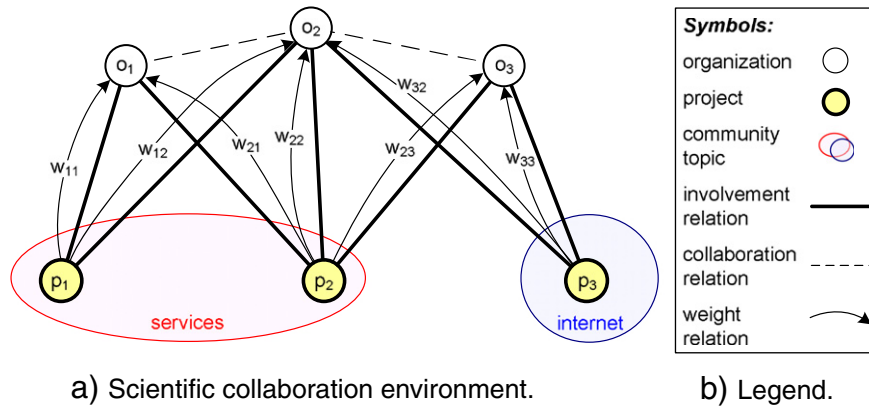


Fig. 1. Scientific collaboration environment and definitions.

roles, funding, project descriptions, date of project contracts, and project duration. Let us define some basic graph structures that are obtained from information in the Project Database and then managed in the Graph Database.

Based on projects, organizations, and involvement relations we define two types of graphs. First, let us define the directed project-organization graph  $G_{PO}(V_P, V_O, E_P)$  that is composed of the set projects  $V_P$  and the set of organizations  $V_O$  ( $V_P$  and  $V_O$  depicting the vertices in

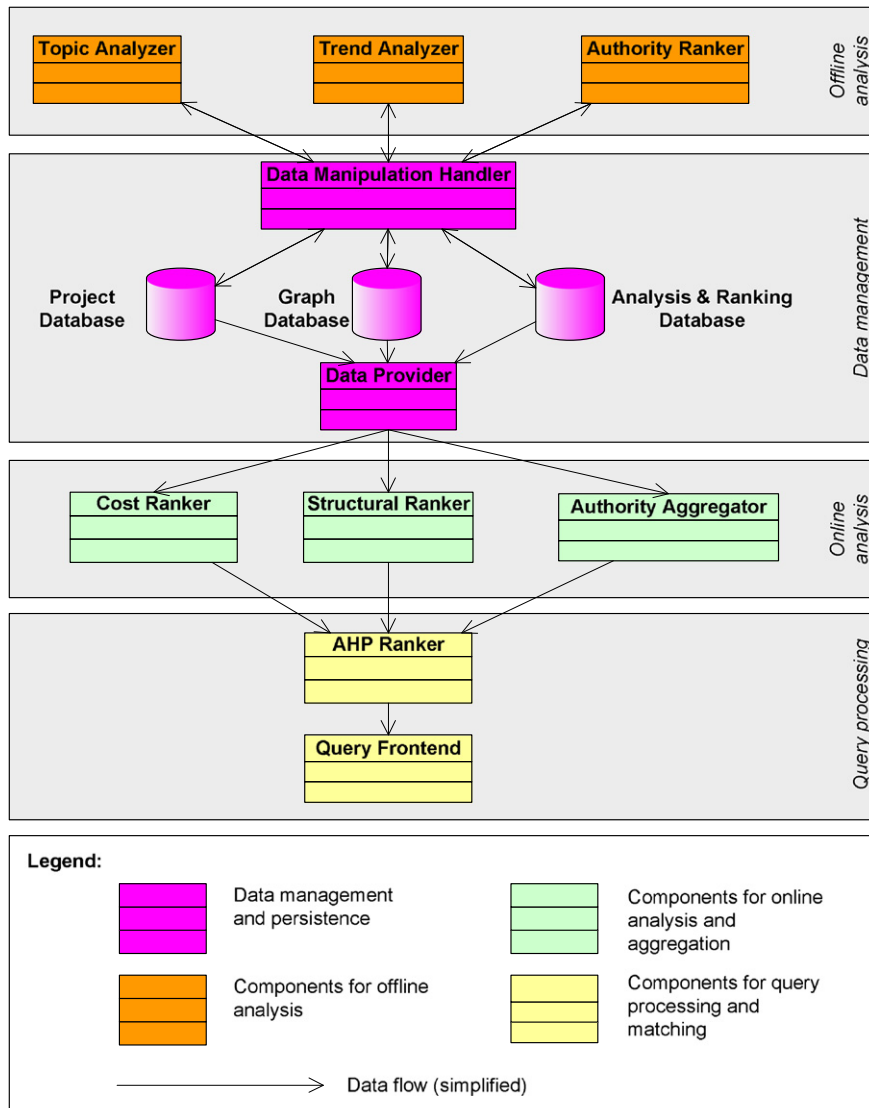


Fig. 2. Solution framework outline.

the graph) and the project involvement relations denoted by the edge set  $E_p$  where an edge  $(p, o) \in E_p$  points from the project  $p$  to organization  $o$ . Each edge  $(p, o) \in E_p$  has a weight  $w_{po}$  associated with it depending on the funding the organization  $o$  receives in project  $p$  divided by the total project funding. This type of graph is being used for *organization authority ranking*. Let us define the second type of graph as the undirected organization-collaboration graph  $G_{OC}(V_O, E_O)$  consisting of the set of organizations  $V_O$  depicting the vertices in the graph and the set of collaboration relations  $E_O$ . Whereas the edges  $E_p$  in  $G_{PO}$  are based on project involvement relations pointing from projects to organizations, the edges  $E_O$  in  $G_{OC}$  are undirected and connect two organizations. This type of graph is being used for *structural importance ranking*. Details regarding these two types of graph structures will be provided in the following.

### 3.2.2. Offline analysis

This layer deals with components that are invoked in an offline manner (e.g., triggered by changes in the Project Database). The Topic Analyzer extracts relevant topics from project descriptions by filtering stop words and combining synonyms to single topics. Essentially, each topic is identified by a single keyword that has a frequency associated with it to identify popularity of topics. Typical topics in the context of ICT research are, for example, 'services' and 'internet'. Sophisticated topic models such as cross-topic relations or hierarchical structures are not within the focus of this work (e.g., see [31] for hierarchical topic models and topic clustering techniques). The Topic Analyzer saves topic information in the Analysis & Ranking Database. If new topics are added, other components such as the Authority Ranker are triggered (see later). The Trend Analyzer calculates trends with regard to organizations' activities in topics. The historical project information is used to calculate a trend in increasing or decreasing number of projects for given topics. In a steady state (neither increasing nor decreasing activity), the trend equals 0. Trend information is utilized by the Authority Ranker to create topic and time-aware authority scores. The detailed mechanisms will be discussed in later sections. The Authority Ranker calculates numeric values for authority scores. To explain the notion of authority as used in this work, participation of an organization in a research project (i.e., involvement relation) is understood as a carrier of authority. By being involved in certain projects, we assume that organizations develop knowledge with regards to the projects' topic(s). An organization is considered to be an authority if it has extensive or specialized knowledge about a topic. In other words, an organization must have collaborated in the context of a topic to be considered as an authority for a given topic.

### 3.2.3. Online analysis

Previously, the offline analysis components were responsible for preparing the information needed to perform online analysis and ranking. Our decision to divide functionality into online and offline analysis was due to computational complexity of link-based authority ranking algorithms. Computation of authority at query time without having performed offline computation would result in unacceptable response times at magnitudes of hours or even longer. All information from the previous steps is made available in the Analysis & Ranking Database. The Cost Ranker is a simple ranker that provides a scoring function based on organizations' average costs. The Structural Ranker calculates numeric values for structural importance scores. The idea of our structural importance metric is drawn from the notion of structural holes as established in a sociological context. To detail the difference between structural importance and authority, the notion of authority captures the importance of an organization with regard to knowledge drawn from past project experience. Structural importance captures a different notion of importance that is based on the lack of information flow and connectedness of parts of the network. As stated by Burt [5], structural holes are an opportunity to broker the flow of information between people and control the projects that bring together people from opposite

sides of the hole. Here the notion of structural holes is not applied to people-based social networks, but to organization collaboration networks (i.e.,  $G_{OC}$ ). The goal of our ranking approach is to identify those organizations that have the ability to bridge structural holes and to allow for the emergence of novel innovative ideas through brokerage of information. The Authority Aggregator combines the authority results of offline computed authority scores.

### 3.2.4. Query processing

Suppose a coordinator attempts to establish a new consortium and thus wants to find collaboration partners who are able to join the consortium. Often, previous collaborators are known from first hand collaboration experience but in today's vibrant and fast-paced research environment it is also useful to see the current community standing of known collaboration partners and to discover potential new collaborators. The coordinator is able to specify a keyword-based query  $Q = \{q_1, q_2, \dots, q_n\}$  (using the Query Frontend) with the goal of finding matching organizations that are ranked according to a set of criteria (i.e., cost, structural importance and authority). The idea of our ranking approach is to compute ranking scores with respect to certain *areas of expertise*. The demanded areas of expertise are specified via the query  $Q$  and matched with topics. Each query keyword  $q_n$  corresponds to a desired area of expertise. A query returns a ranked list of organizations based on the demanded set of expertise areas. The AHP Ranker is used to create a composite ranking score  $S(o; Q)$  of organization  $o$ . The score  $S(o; Q)$  is given as

$$S(o; Q) = AHP(A(o; Q), SI(o; Q), C(o)) \quad (1)$$

where  $A(o; Q)$  is the organization's authority score and  $SI(o; Q)$  the structural importance score with respect to the query  $Q$ , and  $C(o)$  the cost score. The following sections will focus on the calculation of  $S(o; Q)$ .

## 4. Authority model

Here we formalize the notion of organization authority as it will be used in our ranking model. Authority is automatically calculated using network analysis techniques. The novelty of the approach is that authority is put *into context* by considering topic information. Well-established models provide the foundational concepts and basis. Specifically, we base our approach upon the model of hubs and authorities as developed by [21].

### 4.1. Hubs and authorities

Let us apply the notion of hubs and authorities to a collaboration environment as depicted by Fig. 1. A project is regarded to be important if the organizations contributing to it are also regarded to be important (e.g., knowledgeable and reputable). In turn, the importance of an organization is based on its involvement in important projects. This is a recursive definition of importance and can be modeled by using the intuitive notion of hubs and authorities as proposed by [21].

$$A(o) = \sum_{(p,o) \in E_p} H(p) \quad H(p) = \sum_{(p,u) \in E_p} A(u) \quad (2)$$

In the model, an organization  $o$  obtains an authority score depicted by  $A(o)$  and a project  $p$  obtains a hub score denoted by  $H(p)$ . The drawback of this model is the 'stability' of rankings. A ranking algorithm is stable if the algorithm returns similar results upon small disturbances. We follow the *randomized HITS* approach as proposed in [27] and expand the equations in Eq. (2) as follows:

$$A(o) = (1 - \lambda_a) \delta_o(o) + \lambda_a \sum_{(p,o) \in E_p} H(p) \quad (3)$$



$$H(p) = (1 - \lambda_h) \delta_p(p) + \lambda_h \sum_{(p,u) \in E_p} A(u). \quad (4)$$

This adjusted model is a natural way of designing a random-walk based algorithm following the HITS model. The randomized HITS approach is, like PageRank, stable to small perturbations [27]. The symbols  $\delta_o(o)$  and  $\delta_p(p)$  depict personalization vectors that may be assigned uniformly for each node such that  $\delta_o(o) = \frac{1}{|V_o|}$  and  $\delta_p(p) = \frac{1}{|V_p|}$ . Non-uniform personalization vectors result in personalized rankings. The parameters  $\lambda_a$  and  $\lambda_h$  with  $0 \leq \lambda \leq 1$  allow for balancing between authority/hub weights and personalization weights. A typical value for  $\lambda$  is 0.85 [28]. Assigning lower values to  $\lambda$  means that higher importance is given to the personalization weights; thereby reducing the ‘network effect’ of the ranking algorithm.

#### 4.2. Query-sensitive personalization

Let us define the query-sensitive authority score:

$$A(o; Q) = (1 - \lambda_a) \delta_o(o; Q) + \lambda_a \sum_{(p,o) \in E_p} w_{po} H(p; Q) \quad (5)$$

Similarly, let us define the query-sensitive hub score:

$$H(p; Q) = (1 - \lambda_h) \delta_p(p; Q) + \lambda_h \sum_{(p,u) \in E_p} w_{pu} A(u; Q) \quad (6)$$

The edge weights  $w_{po}$  and  $w_{pu}$  are based on the organizations' degree of project involvement. Particularly, the weight  $w_{po}$  is based on the funding received by organization  $o$  in project  $p$  and is calculated as

$$w_{po} = \frac{\text{funding}(p, o)}{\sum_{v \in \text{adj}(p)} \text{funding}(p, v)} \quad (7)$$

where  $\text{adj}(p)$  depicts the set of nodes adjacent to  $p$  (i.e., the set of organizations involved in project  $p$ ). To compute authority scores using a single equation, which is the desired goal of our approach, we substitute  $H(p; Q)$  in Eq. (5) by Eq. (6) and have:

$$A(o; Q) = (1 - \lambda_a) \delta_o(o; Q) + \lambda_a (1 - \lambda_h) \sum_{(p,o) \in E_p} w_{po} \delta_p(p; Q) + \lambda_a \lambda_h \sum_{(p,o) \in E_p} \sum_{(p,u) \in E} w_{po} w_{pu} A(u; Q). \quad (8)$$

Based on Eq. (8), let us define the personalization vector  $\delta'_o(o; Q)$  as follows:

$$\delta'_o(o; Q) = \frac{1 - \lambda_a}{1 - \lambda_h} \delta_o(o; Q) + \lambda_a \sum_{(p,o) \in E_p} w_{po} \delta_p(p; Q). \quad (9)$$

If we use the same parameter values for  $\lambda_a$  and  $\lambda_h$  (due to symmetry of Eqs. (5) and (6)) such that  $\lambda_a = \lambda_h$ , Eq. (9) simplifies to:

$$\delta'_o(o; Q) = \delta_o(o; Q) + \lambda \sum_{(p,o) \in E_p} w_{po} \delta_p(p; Q). \quad (10)$$

In the following step we rewrite Eq. (8) by using the personalization vector  $p'(u; Q)$  as defined in Eq. (10).

$$A(o; Q) = (1 - \lambda) \delta'_o(o; Q) + \lambda^2 \sum_{(p,o) \in E_p} \sum_{(p,u) \in E_p} w_{po} w_{pu} A(u; Q) \quad (11)$$

As one can see, Eq. (11) has a PageRank-like structure. An important concept for personalization based on the PageRank model is the *linearity theorem* as introduced in [19]. The theorem states that for any

personalization vectors  $\delta_1, \delta_2$  and weights  $w_1, w_2$  with  $w_1 + w_2 = 1$ , the following equality holds:

$$PPV(w_1 \delta_1 + w_2 \delta_2) = w_1 PPV(\delta_1) + w_2 PPV(\delta_2). \quad (12)$$

$$\begin{aligned} A(o; Q) &= (1 - \lambda) \delta'_p(o; Q) + \lambda^2 \sum_{(p,o) \in E_p} \sum_{(p,u) \in E_p} w_{po} w_{pu} A(u; Q) \\ &= (1 - \lambda) \sum_{q \in Q} w_q \delta'_p(o; q) + \lambda^2 \sum_{(p,o) \in E_p} \sum_{(p,u) \in E_p} \sum_{q \in Q} w_q w_{po} w_{pu} A(u; q) \\ &= \sum_{q \in Q} w_q (1 - \lambda) \delta'_p(o; q) + \sum_{q \in Q} w_q \lambda^2 \sum_{(p,o) \in E_p} \sum_{(p,u) \in E_p} w_{po} w_{pu} A(u; q) \\ &= \sum_{q \in Q} w_q \left[ (1 - \lambda) \delta'_p(o; q) + \lambda^2 \sum_{(p,o) \in E_p} \sum_{(p,u) \in E_p} w_{po} w_{pu} A(u; q) \right] \\ &= \sum_{q \in Q} w_q [A(o; q)]. \end{aligned} \quad (13)$$

The linearity theorem states that personalized PageRank vectors  $PPV$  can be composed as the weighted sum of PageRank vectors. Eq. (13) shows how to derive the weighted sum of personalized authority ranking scores using Eq. (11). The goal is to obtain a structure as depicted by the right part of Eq. (12). The weight  $w_q$  is associated with a particular keyword  $q$  with  $w_q = \frac{1}{|Q|}$  for uniform weights and  $\sum_q w_q = 1$ .

As stated before, the benefit of the model is the ability to precompute authority scores for particular topics, save them in a database, and aggregate the precomputed authority scores later at query time. Suppose the set of topics, as extracted by the Topic Analyzer, is given as  $T = \{T_1, T_2, \dots, T_n\}$ . For each topic authority scores are calculated  $A(o; T_1), A(o; T_2), \dots, A(o; T_n)$  and utilized by the Authority Aggregator to compute

$$A(o; Q) = \sum_{q \in Q} w_q A(o; T_q) \quad (14)$$

where  $T_q$  is the topic matching query keyword  $q$ . Next, we describe the time-aware authority model.

#### 4.3. Personalization weights and time-aware authority

We have extensively discussed the notion of authority and the idea of computing authority scores for individual topics that can be aggregated at query time. Now we turn to the definition of the personalization vectors  $\delta_p$  and  $\delta_o$ . Recall,  $\delta_p$  holds personalization weights for projects and  $\delta_o$  holds personalization weights for organizations. For  $\delta_p$  we use a straightforward model to calculate personalization weights

$$\delta_p(p) = \frac{\text{funding}(p)}{\sum_{proj \in V_p} \text{funding}(proj)} \quad (15)$$

where  $\text{funding}(p)$  depicts the monetary funding received by project  $p$ . For simplicity, we do not consider the query context  $Q$  for the project-based personalization vector.

The next discussion is related to the concept of time-aware and topic-based authority ranking. Thus, we establish metrics to calculate the personalization weights of  $\delta_o$ . Here topic-based personalization and time-aware weighting is applied. Recall that a topic is identified by a single keyword. Organizations typically perform numerous projects that are related to one or more topic(s). Thus, each organization has a set of topics including topic frequency associated with it. Furthermore, frequencies of topics are counted by year. An example for such data would be (“OrgA”, 2011, “services”, 5) where “OrgA” is the organization name, 2011 the specific year, “services” the given topic and the number

5 an example of a frequency count. As a first step let us define the weight function  $W^T(o, y; T_x)$  that obtains the frequency count of organization  $o$  in year  $y$  for some topic  $T_x$ . The frequency count is based on how many projects related to the given topic the organization has started in the year (i.e., the year when signing the project contract). To establish the notion of positive or negative change in topic specific weights, we define the weight deviation function  $W_{\Delta}^T(o, y; T_x)$  as follows:

$$W_{\Delta}^T(o, y; T_x) = W^T(o, y; T_x) - \frac{1}{|Y|} \sum_{y' \in Y} W^T(o, y'; T_x) \quad (16)$$

Deviation in this context means the weight  $W^T(o, y; T_x)$  in year  $y$  minus the average weight with regard to topic  $T_x$ . Straightforwardly, a positive sign means increasing topic-based weight, a negative sign means decreasing topic-based weight as a result of being below the average, and 0 means no change in topic-based weights (i.e., through constant rate of projects related to topic  $T_x$ ). This definition is quite simple and captures already a notion of ‘trend’ by analyzing the temporal project history of an organization. The positive/negative sign depicts increasing or decreasing trend. However,  $W_{\Delta}^T(o, y; T_x)$  just analyzes the trend with respect to organization  $o$  without considering the weights and thus performance of other organizations. Personalization for authority ranking in collaboration networks must be performed by considering weights in relation to all other organizations. For brevity, let us define the set  $\alpha = \{W^T(o_1, y; T_x), W^T(o_2, y; T_x), \dots, W^T(o_n, y; T_x)\}$  with  $\{o_1, o_2, \dots, o_n\} \in V_O$ . Let us define the trend  $Tr(o; T_x)$  of organization  $o$  with respect to topic  $T_x$  as:

$$Tr(o; T_x) = \sum_{y \in Y} w_y \left[ \frac{W^T(o, y; T_x)}{\max(\alpha)} \times W_{\Delta}^T(o, y; T_x) \right]. \quad (17)$$

$Tr(o; T_x)$  is based on the trend for topic  $T_x$  over the years  $Y = \{y_1, y_2, \dots, y_n\}$  where  $y_n$  is the most recent year,  $y_{n-1}$  the previous year and so forth (ordered by recency). The first term within the square brackets measures the topic based weight in relation to the community performance in  $T_x$  by dividing  $W^T(o, y; T_x)$  by  $\max(\alpha)$ . For the top-performing organizations having the most numbers of projects related to  $T_x$  in year  $y$  the term becomes 1. The term is multiplied by the organization specific weight deviation function  $W_{\Delta}^T(o, y; T_x)$ . The weight  $w_y$  puts more emphasis on recent years (recency factor) by being calculated as  $w_y \in \left\{ \frac{1}{|Y|}, \frac{1}{|Y|-1}, \frac{1}{|Y|-2}, \dots, 1 \right\}$ .

Finally, the personalization vector  $\delta_o$  needs to be assigned by matching organizations having performed projects related to  $T_x$  and trend values  $Tr$  need to be mapped to a positive interval. This is done because  $\delta_o$  represents a probability distribution (for theoretical foundations related to personalized PageRank see, for example, [18]). Let us define the set  $\beta = \{Tr(o_1; T_x), Tr(o_2; T_x), \dots, Tr(o_n; T_x)\}$  with  $\{o_1, o_2, \dots, o_n\} \in V_O$ .

$$\delta_o(o; T_x) = \begin{cases} 1 - \frac{\max(\beta) - Tr(o; T_x)}{\max(\beta) - \min(\beta)} & , \text{if matches}(o; T_x) \\ 0 & , \text{otherwise} \end{cases} \quad (18)$$

The function  $\text{matches}(o; T_x)$  checks if  $o$  has performed projects related to  $T_x$  and evaluates to true or false. To evaluate a query  $Q = \{T_1, T_2\}$  a simple aggregation is performed

$$A(o; \{T_1, T_2\}) = w_1 A(o; T_1) + w_2 A(o; T_2) \quad (19)$$

where  $A(o; T_1)$  is personalized for  $T_1$  and  $A(o; T_2)$  is personalized for  $T_2$ . In other words, both  $A(o; T_1)$  and  $A(o; T_2)$  hold topic-based and time-aware authority scores for all  $o \in V_O$ .

## 5. Structural importance model

The previous section explained in detail the authority model and ranking approach. Here we turn to the second criteria used in our overall ranking model. We define the notion of structural importance and detail a metric to calculate the importance. The obtained ranking scores for structural importance are used as a second parameter in the AHP-based aggregation (i.e., the AHP parameter  $SI(o; Q)$ ). By following the notion of structural holes as coined by Burt [5,6], structural holes are an opportunity to broker the flow of information between people in an organizational or social network. As an example, managers often act as information brokers as they talk to many people in the project.

Structural importance captures the ability of a network node to broker information between its neighbors (in our context organizations). A node can do so if potential ‘information gaps’ (or buffers) arise in the network. A broker can also be seen as a mediator that helps establishing communication between other nodes. A project partner with ‘brokerage’ capabilities is often important in project consortia to help establish and facilitate communication among other partners. As an example, a project consortium may be led by an academic partner who is in charge of coordinating the project from an administrative and scientific point of view. Typically, exploitation and further use of project results is an important issue in research projects. However, the consortium leader may not be the optimal partner for transferring (or ‘translating’) scientific results to business. Thus, there may be a gap between technical/scientific results and exploitation of results within an industrial context (e.g., implementing novel solutions within an industrial environment). With regards to this example, an organization may act as a broker by mediating communication and transferring the knowledge to an industrial partner within the project.

Thus, structural importance essentially focuses on mediation capabilities of an organization as opposed to expertise/authority. Such mediators help running projects more effectively and efficiently by (a) establishing communication between potentially disconnected network segments that have not communicated before and (b) help making communication more fluid and efficient. To be able to act as a broker, gaps must exist in the network because otherwise a node loses its ability to establish communication. The notion of *redundancy* provides means to express the existence of such gaps. If there is high redundancy in terms of network edges and communication paths in a network, the need to fill structural gaps may be very limited. On the contrary, if a network is highly segmented and only few nodes connect individual segments, the need for brokers and mediation opportunities may be very high.

Let us consider a graph as depicted by Fig. 3. Here the graph model  $G_{OC}$  is used that consists of organizations and collaboration relations as undirected edges. Each node depicts an organization with  $\{a, b, c, o, r, u, v, z\}$ .

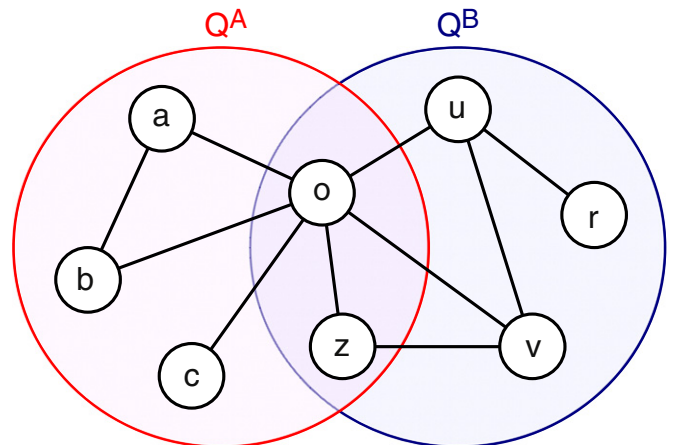


Fig. 3. Network structure to illustrate structural importance metric.

$u, v, z\} \subset V_o$ . A circle surrounds nodes that belong to a particular expertise area or community identified through  $A$  and  $B$ . A query may be formulated to match the nodes and edges in either  $Q^A$  or  $Q^B$  or both  $Q = Q^A \cup Q^B = \{T_A, T_B\}$ . An edge  $(v, u) \in G_{OC}$  has a weight which is based on the number of performed projects between  $v$  and  $u$ . The weight is dynamically assigned depending on the query context  $Q$ . For example, the weight of the edge  $(o, z) \in G_{OC}$  may be different in  $Q^A$  and  $Q^B$  depending on the joint projects performed by  $o$  and  $z$  (i.e., if the projects match  $Q^A$  or  $Q^B$  or both). Suppose  $Q = Q^A \cup Q^B$ , the node tito has the highest number of non-redundant edges in the graph because it connects the node sets  $\{a, b, c\}$  and  $\{u, v, z\}$  which are only reachable via  $o$ . Thus,  $o$  has a unique position within the network because  $o$  is able to control the information flow between both node sets. Furthermore, only  $o$  and  $z$  belong to both expertise areas  $A$  and  $B$  but only  $o$  is connected to  $\{a, b, c\}$  in  $A$ . Let us define  $SI(o; Q)$  as

$$SI(o; Q) = \sum_{u \in N(o)} \left[ 1 - \sum_{v \in N(u)} W_N^Q(o, v; Q) W_M^Q(u, v; Q) \right] \quad (20)$$

where  $v \notin \{u, o\}$  and  $N(o)$  the set of  $o$ 's neighbors. For  $SI(o; Q)$ , we follow Burt's measure of the effective size of a node's network [6]. Here the notion of structural holes is not applied to people-based social networks, but to organization collaboration networks (i.e.,  $G_{OC}$ ). Conceptually, the effective size is the number of nodes  $o$  is connected to, minus the redundancy in the network.

In contrast to Burt's definition of effective size, we compute structural importance with respect to the query  $Q$ . As an example, while  $o$  in Fig. 3 is structurally important in  $Q = Q^A \cup Q^B$  to establish a flow between  $\{a, b, c\}$  and  $\{u, v, z\}$ ,  $o$  is less significant if only  $Q^B$  is considered. Actually, within  $Q^B$   $u$  has a unique position because  $r$  is only reachable via  $u$ .

The weight  $W_N^Q(o, v; Q)$  in Eq. (20) depicts the query-sensitive normalized edge weight between  $o$  and  $v$  and is calculated as

$$W_N^Q(o, v; Q) = \sum_{q \in Q} \frac{w_{ov}^q}{\sum_{u \in N(o)} w_{ou}^q} \quad (21)$$

where  $w_{ov}^q$  is the weight associated with  $(o, v) \in E_o$  and calculated as the number of joint projects between  $o$  and  $v$  matching the query keyword  $q$ . Furthermore, the weight  $W_M^Q(u, v; Q)$  in Eq. (20) depicts the query-sensitive marginal edge weight between  $u$  and  $v$  and is calculated as follows:

$$W_M^Q(u, v; Q) = \sum_{q \in Q} \frac{w_{uv}^q}{\max(\{w_{un}^q | \forall n \in N(u)\})} \quad (22)$$

The marginal weight of  $u$  with neighbor  $v$  is the weight  $w_{uv}^q$  (also based on the number of matching joint projects between them) divided by  $u$ 's strongest weight with anyone of its neighbors  $N(u)$ . If none of the projects match  $q$ , the weight  $w_{uv}^q = 0$ .

## 6. Multi-criteria ranking algorithm

Here we discuss the computation of the final ranking score. Recall that the composite ranking score  $S(o; Q)$  of organization  $o$  is obtained through  $AHP(A(o; Q), SI(o; Q), C(o))$ . Previously we have defined the authority  $A(o; Q)$  and the structural importance  $SI(o; Q)$ . Cost  $C(o)$  is calculated as the average funding organization  $o$  receives:

$$C(o) = \frac{1}{\text{num.projects}(o)} \sum_{(p, o) \in E_p} \text{funding}(p, o) \quad (23)$$

The final aggregation and computation of a composite ranking score is done using the AHP algorithm. AHP is a technique for making complex decisions in a structured way. AHP has been successfully applied in a number of fields including transportation [12], maintenance and

configurations [8], and service quality assessment [12]. The theoretical background will not be covered in this work since AHP is a well explored technique. We refer the reader to [29] for details regarding AHP as a decision making technique.

Algorithm 1 shows the main steps at a high level. The input of the algorithm is given as the query  $Q$  and the organization-collaboration graph  $G_{OC}$ . The graph  $G_{OC}$  is used to compute the structural importance scores in an online manner. Next four essential steps are performed: (1) create map with criteria input scores, (2) set up AHP, (3) perform AHP ranking, and (4) assign final AHP ranking scores to output map.

### Algorithm 1. Multi-criteria ranking algorithm

*Input:* The query  $Q$  and the undirected organization-collaboration graph  $G_{OC}$ .

*Compute:*

1. Create map for org with individual scores. For each organization  $o \in V_o$  do:
  - $A(o) \leftarrow au\_score(o, Q)$  //authority
  - $SI(o) \leftarrow si\_score(o, G_{OC}, Q)$  //struct. imp.
  - $C(o) \leftarrow avg\_cost(o)$  //cost
  - Add to map  $(o, \{A(o), SI(o), C(o)\})$
2. Setup AHP attributes weights and desirability.
  - Auth. attributes ("authority",  $\{w_{au}, +1\}$ )
  - Struct. attributes ("structure",  $\{w_{si}, +1\}$ )
  - Cost attributes ("cost",  $\{w_{cost}, -1\}$ )
3. Perform AHP ranking using output from previous steps.
  - Compute the vector of criteria weights.
  - Compute the matrix of organization scores.
  - Rank the organizations.
4. Assign final AHP ranking scores to map  $S$ . For each organization  $o \in V_o$  do:
  - $S(o) \leftarrow ahp\_score(o)$  //final score

*Output:* Ranked organizations based on query  $Q$  and according to composite AHP ranking score.

First, the ranking criteria scores are obtained as described in the previous sections (authority Section 4 and structural importance Section 5 respectively). These include authority, structural importance and cost. Using a map, each criteria score is associated with an organization. The map generated in this step is passed as an argument to the AHP ranking in step 3.

Second, AHP attributes are setup by assigning the weights  $w_{au}$ ,  $w_{si}$ ,  $w_{cost}$  to each criteria with  $[\sum_w w] = 1$ . In addition, the desirability attribute is assigned to denote if a certain criterion is desired or not. In particular, authority and structural importance should be high (desirability = +1) to obtain a better AHP ranking score whereas cost should be low to obtain a better ranking score (desirability = -1).

Third, AHP ranking is performed by using the previously setup attributes and the output map of step 1. The step 3 of Algorithm 1 is decomposed into the following steps:

- Compute the vector of criteria weights: In this step rating of the relative priority of the criteria is done by assigning a weight value to the more important criteria. The weight values are taken from the previous step of the algorithm (step 2). The weight assignment is done through a pairwise comparison of the criteria. After that, the resulting weights are normalized and the average is computed for each criterion.
- Compute the matrix of organizations scores: Here the score for each organization is determined by computing how well organization  $o$  meets some criterion  $Y$ . Afterwards, the organizations' scores are normalized and averaged.
- Rank the organizations: In a final step the organizations' scores are combined with the criterion weights to produce an overall score for each organization. The extent to which the organizations satisfy the

**Table 1**  
Popular project topics and frequencies.

| Topic      | Frequency |
|------------|-----------|
| Systems    | 4126      |
| Internet   | 2729      |
| Networks   | 1771      |
| Services   | 1247      |
| Software   | 1224      |
| Health     | 1115      |
| Embedded   | 1054      |
| Transport  | 890       |
| Efficiency | 849       |
| Energy     | 849       |

criteria is weighted according to the relative importance of the criteria. The final score is simply computed as a weighted sum.

Note, in our case criteria are contrasting by demanding that organizations should have high authority but low cost. In general, the organization that is recommended for selection (top-ranked in final output  $S$ ) is not necessarily the one which optimizes each single criterion, but rather the organization which achieves the most suitable *trade-off among the different criteria*. This behavior makes AHP a very flexible and powerful tool for multi-criteria partner selection.

Fourth, the AHP scores are saved in a final score map  $S$ . Organizations are ranked in descending order by ranking score.

## 7. Evaluation

Here the evaluation of the proposed concepts and model is presented. We have selected a dataset of a scientific collaboration environment to test the concepts.

### 7.1. Description of dataset

The data is based on ICT research projects having received grants under the EU's Seventh Framework Programme (FP7). The data as described in detail in [25] and covers a period from 2007 to 2011. Research projects have multiple partners and an organization can be the partner of multiple projects. To date, the FP7 ICT program has allocated funding to 1469 projects for a total Union funding of 4,979,301,152 Euro. This results in 14,781 participations by 4718 distinct legal entities.

**Table 2**  
Top-20 organizations ranked by degree.

| PNr | Name  | Cost      | Degree | Structural rank score |
|-----|---|-----------|--------|-----------------------|
| 1   | Fraunhofer-Gesellschaft zur Foerderung der Angewandten Forschung E.V. | 524,515   | 272    | 516                   |
| 2   | Centre National de la Recherche Scientifique                          | 159,983   | 153    | 175                   |
| 3   | Commissariat à l'Énergie Atomique et Aux Energies Alternatives        | 235,911   | 137    | 201                   |
| 4   | Ecole Polytechnique Federale de Lausanne                              | 290,827   | 97     | 146                   |
| 5   | Consiglio Nazionale delle Ricerche                                    | 455,650   | 96     | 154                   |
| 6   | Valtion Teknillinen Tutkimuskeskus                                    | 293,240   | 95     | 190                   |
| 7   | Institut National de Recherche en Informatique et en Automatique      | 799,995   | 94     | 127                   |
| 8   | Interuniversitair Micro-Electronica Centrum Vzw                       | 964,195   | 90     | 140                   |
| 9   | Eidgenoessische Technische Hochschule Zurich                          | 389,544   | 90     | 89                    |
| 10  | Telefonica Investigacion Y Desarrollo S.A.                            | 636,818   | 76     | 131                   |
| 11  | Katholieke Universiteit Leuven  | 711,085   | 69     | 98                    |
| 12  | SAP AG  | 1,221,665 | 68     | 168                   |
| 13  | Universidad Politecnica de Madrid                                     | 331,315   | 65     | 136                   |
| 14  | Atos Origin Sociedad Anonima Espanola                                 | 628,296   | 62     | 215                   |
| 15  | Imperial College of Science, Technology and Medicine                  | 286,930   | 61     | 56                    |
| 16  | Politecnico di Milano   | 531,975   | 61     | 98                    |
| 17  | Kungliga Tekniska Hoegskolan  | 415,684   | 59     | 85                    |
| 18  | Technische Universiteit Delft   | 622,286   | 58     | 88                    |
| 19  | Karlsruher Institut Fuer Technologie                                  | 250,599   | 56     | 76                    |
| 20  | Technische Universitaet Wien  | 287,363   | 55     | 87                    |

Our evaluation is performed as follows. First we select the top-20 organizations (ranked by degree and given in Table 2) and compute metrics for those 20 organizations with regard to popular topics. This evaluation is called top-k rank evaluation and is presented in Section 7.2. Second we compute cross topic ranking statistics such as overlap similarity and Kendall's  $\tau$  rank difference. This evaluation is presented in Section 7.3 along with the definition of relevant ranking metrics.

Table 1 gives an overview of popular (project) topics extracted from project information (see [25] for details). Frequency is measured by counting appearance of the topic string within project names and project short descriptions of each project partner involvement record (association of organization to project including received funding). In total, we extracted 170 topics after performing some automatic and manual processing of the data. Table 1 shows the top-10 topics with the highest frequencies among the 170 topics.

### 7.2. Top-k rank evaluation

Table 2 shows the top-20 organizations ranked by their degree in  $G_{PO}$  (project-organization graph). The first column (PNr column) is a unique key associated with an organization and used throughout this section to identify a top-20 organization. The second column (Name column) depicts the organizations' legal name. The third column (Cost column) shows the average organization cost using Eq. (23). The organization in degree (Degree column) is analog to the project count as projects  $p \in V_P$  point to organizations  $o \in V_O$ . The degree-based rank will be used as a baseline ranking. This baseline results will be compared with AHP-based rankings. We have selected the degree-based rank as a baseline algorithm to show the impact of personalization based on topic information and time-aware authority ranking. Notice, the degree-based rank has no topic bias. In addition, the degree-based rank was selected because it already captures some notion of importance with regard to organization reputation. The last column (Structural Rank Score column) shows the structural rank score (using Eq. (20)) over all topics in Table 1. A higher score is better.

It is noticeable that the organization 14 has a particular high structural rank score in relation to its degree-based rank position. Organization 1 has an exceptionally high structural rank score but has also the most projects within the ICT framework program. Notice, however, the degree is calculated using  $G_{PO}$  and the structural rank using  $G_{OC}$ .

To compare AHP results with the rankings in Tables 2, 3 and 4 list detailed metrics for selected topics. We have selected eight out of the ten topics from Table 1 due to space reasons.



**Table 3**  
Top-20 list of organizations: topics include 'networks', 'systems', 'software', and 'services'.

| PNr | Networks |      |            |      |       | Systems |      |           |      |       | Software |      |            |      |       | Services |      |            |      |       |
|-----|----------|------|------------|------|-------|---------|------|-----------|------|-------|----------|------|------------|------|-------|----------|------|------------|------|-------|
|     | D        | A    | P          | Ch   | Tr    | D       | A    | P         | Ch   | Tr    | D        | A    | P          | Ch   | Tr    | D        | A    | P          | Ch   | Tr    |
| 1   | 29       | 0.11 | <b>1</b>   | 2    | 1.08  | 70      | 0.86 | <b>1</b>  | 0    | 7.97  | 7        | 0.15 | <b>2</b>   | 3    | 0.23  | 11       | 0.04 | <b>5</b>   | 28   | -0.03 |
| 2   | 10       | 0.03 | <b>31</b>  | 677  | -0.13 | 39      | 0.00 | <b>16</b> | 196  | -0.03 | 3        | 0.04 | <b>12</b>  | 27   | 0.00  | 3        | 0.04 | <b>15</b>  | 35   | 0.00  |
| 3   | 20       | 0.05 | <b>12</b>  | 25   | 0.19  | 44      | 0.10 | <b>5</b>  | 4    | 0.79  | 1        | 0.04 | <b>7</b>   | 20   | 0.00  | 1        | 0.04 | <b>10</b>  | 28   | 0.00  |
| 4   | 5        | 0.03 | <b>30</b>  | 665  | -0.05 | 26      | 0.01 | <b>19</b> | 87   | 0.01  | 1        | 0.04 | <b>16</b>  | 27   | 0.00  | 2        | 0.04 | <b>17</b>  | 34   | 0.00  |
| 5   | 4        | 0.05 | <b>16</b>  | 27   | 0.17  | 19      | 0.15 | <b>3</b>  | -1   | 1.44  | 6        | 0.06 | <b>10</b>  | 6    | 0.04  | 7        | 0.04 | <b>11</b>  | 13   | 0.06  |
| 6   | 12       | 0.07 | <b>6</b>   | 5    | 0.56  | 20      | 0.05 | <b>8</b>  | 15   | 0.39  | 1        | 0.04 | <b>9</b>   | 19   | 0.00  | 7        | 0.02 | <b>25</b>  | 630  | -0.18 |
| 7   | 16       | 0.09 | <b>5</b>   | -2   | 0.85  | 25      | 0.08 | <b>10</b> | 5    | 0.73  | 7        | 0.04 | <b>18</b>  | 44   | -0.01 | 9        | 0.04 | <b>22</b>  | 618  | -0.02 |
| 8   | 6        | 0.04 | <b>27</b>  | 61   | 0.01  | 27      | 0.14 | <b>6</b>  | -3   | 1.21  | 0        | 0.01 | <b>42</b>  | 357  | 0.00  | 1        | 0.04 | <b>19</b>  | 25   | 0.00  |
| 9   | 7        | 0.04 | <b>32</b>  | 31   | 0.10  | 33      | 0.09 | <b>12</b> | -1   | 0.87  | 1        | 0.04 | <b>29</b>  | 18   | 0.00  | 1        | 0.04 | <b>34</b>  | 27   | 0.00  |
| 10  | 38       | 0.00 | <b>681</b> | 671  | -0.55 | 6       | 0.01 | <b>26</b> | 67   | 0.06  | 11       | 0.01 | <b>37</b>  | 353  | -0.07 | 12       | 0.04 | <b>18</b>  | 16   | 0.03  |
| 11  | 3        | 0.03 | <b>57</b>  | 661  | -0.05 | 23      | 0.05 | <b>18</b> | 8    | 0.48  | 2        | 0.04 | <b>25</b>  | 25   | 0.00  | 2        | 0.04 | <b>32</b>  | 32   | 0.00  |
| 12  | 7        | 0.07 | <b>7</b>   | 2    | 0.51  | 6       | 0.01 | <b>15</b> | 56   | 0.06  | 17       | 0.22 | <b>3</b>   | -10  | 0.43  | 18       | 0.17 | <b>2</b>   | -10  | 1.25  |
| 13  | 6        | 0.06 | <b>13</b>  | 6    | 0.38  | 12      | 0.01 | <b>22</b> | 67   | 0.06  | 10       | 0.03 | <b>20</b>  | 349  | -0.03 | 11       | 0.00 | <b>636</b> | 625  | -0.34 |
| 14  | 9        | 0.10 | <b>3</b>   | -10  | 1.06  | 5       | 0.02 | <b>9</b>  | 38   | 0.12  | 7        | 0.05 | <b>5</b>   | 1    | 0.03  | 11       | 0.11 | <b>3</b>   | -9   | 0.72  |
| 15  | 0        | 0.00 | <b>689</b> | 667  | 0.00  | 19      | 0.04 | <b>42</b> | 13   | 0.38  | 1        | 0.04 | <b>57</b>  | 28   | 0.00  | 2        | 0.04 | <b>59</b>  | 31   | 0.00  |
| 16  | 3        | 0.04 | <b>44</b>  | 88   | 0.00  | 22      | 0.07 | <b>14</b> | -3   | 0.66  | 5        | 0.05 | <b>19</b>  | -2   | 0.03  | 5        | 0.04 | <b>30</b>  | 16   | 0.02  |
| 17  | 9        | 0.04 | <b>42</b>  | 48   | 0.04  | 19      | 0.09 | <b>13</b> | -8   | 0.83  | 1        | 0.04 | <b>34</b>  | 21   | 0.00  | 1        | 0.04 | <b>37</b>  | 25   | 0.00  |
| 18  | 3        | 0.03 | <b>60</b>  | 642  | -0.03 | 25      | 0.15 | <b>7</b>  | -15  | 1.48  | 0        | 0.00 | <b>101</b> | 348  | 0.00  | 2        | 0.04 | <b>36</b>  | 26   | 0.00  |
| 19  | 2        | 0.04 | <b>65</b>  | 84   | 0.00  | 17      | 0.02 | <b>49</b> | 26   | 0.18  | 1        | 0.04 | <b>45</b>  | 33   | 0.00  | 1        | 0.04 | <b>47</b>  | 33   | 0.00  |
| 20  | 3        | 0.04 | <b>52</b>  | 96   | 0.00  | 17      | 0.02 | <b>40</b> | 30   | 0.15  | 3        | 0.04 | <b>32</b>  | 6    | 0.01  | 5        | 0.06 | <b>13</b>  | -9   | 0.20  |
| D   | 1.0      | 0.2  | 0.3        | 0.0  | 0.1   | 1.0     | 0.8  | -0.4      | -0.1 | 0.8   | 1.0      | 0.6  | -0.4       | 0.1  | 0.6   | 1.0      | 0.5  | 0.2        | 0.3  | 0.5   |
| A   |          | 1.0  | -0.6       | -0.7 | 1.0   |         | 1.0  | -0.4      | -0.3 | 1.0   |          | 1.0  | -0.5       | -0.5 | 1.0   |          | 1.0  | -0.4       | -0.4 | 1.0   |
| P   |          |      | 1.0        | 0.6  | -0.5  |         |      | 1.0       | 0.2  | -0.4  |          |      | 1.0        | 0.5  | -0.4  |          |      | 1.0        | 0.5  | -0.3  |
| Ch  |          |      |            | 1.0  | -0.6  |         |      |           | 1.0  | -0.3  |          |      |            | -0.3 | 0.0   |          |      |            | 1.0  | -0.4  |

Each metric is computed for each topic in Tables 3 and 4 respectively. Using  $G_{PO}$ , the degree  $D$  is based on matching projects only. Projects are matched against the given topic as depicted in the headings of Tables 3 and 4. The authority  $A$  is calculated for respective topics. The position  $P$  is the rank position index as obtained by the AHP rank using Eq. (1).

AHP is setup with the weights 0.4 for authority, 0.2 for the structural importance rank and 0.4 for cost. Thus, authority and cost are given slightly higher weights than structural importance. We regard authority as highly desirable but at the same time cost should be kept at an acceptable level. After that structural importance is also a desirable

property but not equally important as the other criteria. However, since our approach is flexible weights can be adjusted as demanded.

The position change  $Ch$  is computed between degree-based ranking positions and authority based ranking positions in the following manner

$$Ch(o) = \text{pos}(A(o; T_x)) - \text{pos}(\text{degree\_rank}(o)) \tag{24}$$

where  $\text{pos}()$  retrieves the position index by ranking score. This lets us show how rankings are influenced by authority. Finally, the trend  $Tr$  is

**Table 4**  
Top-20 list of organizations: topics include 'health', 'embedded', 'internet', and 'energy'.

| PNr | Health |      |            |      |       | Embedded |      |           |      |       | Internet |      |             |      |       | Energy |      |            |      |      |
|-----|--------|------|------------|------|-------|----------|------|-----------|------|-------|----------|------|-------------|------|-------|--------|------|------------|------|------|
|     | D      | A    | P          | Ch   | Tr    | D        | A    | P         | Ch   | Tr    | D        | A    | P           | Ch   | Tr    | D      | A    | P          | Ch   | Tr   |
| 1   | 12     | 0.05 | <b>3</b>   | 22   | 0.07  | 11       | 0.12 | <b>2</b>  | 10   | 0.51  | 47       | 0.04 | <b>3</b>    | 11   | 0.47  | 17     | 0.43 | <b>1</b>   | 0    | 0.56 |
| 2   | 7      | 0.06 | <b>10</b>  | 9    | 0.26  | 12       | 0.23 | <b>4</b>  | 1    | 1.22  | 14       | 0.03 | <b>24</b>   | 1040 | -0.21 | 0      | 0.00 | <b>25</b>  | 609  | 0.00 |
| 3   | 3      | 0.06 | <b>11</b>  | 14   | 0.19  | 15       | 0.27 | <b>3</b>  | -1   | 1.38  | 23       | 0.03 | <b>8</b>    | 33   | 0.12  | 3      | 0.05 | <b>10</b>  | 12   | 0.01 |
| 4   | 5      | 0.07 | <b>9</b>   | 5    | 0.30  | 8        | 0.04 | <b>23</b> | 41   | 0.04  | 8        | 0.03 | <b>21</b>   | 114  | -0.02 | 4      | 0.03 | <b>14</b>  | 24   | 0.00 |
| 5   | 8      | 0.06 | <b>14</b>  | 11   | 0.21  | 1        | 0.03 | <b>25</b> | 49   | 0.00  | 11       | 0.03 | <b>15</b>   | 36   | 0.13  | 3      | 0.04 | <b>12</b>  | 14   | 0.01 |
| 6   | 4      | 0.04 | <b>20</b>  | 37   | 0.00  | 3        | 0.03 | <b>19</b> | 46   | 0.01  | 15       | 0.05 | <b>6</b>    | 5    | 0.67  | 14     | 0.08 | <b>7</b>   | 3    | 0.07 |
| 7   | 6      | 0.03 | <b>37</b>  | 699  | -0.05 | 12       | 0.21 | <b>5</b>  | -2   | 1.07  | 27       | 0.05 | <b>7</b>    | 0    | 0.83  | 2      | 0.03 | <b>19</b>  | 24   | 0.00 |
| 8   | 0      | 0.00 | <b>714</b> | 706  | 0.00  | 10       | 0.05 | <b>20</b> | 25   | 0.11  | 6        | 0.03 | <b>20</b>   | 65   | 0.01  | 0      | 0.00 | <b>49</b>  | 602  | 0.00 |
| 9   | 9      | 0.02 | <b>710</b> | 703  | -0.16 | 11       | 0.02 | <b>59</b> | 506  | -0.05 | 9        | 0.03 | <b>33</b>   | 57   | 0.05  | 1      | 0.03 | <b>32</b>  | 20   | 0.00 |
| 10  | 1      | 0.04 | <b>28</b>  | 34   | 0.00  | 2        | 0.03 | <b>33</b> | 48   | 0.00  | 54       | 0.00 | <b>1044</b> | 1034 | -1.41 | 2      | 0.03 | <b>17</b>  | 20   | 0.00 |
| 11  | 5      | 0.04 | <b>25</b>  | 15   | 0.07  | 2        | 0.03 | <b>41</b> | 48   | 0.00  | 5        | 0.03 | <b>29</b>   | 52   | 0.07  | 2      | 0.03 | <b>26</b>  | 23   | 0.00 |
| 12  | 0      | 0.00 | <b>713</b> | 703  | 0.00  | 5        | 0.04 | <b>17</b> | 28   | 0.08  | 28       | 0.09 | <b>2</b>    | -10  | 2.60  | 6      | 0.21 | <b>2</b>   | -10  | 0.27 |
| 13  | 12     | 0.06 | <b>13</b>  | 0    | 0.25  | 5        | 0.07 | <b>14</b> | 7    | 0.26  | 18       | 0.03 | <b>23</b>   | 61   | 0.04  | 2      | 0.03 | <b>16</b>  | 22   | 0.00 |
| 14  | 4      | 0.04 | <b>19</b>  | 38   | 0.00  | 0        | 0.00 | <b>24</b> | 510  | 0.00  | 20       | 0.12 | <b>1</b>    | -13  | 4.42  | 5      | 0.04 | <b>9</b>   | 6    | 0.01 |
| 15  | 9      | 0.11 | <b>4</b>   | -11  | 0.67  | 5        | 0.06 | <b>29</b> | 7    | 0.23  | 1        | 0.03 | <b>77</b>   | 95   | 0.00  | 3      | 0.06 | <b>18</b>  | -2   | 0.05 |
| 16  | 4      | 0.04 | <b>34</b>  | 32   | 0.00  | 9        | 0.04 | <b>32</b> | 23   | 0.10  | 8        | 0.03 | <b>27</b>   | 33   | 0.12  | 3      | 0.03 | <b>24</b>  | 16   | 0.00 |
| 17  | 0      | 0.00 | <b>720</b> | 699  | 0.00  | 9        | 0.05 | <b>28</b> | 14   | 0.15  | 10       | 0.03 | <b>39</b>   | 52   | 0.05  | 0      | 0.00 | <b>163</b> | 597  | 0.00 |
| 18  | 3      | 0.04 | <b>31</b>  | 12   | 0.04  | 10       | 0.15 | <b>10</b> | -9   | 0.72  | 4        | 0.03 | <b>42</b>   | 88   | 0.00  | 0      | 0.00 | <b>137</b> | 595  | 0.00 |
| 19  | 3      | 0.04 | <b>44</b>  | 40   | 0.00  | 6        | 0.11 | <b>12</b> | -5   | 0.54  | 3        | 0.03 | <b>60</b>   | 120  | 0.00  | 4      | 0.05 | <b>21</b>  | -5   | 0.02 |
| 20  | 4      | 0.06 | 15         | -10  | 0.27  | 11       | 0.02 | <b>60</b> | 494  | -0.04 | 8        | 0.03 | <b>52</b>   | 1004 | -0.04 | 2      | 0.03 | <b>36</b>  | 35   | 0.00 |
| D   | 1.0    | 0.6  | -0.4       | -0.3 | 0.4   | 1.0      | 0.7  | -0.2      | -0.1 | 0.6   | 1.0      | 0.1  | 0.6         | 0.2  | 0.1   | 1.0    | 0.8  | -0.4       | -0.4 | 0.8  |
| A   |        | 1.0  | -0.8       | -0.8 | 0.9   |          | 1.0  | -0.7      | -0.4 | 1.0   |          | 1.0  | -0.4        | -0.4 | 1.0   |        | 0.8  | -0.4       | -0.4 | 0.8  |
| P   |        |      | 1.0        | 0.9  | -0.4  |          |      | 1.0       | 0.7  | -0.7  |          |      | 1.0         | 0.6  | -0.4  |        |      | 1.0        | 0.7  | -0.3 |
| Ch  |        |      |            | 1.0  | -0.5  |          |      |           | 1.0  | -0.4  |          |      |             | 1.0  | -0.4  |        |      |            | 1.0  | -0.2 |

The rank position index as obtained by the AHP rank is depicted as bold face entry.

computed by using the Eq. (17). As state before, the sign has the following meaning:

$$Tr = \begin{cases} \text{positive sign} & , \text{if trend is increasing} \\ \text{negative sign} & , \text{if trend is decreasing} \\ 0 & , \text{otherwise.} \end{cases}$$

To show the relationship between two metrics, at the bottom of Tables 3 and 4 we show the correlation coefficient among various metrics. As usual, the correlation coefficient takes a value between  $[-1, 1]$ , with 1 or  $-1$  indicating perfect correlation. A positive correlation depicts a positive association between the variables. Thus, increasing values of one variable correspond to increasing values of the other variable. On the other hand, negative correlation indicates a negative association between the variables. Thus, increasing values of one variable correspond to decreasing values of the other variable. A correlation value close to 0 indicates no association between the variables.

Table 3 shows the results for the topics 'networks', 'systems', 'software', and 'services'. The organization *PNr* 1 has been ranked by AHP at position 1 in 'networks', position 1 in 'systems', position 2 in 'software', and position 5 in 'services'. With regards to 'services', a negative trend is shown for *PNr* 1 and thus the position has dropped in this topic. In the other topics, positive trend can be observed and thus the ranking position was mostly preserved. With regards to the topic 'systems', a very good trend of 7.97 can be observed and highest authority score of 0.86 within the table. As one can see, by applying our approach, much more fine-grained ranking can be performed by considering topic information.

With regards to correlation, *A* always correlates perfectly with *Tr* because time-aware authority takes trend through personalization into account. *D* shows good correlation with *Tr* in the topic 'systems'. This is a result of the broad scope of 'systems' and the high frequency of the topic within projects (see also Table 1).

Table 4 shows the results for the topics 'health', 'embedded', 'internet', and 'energy'. The organization *PNr* 1 was only ranked in 'energy' at position 1 but not for the other topics. One exceptionally high change in the ranking position can be seen for organization *PNr* 10 in 'internet' which ranks by AHP at 1044. *PNr* 10 had some substantial amounts of projects with regard to 'internet' in the past (54 matching projects as indicated by *D*) but the trend is highly negative (*Tr* is  $-1.41$ , which is the lowest in the table) and time-aware *A* is 0.00. Thus, we believe that negative trend and limited recent activity in the context 'internet' justifies a change in the rank position.

With regards to correlation, *A* correlates perfectly only in 'embedded' and 'internet' but not for the other topics (although a high correlation is still achieved). *D* shows good correlation with *Tr* in the topic 'energy'. As in Table 3, *A* shows good correlation with *P*. Indeed, authority is part of AHP's ranking criteria so a correlation can be expected. Recall, higher authority yields better positions. Thus, negative correlation means increasing values of authority correspond to decreasing values of the rank position (lower position value is better).

Based on the data in Tables 3 and 4, average values of degree, position, and change are depicted in Fig. 4 and average values of authority and trend are shown in Fig. 5. Average values are based on the metric values of the top-20 list of organizations. Again, the baseline algorithm for ranking is the degree-based rank.

The topic 'networks' has the highest average value with regard to change. Thus, AHP rankings based on topic information have significant impact on the ranking position of organizations and a lot of changes are observed within the top-20 list. Also the topics 'health' and 'internet' yield high changes on average. However, only 'health' yields also high average values with regards to position. This means that organizations ranked within top-20 positions by the degree-based rank would be ranked at much higher positions by AHP in the 'health' topic. As mentioned before, since 'systems' is a very broad topic also the positions by AHP are quite similar when compared with the degree-based rank

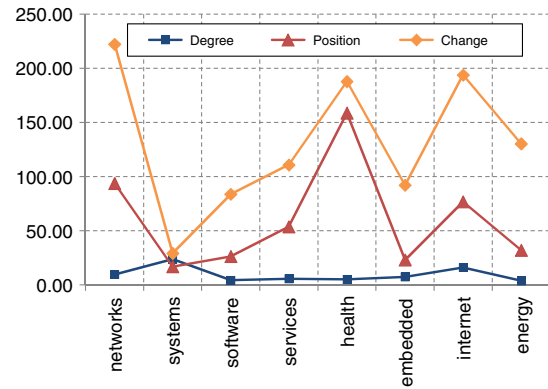


Fig. 4. Average degree, position, and change.

(the lowest average value as depicted by Fig. 4). The average degree does not significantly change across topics. Generally, topic based personalization has the effect that significant changes of rank position can be expected.

Next, Fig. 5 shows the average values for authority and trend. The topic 'systems' shows the highest average authority and the highest average trend. This observation is also consistent with the previous discussion. The topic 'software' shows the lowest trend and also a low average value for authority. In general, deviations in authority across topics are very high.

To summarize the main observations in this section:

- Our proposed model enables more fine-grained ranking by considering topic information.
- Authority correlates to a high degree with trend because time-aware authority takes trend through personalization into account.
- Generally, topic based personalization has the effect that significant changes of rank position can be observed.
- Topics that play a role in many projects (having a broad scope) correlate better with degree-based ranking. Thus, no significant changes through personalization can be expected.
- As a consequence of the previous observation, by focusing on narrow and more specialized topics organizations with fewer projects are able to build up authority and are thereby ranked at better positions in those topics.

### 7.3. Statistical comparison

Here a statistical comparison of ranking techniques is performed. In the previous section, a top-20 list of organizations was selected (as ranked by the organizations' degree) and evaluated by using different metrics. In this section we use a set overlap and distance based ranking metric to compare the AHP based results with non-personalized

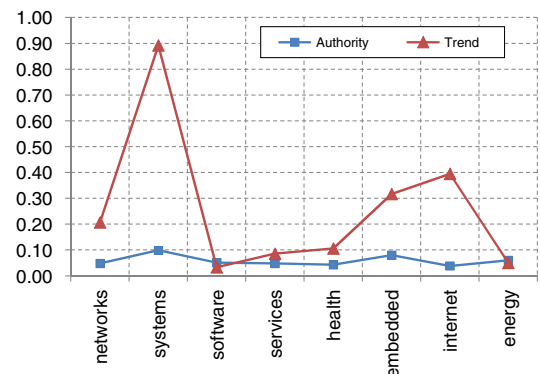


Fig. 5. Average authority score and trend.

rankings including the degree-based rank, a funding based rank, and the structural rank.

The funding based rank uses the total amount of funding received by an organization to perform ranking (the higher the total funding the better the rank). The structural importance rank is used in isolation of AHP and compared with the regular AHP using the criteria authority, structural importance, and cost. After that a cross topic comparison is performed by using AHP and authority based rankings and AHP-based rankings personalized for different topics. AHP is setup with the weights 0.4 for authority, 0.2 for the structural importance rank and 0.4 for cost.

To systematically compare results of two ranking algorithms, let us define two standard ranking metrics.

7.3.1. OSim@k

To measure similarity of top-k sets, let us define overlap similarity as follows:

$$OSim@k = \frac{O_{k1} \cap O_{k2}}{k} \tag{25}$$

OSim@k defines the overlap similarity of the top-k sets  $O_k$  ranked by two algorithms. Each set consists of organizations such that  $O_k \subset V_O$ . The first algorithm is always AHP, which has been parameterized using the same weights as defined previously.

7.3.2. Kendall's  $\tau$

The next ranking metric used in this work is the well-known Kendall's  $\tau$  metric (for example, see [30]):

$$Kendall's \tau = \frac{2(\text{num\_concordant} - \text{num\_discordant})}{|V_O|(|V_O| - 1)} \tag{26}$$

Consider the pair of nodes  $o, u$ . The pair is concordant if two rankings agree on the order and discordant if both rankings disagree on the order. Denote the number of these pairs by num\_concordant and num\_disdiscordant respectively. The total number of pairs is given as  $\frac{|V_O|(|V_O|-1)}{2}$ . Kendall's  $\tau$  is defined between the interval  $\tau \in [-1, 1]$ . Kendall's  $\tau$  helps in analyzing if two ranking algorithms are rank similar. If  $\tau$  equals 1, there are no cases where the pair  $o, u$  is ranked in a different order.

Table 5 shows the comparison results of AHP-based rankings (for the top-10 topics in Table 1) and the degree-based, funding-based, and structural importance rank. The highest values for OSim and Kendall's  $\tau$  are depicted as bold-face numbers. The topic 'systems' clearly shows the highest overlap with the other (non-topic based) rankings. OSim@10, OSim@20, and OSim@50 show the highest overlap in each topic. This observation is again in line with the previous discussion. Previously 'systems' showed the highest average authority and the highest average trend within the top-20 list of organizations. The structural importance

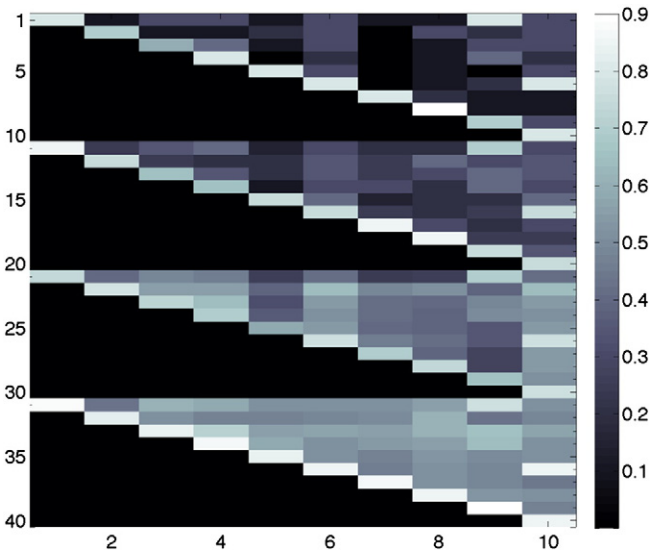


Fig. 6. OSim and Kendall's  $\tau$  for comparison of AHP with authority-based rankings (detailed numbers are available in Table 6).

rank shows the highest overlap of 0.70 in the top-10 segment (depicted as OSim@10). However, a higher agreement in the rank order as measured through Kendall's  $\tau$  is given in the topic 'software'. Kendall's  $\tau$  is calculated by using the whole list of ranked organizations. Whereas the highest overlap of AHP-based rankings with the degree-based, funding-based, and structural importance rank is given in 'systems', higher agreement in terms of Kendall's  $\tau$  is given in 'software'.

Fig. 6 shows the comparison results of AHP-based rankings (again for the top-10 topics in Table 1) and the authority-based rankings. Here, for each topic ranking is performed using AHP as defined in Eq. (1) and authority as defined in Eq. (19). The results are then compared using OSim and Kendall's  $\tau$ . Further details are provided in Table 6. The first set of rows (1–10) depicts OSim@10, the second set of rows (11–20) depicts OSim@20, the third set of rows (21–30) depicts OSim@50, and the fourth set of rows (31–40) depicts Kendall's  $\tau$ .

The values below the matrix diagonal (from top-left to bottom right corner) are all set to 0 because of symmetry. For example, overlap similarity OSim for the topics 'networks' and 'systems' yields the same results as 'systems' and 'networks'. At the diagonal values comparison of AHP and authority rankings for the same topic was performed. Thus, high overlap and agreement with regard to OSim and Kendall's  $\tau$ , respectively, can be observed. Fig. 7 shows the average values of OSim@10, OSim@20, OSim@50, and Kendall's  $\tau$  for each topic. With regard to OSim@10, 'health' yields the lowest average overlap similarity. The topics 'efficiency' and 'energy' have the highest overlap similarities in the top-10 segment.

Table 5 OSim and Kendall's  $\tau$  for comparison of AHP with degree, funding, and structural rank.

|            |                  | Networks | Systems     | Software    | Services | Transport | Efficiency | Health | Embedded | Internet | Energy |
|------------|------------------|----------|-------------|-------------|----------|-----------|------------|--------|----------|----------|--------|
| Degree     | OSim@10          | 0.30     | <b>0.60</b> | 0.40        | 0.20     | 0.30      | 0.30       | 0.30   | 0.40     | 0.40     | 0.30   |
|            | OSim@20          | 0.40     | <b>0.75</b> | 0.55        | 0.50     | 0.40      | 0.55       | 0.50   | 0.50     | 0.40     | 0.55   |
|            | OSim@50          | 0.56     | <b>0.86</b> | 0.72        | 0.74     | 0.70      | 0.78       | 0.66   | 0.68     | 0.70     | 0.78   |
|            | Kendall's $\tau$ | 0.44     | 0.42        | <b>0.46</b> | 0.45     | 0.44      | 0.41       | 0.41   | 0.45     | 0.43     | 0.41   |
| Funding    | OSim@10          | 0.30     | <b>0.50</b> | <b>0.50</b> | 0.30     | 0.30      | 0.40       | 0.30   | 0.30     | 0.40     | 0.40   |
|            | OSim@20          | 0.40     | <b>0.70</b> | 0.50        | 0.55     | 0.40      | 0.55       | 0.40   | 0.40     | 0.60     | 0.55   |
|            | OSim@50          | 0.56     | <b>0.84</b> | 0.76        | 0.72     | 0.70      | 0.76       | 0.74   | 0.66     | 0.70     | 0.76   |
|            | Kendall's $\tau$ | 0.63     | 0.54        | <b>0.65</b> | 0.64     | 0.60      | 0.58       | 0.61   | 0.61     | 0.62     | 0.58   |
| Structural | OSim@10          | 0.40     | <b>0.70</b> | 0.60        | 0.40     | 0.40      | 0.60       | 0.30   | 0.30     | 0.50     | 0.60   |
|            | OSim@20          | 0.40     | <b>0.70</b> | 0.65        | 0.55     | 0.45      | 0.60       | 0.40   | 0.45     | 0.60     | 0.60   |
|            | OSim@50          | 0.56     | <b>0.86</b> | 0.80        | 0.82     | 0.80      | 0.84       | 0.66   | 0.72     | 0.68     | 0.84   |
|            | Kendall's $\tau$ | 0.48     | 0.47        | <b>0.51</b> | 0.49     | 0.49      | 0.46       | 0.46   | 0.50     | 0.46     | 0.46   |

The rank position index as obtained by the AHP rank is depicted as bold face entry.

**Table 6**  
OSim and Kendall's  $\tau$  for comparison of AHP with authority-based rankings.

|                  |            | Networks | Systems | Software | Services | Transport | Efficiency | Health | Embedded | Internet | Energy |
|------------------|------------|----------|---------|----------|----------|-----------|------------|--------|----------|----------|--------|
| OSim@10          | Networks   | 0.80     | 0.10    | 0.30     | 0.30     | 0.10      | 0.30       | 0.10   | 0.10     | 0.80     | 0.30   |
|                  | Systems    | 0.00     | 0.70    | 0.10     | 0.10     | 0.20      | 0.30       | 0.00   | 0.30     | 0.20     | 0.30   |
|                  | Software   | 0.00     | 0.00    | 0.60     | 0.40     | 0.10      | 0.30       | 0.00   | 0.10     | 0.30     | 0.30   |
|                  | Services   | 0.00     | 0.00    | 0.00     | 0.80     | 0.00      | 0.20       | 0.00   | 0.10     | 0.40     | 0.20   |
|                  | Transport  | 0.00     | 0.00    | 0.00     | 0.00     | 0.80      | 0.30       | 0.00   | 0.10     | 0.00     | 0.30   |
|                  | Efficiency | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.80       | 0.00   | 0.10     | 0.20     | 0.80   |
|                  | Health     | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.00       | 0.80   | 0.20     | 0.10     | 0.10   |
|                  | Embedded   | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.00       | 0.00   | 0.90     | 0.10     | 0.10   |
|                  | Internet   | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.00       | 0.00   | 0.00     | 0.70     | 0.30   |
|                  | Energy     | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.00       | 0.00   | 0.00     | 0.00     | 0.80   |
| OSim@20          | Networks   | 0.85     | 0.25    | 0.35     | 0.40     | 0.15      | 0.30       | 0.20   | 0.20     | 0.70     | 0.30   |
|                  | Systems    | 0.00     | 0.75    | 0.25     | 0.20     | 0.20      | 0.35       | 0.25   | 0.40     | 0.30     | 0.35   |
|                  | Software   | 0.00     | 0.00    | 0.65     | 0.35     | 0.20      | 0.35       | 0.25   | 0.30     | 0.40     | 0.35   |
|                  | Services   | 0.00     | 0.00    | 0.00     | 0.65     | 0.10      | 0.30       | 0.30   | 0.20     | 0.40     | 0.30   |
|                  | Transport  | 0.00     | 0.00    | 0.00     | 0.00     | 0.75      | 0.40       | 0.15   | 0.20     | 0.20     | 0.40   |
|                  | Efficiency | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.75       | 0.25   | 0.20     | 0.25     | 0.75   |
|                  | Health     | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.00       | 0.85   | 0.30     | 0.20     | 0.30   |
|                  | Embedded   | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.00       | 0.00   | 0.85     | 0.25     | 0.25   |
|                  | Internet   | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.00       | 0.00   | 0.00     | 0.75     | 0.35   |
|                  | Energy     | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.00       | 0.00   | 0.00     | 0.00     | 0.75   |
| OSim@50          | Networks   | 0.74     | 0.40    | 0.48     | 0.46     | 0.26      | 0.42       | 0.24   | 0.26     | 0.70     | 0.42   |
|                  | Systems    | 0.00     | 0.78    | 0.56     | 0.56     | 0.38      | 0.64       | 0.48   | 0.52     | 0.38     | 0.64   |
|                  | Software   | 0.00     | 0.00    | 0.72     | 0.64     | 0.32      | 0.54       | 0.42   | 0.40     | 0.48     | 0.54   |
|                  | Services   | 0.00     | 0.00    | 0.00     | 0.70     | 0.36      | 0.52       | 0.42   | 0.38     | 0.50     | 0.52   |
|                  | Transport  | 0.00     | 0.00    | 0.00     | 0.00     | 0.58      | 0.54       | 0.40   | 0.38     | 0.34     | 0.54   |
|                  | Efficiency | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.76       | 0.46   | 0.42     | 0.34     | 0.76   |
|                  | Health     | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.00       | 0.70   | 0.38     | 0.28     | 0.54   |
|                  | Embedded   | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.00       | 0.00   | 0.74     | 0.28     | 0.54   |
|                  | Internet   | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.00       | 0.00   | 0.00     | 0.66     | 0.52   |
|                  | Energy     | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.00       | 0.00   | 0.00     | 0.00     | 0.76   |
| Kendall's $\tau$ | Networks   | 0.88     | 0.43    | 0.61     | 0.58     | 0.52      | 0.51       | 0.51   | 0.56     | 0.77     | 0.51   |
|                  | Systems    | 0.00     | 0.82    | 0.52     | 0.47     | 0.47      | 0.48       | 0.50   | 0.61     | 0.43     | 0.48   |
|                  | Software   | 0.00     | 0.00    | 0.84     | 0.71     | 0.55      | 0.57       | 0.55   | 0.61     | 0.66     | 0.57   |
|                  | Services   | 0.00     | 0.00    | 0.00     | 0.87     | 0.59      | 0.52       | 0.54   | 0.55     | 0.64     | 0.52   |
|                  | Transport  | 0.00     | 0.00    | 0.00     | 0.00     | 0.83      | 0.52       | 0.47   | 0.53     | 0.50     | 0.52   |
|                  | Efficiency | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.85       | 0.46   | 0.52     | 0.49     | 0.85   |
|                  | Health     | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.00       | 0.87   | 0.52     | 0.48     | 0.44   |
|                  | Embedded   | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.00       | 0.00   | 0.85     | 0.53     | 0.51   |
|                  | Internet   | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.00       | 0.00   | 0.00     | 0.89     | 0.47   |
|                  | Energy     | 0.00     | 0.00    | 0.00     | 0.00     | 0.00      | 0.00       | 0.00   | 0.00     | 0.00     | 0.85   |

Fig. 9 shows the comparison results of AHP-based rankings for the top-10 topics in Table 1 across topics. This comparison shows how ranking results change by considering different topics. Further details are provided in Table 7.

Rows are segmented in the same manner as already described previously. Values at the matrix diagonal (from top-left to bottom right corner) are all 1. The values below the matrix diagonal are all set to 0 for the previously mentioned reason. Fig. 8 shows the average values of OSim@10, OSim@20, OSim@50, and Kendall's  $\tau$  for each topic.

In OSim@10 the topic 'health' results in the lowest average overlap similarity followed by the topic 'embedded', which has also low overlap similarity. In general, higher average values of OSim@10, OSim@20, OSim@50 as well as Kendall's  $\tau$  can be observed when compared with the previous discussion. Higher values are the result of the same ranking technique being used (AHP-based rankings) and results being compared across topics. Before the AHP-based rankings were compared with authority, which is only one of the ranking criteria being used in AHP.

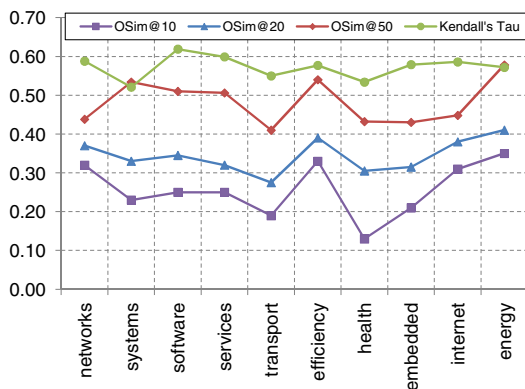


Fig. 7. Average values of OSim@10, OSim@20, OSim@50, and Kendall's  $\tau$  based on Table 6.

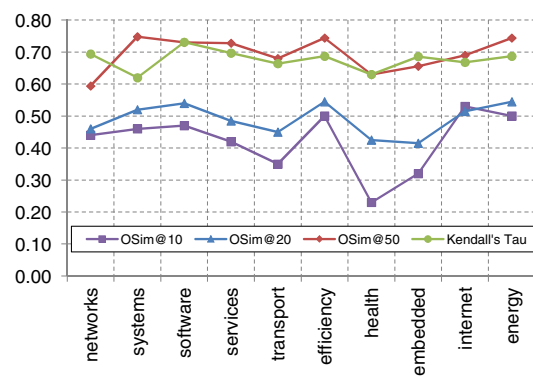


Fig. 8. Average values of OSim@10, OSim@20, OSim@50, and Kendall's  $\tau$  based on Table 7.





## References

- [1] A. Anagnostopoulos, L. Becchetti, C. Castillo, A. Gionis, S. Leonardi, Online team formation in social networks, Proceedings of the 21st International Conference on World Wide Web. WWW '12, ACM, New York, NY, USA, 2012, pp. 839–848.
- [2] L. Backstrom, D. Huttenlocher, J. Kleinberg, X. Lan, Group formation in large social networks: membership, growth, and evolution, Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '06, ACM, New York, NY, USA, 2006, pp. 44–54.
- [3] K. Berberich, M. Vazirgiannis, G. Weikum, T-rank: time-aware authority ranking, in: S. Leonardi (Ed.), Algorithms and Models for the Web-Graph, Lecture Notes in Computer Science, vol. 3243, Springer, Berlin Heidelberg, 2004, pp. 131–142.
- [4] P. Berkhin, Survey: a survey on PageRank computing, Internet Mathematics 2 (1) (2005) 73–120.
- [5] R.S. Burt, Structural Holes: The Social Structure of Competition, Harvard University Press, 1992.
- [6] R.S. Burt, Structural holes and good ideas, American Journal of Sociology 110 (2) (Sept. 2004) 349–399.
- [7] L.M. Camarinha-Matos, H. Afsarmanesh, Collaborative networks, PROLAMAT, 2006, pp. 26–40.
- [8] A. Certa, M. Enea, T. Lupo, ELECTRE III to dynamically support the decision maker about the periodic replacements configurations for a multi-component system, Decision Support Systems 55 (1) (Apr. 2013) 126–134.
- [9] S. Chakrabarti, Dynamic personalized PageRank in entity-relation graphs, Proceedings of the 16th International Conference on World Wide Web. WWW '07, ACM, New York, NY, USA, 2007, pp. 571–580, <http://dx.doi.org/10.1145/1242572.1242650>.
- [10] H. Deng, M.R. Lyu, I. King, A generalized co-hits algorithm and its application to bipartite graphs, Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '09, ACM, New York, NY, USA, 2009, pp. 239–248, <http://dx.doi.org/10.1145/1557019.1557051>.
- [11] Y. Ding, Scientific collaboration and endorsement: network analysis of coauthorship and citation networks, Journal of Informetrics 5 (1) (2011) 187–203.
- [12] P. Ferrari, A method for choosing from among alternative transportation projects, European Journal of Operational Research 150 (1) (2003) 194–203.
- [13] D. Fogaras, K. Csalogany, B. Racz, T. Sarlos, Towards scaling fully personalized PageRank: algorithms, lower bounds, and experiments, Internet Mathematics 2 (3) (2005) 333–358.
- [14] F. Fu, C. Hauert, M.A. Nowak, L. Wang, Reputation-based partner choice promotes cooperation in social networks, Physical Review E 78 (Aug 2008) 026117.
- [15] S. Goyala, F. Vega-Redondo, Structural holes in social networks, Journal of Economic Theory 137 (1) (November 2007) 460–492.
- [16] M.S. Granovetter, The strength of weak ties, The American Journal of Sociology 78 (6) (1973) 1360–1380.
- [17] R. Guns, Y. Liu, D. Mahbuba, Q-measures and betweenness centrality in a collaboration network: a case study of the field of informetrics, Scientometrics 87 (1) (2011) 133–147, <http://dx.doi.org/10.1007/s11192-010-0332-3>.
- [18] T. Haveliwala, S. Kamvar, G. Jeh, An analytical comparison of approaches to personalizing PageRank, Tech. Rep, Stanford University, 2003.
- [19] T.H. Haveliwala, Topic-sensitive PageRank, Proceedings of the 11th International Conference on World Wide Web. WWW '02, ACM, New York, NY, USA, 2002, pp. 517–526, <http://dx.doi.org/10.1145/511446.511513>.
- [20] G. Jeh, J. Widom, Scaling personalized web search, Proceedings of the 12th International Conference on World Wide Web. WWW '03, ACM, New York, NY, USA, 2003, pp. 271–279, <http://dx.doi.org/10.1145/775152.775191>.
- [21] J. Kleinberg, Authoritative sources in a hyperlinked environment, Journal of the ACM 46 (1999) 668–677.
- [22] J. Kleinberg, S. Suri, E. Tardos, T. Wexler, Strategic network formation with structural holes, SIGecom Exchanges 7 (3) (2008).
- [23] N. Lavrac, P. Ljubic, T. Urbancic, G. Papa, M. Jermol, S. Bollhalter, Trust modeling for networked organizations using reputation and collaboration estimates, IEEE Transactions on Systems, Man, and Cybernetics. Part C: Applications and Reviews 37 (3) (May 2007) 429–439.
- [24] S. Milojević, Modes of collaboration in modern science: beyond power laws and preferential attachment, Journal of the American Society for Information Science and Technology 61 (7) (Jul. 2010) 1410–1423.
- [25] F. Munisteri, ICT statistical report for annual monitoring 2011, [http://ec.europa.eu/digital-agenda/sites/digital-agenda/files/stream\\_2012\\_0.pdf](http://ec.europa.eu/digital-agenda/sites/digital-agenda/files/stream_2012_0.pdf), Feb. 2012.
- [26] M.E.J. Newman, Coauthorship networks and patterns of scientific collaboration, Proceedings of the National Academy of Sciences of the United States of America 101 (Suppl. 1) (2004) 5200–5205.
- [27] A.Y. Ng, A.X. Zheng, M.I. Jordan, Stable algorithms for link analysis, Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in x. SIGIR '01, ACM, New York, NY, USA, 2001, pp. 258–266.
- [28] L. Page, S. Brin, R. Motwani, T. Winograd, The PageRank citation ranking: bringing order to the web, Tech. Rep, Stanford University, 1998.
- [29] T.L. Saaty, Decision making with the analytic hierarchy process, International Journal of Services Sciences 1 (Nov. 2008) 83–98.
- [30] D. Schall, Expertise ranking using activity and contextual link measures, Data & Knowledge Engineering 71 (1) (2012) 92–113.
- [31] D. Schall, Service oriented crowdsourcing: architecture, protocols and algorithms, Springer Briefs in Computer Science, Springer, New York, New York, NY, USA, 2012.
- [32] D. Schall, Measuring contextual partner importance in scientific collaboration networks, Journal of Informetrics 7 (3) (July 2013) 730–736.
- [33] D.H. Sonnenwald, B. Cronin, Anonymous, Scientific collaboration: a synthesis of challenges and strategies, Annual Review of Information Science and Technology, vol. 4, Information Today, 2007, pp. 2–37.
- [34] W. Tsai, Social capital, strategic relatedness, and the formation of intra-organizational strategic linkages, Strategic Management Journal 21 (9) (2000) 925–939.
- [35] C.S. Wagner, L. Leydesdorff, Network structure, self-organization, and the growth of international collaboration in science, Research Policy 34 (10) (2005) 1608–1618.

**Daniel Schall** is currently employed as a senior research scientist at Siemens Corporate Technology. From 03/2009 to 10/2011, he worked as a senior research scientist at the Vienna University of Technology. Daniel defended his PhD in 02/2009 with a thesis on 'Human Interactions in Mixed Systems — Architecture, Protocols, and Algorithms', which he performed while working as a research assistant at the Vienna University of Technology (05/2006–02/2009). Prior to that, he worked at Siemens Corporate Research in Princeton, New Jersey, USA.