Random Walks on the Reputation Graph

ABSTRACT

The identification of reputable entities is an important task in business, education, and many other fields. On the other hand, as an arguably subjective, multi-faceted concept, quantifying reputation is challenging. In this paper, instead of relying on a single, precise definition of reputation, we propose to exploit the *transference* of reputation among entities in order to identify the most reputable ones. To this end, we propose a novel random walk model to infer the reputation of a target set of entities with respect to suitable sources of reputation. We instantiate our model in an academic search setting, by modeling research groups as reputation sources and publication venues as reputation targets. By relying on publishing behavior as a reputation signal, we demonstrate the effectiveness of our model in contrast to standard citation-based approaches for identifying reputable venues as well as researchers in the broad area of computer science. In addition, we demonstrate the robustness of our model to perturbations in the selection of reputation sources. Finally, we show that effective reputation sources can be chosen via the proposed model itself in a semi-automatic fashion.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: *Stochastic processes*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Retrieval models*

Keywords

Reputation flows, random walks, academic search

1. INTRODUCTION

Reputation is a widespread notion in society, albeit an arguably ill-defined one. In general, the reputation of an entity reflects the public perception about this entity developed over time. This public perception may be either good or bad, and touches a variety of aspects that may impact the identity of the entity before the public, such as its competence, integrity, and trustworthiness. Moreover, the reputation of an entity can change rapidly following an event in which the entity is involved, by means of word-of-mouth dissemination—whether traditional or electronic. As a result, reputation has been subject of professional management by public relations departments as well as of collective management by members of online communities, such as question-answering forums and online marketplaces [13].

The identification of reputable entities is an important task in many fields. Indeed, more reputable entities are presumably a better fit for most purposes. However, the subjective nature of reputation makes its quantification—and hence the identification of reputable entities—challenging. As a result, existing attempts to quantify the reputation of an entity rely on either manual assessments or on a restrictive definition of reputation, e.g., in terms of authority [17, 28], influence [3], or expertise [4]. In contrast, in this paper, we take an agnostic view of reputation. In particular, instead of relying on a single, precise definition of reputation, we propose to exploit the *transference* of reputation among entities in order to identify the most reputable ones.

In this paper, we propose a novel random walk model for ranking a target set of entities with respect to suitable sources of reputation. To this end, we model reputation sources and reputation targets as nodes in a heterogeneous graph, with edges connecting any pair of nodes whenever there is a reputation transfer between them. To validate our proposed model, we instantiate this so-called *reputation graph* in an academic search setting. In particular, we model research groups as reputation sources and publication venues as reputation targets, with edges running from a source to a target and back again to indicate the transference of reputation through one or more publications. Through a series of experiments, we empirically demonstrate the effectiveness of our model in contrast to standard citation-based approaches for identifying reputable venues as well as individual researchers in the broad area of computer science. In addition, we demonstrate the robustness of our model to random perturbations in the selection of reputation sources, and the suitability of semi-automatically choosing effective reputation sources using the model itself.

In summary, our main contributions are:

1. A novel random walk model for ranking entities according to the reputation collectively transferred to them by other entities in a reputation graph.

ICTIR'15 , September 27–30, Northampton, MA, USA.

2. An empirical validation of the effectiveness and robustness of our proposed model for two academic search tasks, namely, venue and researcher ranking.

3. A preliminary investigation of the suitability of automatically choosing effective reputation sources.

In the remainder of this paper, Section 2 presents related work on ranking based on random walks as well as in ranking in an academic setting. Section 3 introduces our proposed random walk model for reputation-oriented ranking, whereas Section 4 describes its instantiation in an academic search setting. Sections 5 and 6 describe the setup and the results of the empirical evaluation of our model. Lastly, Section 7 provides our concluding remarks.

2. RELATED WORK

In this section, we review the related literature on ranking based on random walks, as well as approaches devoted to generating rankings in an academic search setting.

2.1 Ranking with Random Walks

Page and Brin [28] designed the PageRank algorithm to calculate the importance of pages on the Web. PageRank simulates a web surfer's behavior. In particular, with probability $p < 1$, the surfer randomly chooses one of the hyperlinks of the current page and jumps to the page it links to; otherwise, with probability $1 - p$, the user jumps to a web page chosen uniformly at random from the collection. This defines a Markov chain on the web graph, where each probability of the stationary distribution corresponds to the rank of a web page, referred to as its *pagerank*.

Kleinberg [17] divided the notion of "importance" of a web page into two related attributes: *hub*, measured by the authority score of other pages that the page links to, and *authority*, measured by the hub score of the pages that link to the page. These attributes are calculated in his Hyperlinked-Induced Topic Search (HITS) algorithm. Both algorithms, PageRank and HITS, have been successfully applied to rank the importance of different web pages through analyzing the link structure of the web graph.

Extensions of the random walk model were also studied for scoring several types of objects—e.g., products, people and organizations—in different applications. For instance, Nie et al. [27] presented PopRank, a domain-independent object-level link analysis model to rank objects within a specific domain, by assigning a popularity propagation factor to each type of object relationship. Different popularity propagation factors for these heterogeneous relationships were assessed with respect to their impact on the global popularity ranking. Xi et al. [36] proposed a unified link analysis framework, called Link Fusion, which considers two different categories of links: intra-type links, which represent the relationship of data objects of a homogeneous data type (e.g., web pages), and inter-type links, which represent the relationship of data objects of different data types (e.g., between users and web pages). Regarding the recommendation of generic types of object, Jamali and Ester [14] proposed TrustWalker, a random walk method that combines trust-based and item-based recommendation, considering not only ratings of the target item, but also those of similar items.

Under the context of social networking systems, social friendship and random walks have been shown to be beneficial for collaborative filtering-based recommendation systems. These works argue that social friends—for instance, in Facebook or Twitter—tend to share common interests and thus their relationships should be considered in the process of collaborative filtering [39]. In this context, a random walk sees a social network as a graph with probabilistically weighted links that represent social relations and thus is able to accurately predict users' preferences to items and their social influence with respect to other users. Backstrom and Leskovec [2] proposed an algorithm based on supervised random walks that combines the information from the network structure with node and edge level attributes, using these attributes to guide the random walk on the graph. Weng et al. [34] proposed TwitterRank to measure the influence of users in Twitter, considering both the topical similarity between users and the link structure of the social network.

2.2 Ranking in Academic Search

Ranking has traditionally played an important role in academic search, particularly for tasks related to assessing the scientific productivity of academic entities. In particular, one of the earliest metrics proposed to quantify academic impact was Garfield's Impact Factor [9]. Despite its wide usage since it was proposed in 1955, it has been largely criticized [31]. As a result, many alternatives have been proposed in the literature, such as other citation-based metrics like the H-Index [12], download-based metrics [5], and PageRank-like metrics [38]. However, as argued by Leydesdorff [20], each metric has its own bias and there are both advantages and disadvantages associated with each one.

Citation-based metrics have been applied to rank computer and information science journals [16, 25]. Also, several citation-based metrics have been proposed to measure the quality of a small set of conferences and journals in the database field [30], and to rank documents retrieved from a digital library [19]. Mann et al. [22] introduced topic modeling to further complement the citation-based bibliometric indicators, producing more fine-grained impact measures. Yan and Lee [38] proposed two measures for ranking the impact of academic venues which aim at efficiency and at mimicking the results of the widely accepted Impact Factor. An alternative method was presented by Zhuang et al. [41], who proposed a set of heuristics to automatically discover prestigious and low-quality conferences by mining the characteristics of program committee members.

Piwowar [29] recently claimed that citation-based metrics are useful, but not sufficient to evaluate research. In particular, he observed that metrics like the H-Index are slow. Indeed, the first citation of a scientific article can take years. As a result, he argued for the development of alternative metrics to complement citation analysis. In a similar vein, Lima et al. [21] argued that productivity indices should account for the singularities of the publication patterns of different research areas, in order to produce an unbiased assessment of the impact of academic output. Accordingly, they proposed to assess a researcher's productivity by aggregating his or her impact indicators across multiple areas. Finally, Gonçalves et al. [11] investigated the importance of various academic features to scholar popularity and concluded that only two features are needed to explain all the variation in popularity across different scholars: (i) the number of publications and (ii) the average quality of the scholar's publication venues. In this paper, we validate our proposed approach by exploiting exactly these these two features to rank different venues and different researchers.

The idea of reputation, instead of citations, was discussed by Nelakuditi et al. [24]. In particular, they proposed a metric called peers' reputation, which measures the selectivity of a publication venue based upon the reputation of its authors' institutions. The proposed metric was shown to be a better indicator of selectivity than the acceptance ratio. In addition, the authors observed that many conferences have similar or better peers' reputation than journals. Another approach related to ours was proposed by Cormode et al. [7], who attempted to rank authors according to their similarity with respect to a reference author. As discussed in Section 3, our approach also explores the notion of a reference source of reputation. However, in contrast to the aforementioned approaches, we explore this notion as part of a stochastic Markov process as a means to produce a global reputation-oriented ranking of multiple interconnected entities.

2.3 Random Walks in Academic Search

Earlier works have studied the application of random walks for ranking authors, papers and venues in an academic setting. For instance, Sun and Giles [32] proposed a popularity weighted ranking algorithm for academic digital libraries that uses the popularity factor of a publication venue. Their approach overcomes some limitations of the Impact Factor and performs better than PageRank, citation counting and HITS. Relatedly, Zhou et al. [40] proposed a method for co-ranking authors and their publications using several networks. Similarly, Yan et al. [37] presented a new informetric indicator, P-Rank, for measuring prestige in heterogeneous scholarly networks containing articles, authors and journals. P-Rank differentiates the weight of each citation based on its citing papers, citing journals and citing authors.

In a narrower perspective, random walks have also been used for the task of expert finding in academic search collections. For instance, Deng et al. [8] proposed a joint regularization framework to enhance expertise retrieval in academia by modeling heterogeneous networks as regularization constraints on top of a document-centric model [4]. Relatedly, Wu et al. [35] proposed to model authors and publications as nodes of a publication network, with additional edges representing co-authorship information (author-author edges). In a similar vein, Tang et al. [33] proposed a probabilistic topic modeling approach to enrich a heterogeneous graph comprising multiple academic entities as nodes, including authors, papers, and publication venues, with directed edges representing a variety of relationships such as "written by" and "published in". The stationary distribution computed after a random walk on this graph was then used to rank these entities with respect to an input query. A very similar approach was proposed by Gollapalli et al. [10], by assigning topics to nodes and then computing the unique stationary distribution of the associated Markov chain.

In contrast to the aforementioned works, we use random walks to model the *transference* of reputation from *multiple* reference sources to selected targets in a reputation graph, as discussed in Section 3. In order to validate our model, we instantiate it in an academic search setting by using research groups as reputation sources and publication venues as reputation targets. Moreover, while previous approaches have exploited multiple ranking signals, we demonstrate the power of the notion of reputation transfer by relying on publishing behavior as the only reputation signal.

3. REPUTATION FLOWS

Identifying reputable entities is an important task in many domains. While quantifying the reputation of a given entity is a challenging task, we argue that the flow of reputation among entities can be accurately modeled as a stochastic process. To this end, in this section, we propose a conceptual framework for ranking entities that interact with (and hence convey reputation to) one another in some manner. To formalize our approach, in Section 3.1, we introduce the reputation graph, a data structure that models the flow of reputation from selected sources to multiple targets. In Section 3.2, we formalize a stochastic process to estimate the amount of reputation transferred to target entities. Lastly, in Section 3.3, we discuss a simple mechanism to rank entities according to their inferred reputation.

3.1 The Reputation Graph

We define a *reputation graph* as a graph with three node types: reputation sources, reputation targets, and reputation collaterals, as illustrated in Figure 1. The reputation graph models the transference of reputation from a reference set of reputation sources to reputation targets, and then to reputation collaterals. To refer to the reputation graph, we adopt the following notation: S is the set of reputation sources, T is the set of reputation targets, and C is the set of reputation collaterals.

Figure 1: Structure of the reputation graph.

The reputation of the source nodes influences the reputation of the target nodes and the reputation of the target nodes influences the reputation of the source nodes. Notice that the reputation of the target nodes influences the reputation of the collaterals, but the reputation of the collaterals has no impact in the reputation of the sources and targets.

Given that the reputation of the collaterals has no effect on the reputation of the nodes of other types, we can split the model in two phases. In the first phase, we propagate the reputation of the sources to the targets. In the second phase, we propagate the reputation of the targets to the collaterals. These phases are discussed in the following sections.

3.2 Reputation Flows

The interaction between reputation sources and reputation targets is inspired by the notion of *eigenvalue centrality* in complex networks [26], which also provides the foundation to PageRank [18, 6]. In the reputation graph, if we consider only sources and targets, it is easy to identify reputation flows from sources to sources, from sources to targets, from targets to sources, and from targets to targets. These reputation flows can be modeled as a stochastic process as we now discuss. In particular, let $\mathbf{P}$ be a *right stochastic* matrix of size $(|S|+|T|) \times (|S|+|T|)$ with the following structure:

$$\mathbf{P} = \left[\begin{array}{c|c} (d^{\langle S \rangle}).\mathbf{P}^{\langle SS \rangle} & (1-d^{\langle S \rangle}).\mathbf{P}^{\langle ST \rangle} \\ \hline (1-d^{\langle T \rangle}).\mathbf{P}^{\langle TS \rangle} & (d^{\langle T \rangle}).\mathbf{P}^{\langle TT \rangle} \end{array} \right], \qquad (1)$$

where each quadrant represents a distinct type of reputation flow. Matrix $\mathbf{P}$ depends on the following matrices:

$\mathbf{P}^{\langle SS \rangle}$: right stochastic matrix of size $|S| \times |S|$ representing the transition probabilities between reputation sources;

$\mathbf{P}^{\langle ST \rangle}$: matrix of size $|S| \times |T|$ representing the transition probabilities from reputation sources to targets;

$\mathbf{P}^{\langle TS \rangle}$: matrix of size $|T| \times |S|$ representing the transition probabilities from reputation targets to sources;

$\mathbf{P}^{\langle TT \rangle}$: right stochastic matrix of size $|T| \times |T|$ representing the transition probabilities between reputation targets.

The parameters $d^{\langle S \rangle}$ and $d^{\langle T \rangle}$ control the relative importance of the reputation sources and targets, which are modeled in the four matrices above. Specifically, $d^{\langle S \rangle}$ is the fraction of reputation one wants to transfer between source nodes and $d^{\langle T \rangle}$ is the fraction of reputation one wants to transfer between target nodes. These are useful parameters and the ability to set them is important to calibrate the impact of different reputation flows in the final score. If we do not want to consider reputation flows between nodes of the same type, it is sufficient to set both parameters to zero. If, instead, we want to consider reputation flows between nodes of the same type, we may increase these parameters according to the desired relative importance. Note that, as the sub-matrices $\mathbf{P}^{\langle SS \rangle}$ and $\mathbf{P}^{\langle TT \rangle}$ are *right stochastic*, i.e., each of the rows of matrices $\mathbf{P}^{\langle ST \rangle}$ and $\mathbf{P}^{\langle TS \rangle}$ sums to 1, and the parameters $d^{\langle S \rangle}$ and $d^{\langle T \rangle}$ are both in the range [0,1), then $\mathbf{P}$ defines a Markov chain. Assuming that the transition matrix $\mathbf{P}$ is ergodic, we can compute the steady state probability of each node and use it as a reputation score. Specifically, we can obtain values for ranking the set of nodes by solving:

$$\boldsymbol{\gamma} = \boldsymbol{\gamma}\mathbf{P}, \tag{2}$$

where $\boldsymbol{\gamma}$ is a row matrix with $|S| + |T|$ elements, where each one represents the probability of a node in the set $S \cup T$. This system of linear equations can be easily solved by standard Markov chain techniques. Then, from Equation (2), we obtain the steady state probabilities of all nodes in $S \cup T$, a.k.a. reputation sources and reputation targets.

3.2.1 Flow Equations

We recursively define the reputation of sources in terms of the reputation of targets, and the reputation of targets in terms of the reputation of sources. Specifically, the reputation γ_s of a source s is defined as:

$$\gamma_s = \sum_{t \in T} (1 - d^{\langle T \rangle}).\mathbf{P}_{ts}{}^{\langle TS \rangle}\gamma_t + \sum_{s' \in S} (d^{\langle S \rangle}).\mathbf{P}_{s's}{}^{\langle SS \rangle}\gamma_{s'}. \tag{3}$$

In the summation, $\mathbf{P}_{ts}{}^{\langle TS \rangle}$ is the transition probability from t to s, given by $\mathbf{P}_{ts}{}^{\langle TS \rangle} = n_{ts}/n_t$, where n_{ts} is the number of edges running from t to s and n_t is the total number of edges running from t. Finally, γ_t is the reputation of target t, defined recursively as:

$$\gamma_t = \sum_{s \in S} (1 - d^{\langle S \rangle}).\mathbf{P}_{st}{}^{\langle ST \rangle}\gamma_s + \sum_{t' \in T} (d^{\langle T \rangle}).\mathbf{P}_{t't}{}^{\langle TT \rangle}\gamma_{t'}. \tag{4}$$

Similarly, in the summation, $\mathbf{P}_{st}{}^{\langle ST \rangle}$ is the transition probability from s to t, given by $\mathbf{P}_{st}{}^{\langle ST \rangle} = n_{st}/n_s$, where n_{st} is the number of edges running from s to t and n_s is the total number of edges running from s. Recall that γ_s is the reputation of source s, defined according to Equation (3).

3.2.2 Bipartite Reputation Graph

Some scenarios can be represented as a bipartite reputation graph. In these cases, the transition matrix $\mathbf{P}$ is reduced to a periodic Markov chain with the following structure:

$$\mathbf{P} = \left[\begin{array}{c|c} \mathbf{0} & \mathbf{P}^{\langle ST \rangle} \\ \hline \mathbf{P}^{\langle TS \rangle} & \mathbf{0} \end{array}\right]. \tag{5}$$

From decomposition theory [23], we can obtain values for ranking the set of reputation *sources* by solving:

$$\boldsymbol{\gamma}^{\langle S \rangle} = \boldsymbol{\gamma}^{\langle S \rangle}\mathbf{P}', \tag{6}$$

where $\mathbf{P}' = \mathbf{P}^{\langle ST \rangle} \times \mathbf{P}^{\langle TS \rangle}$ is a stochastic matrix and $\boldsymbol{\gamma}^{\langle S \rangle}$ is a row matrix with $|S|$ elements, where each one represents the probability of a node in the set S of reputation sources.

Note that matrix $\mathbf{P}'$ has dimension $|S| \times |S|$ only and can be easily solved by standard Markov chain techniques. Then, from Equation (7), we obtain the reputation of all reputation *targets* linked by the reputation sources:

$$\boldsymbol{\gamma}^{\langle T \rangle} = \boldsymbol{\gamma}^{\langle S \rangle} \times \mathbf{P}^{\langle ST \rangle}. \tag{7}$$

By modeling a scenario as a bipartite reputation graph instead of a general reputation graph, we reduce the network from a graph of size $(|S| + |T|) \times (|S| + |T|)$ to a graph of size $|S| \times |S|$, which allows us to compute the steady state probabilities much more efficiently. However, by using a bipartite graph, we are certainly losing some information, which may be critical for some applications. It is important to consider this trade-off when instantiating our framework.

3.3 Reputation-based Ranking

The steady state probability of a node can be interpreted as its relative reputation, as transferred from other nodes in the reputation graph. Thus, we can directly use the value of this probability to rank reputation sources or reputation targets. Additionally, this probability can be further propagated to nodes we want to compare, which are in the collateral set. This propagation depends on a matrix $\mathbf{P}^{\langle TC \rangle}$ of size $|T| \times |C|$ representing the transitions from reputation targets to collateral nodes. More generally, we can define the reputation score of an entity e according to:

$$\text{P-score}(e) = \begin{cases} \sum_{t \in T} \mathbf{P}_{te}{}^{\langle TC \rangle}\gamma_t & \text{if } e \in C, \\ \gamma_e & \text{otherwise,} \end{cases} \tag{8}$$

where $\mathbf{P}_{te}{}^{\langle TC \rangle}$ is the transition weight from a target node t to a collateral node $e \in C$. The P-score of all candidate entities (targets or collaterals) can then be used to produce an overall reputation-oriented ranking of these entities.

4. REPUTATION FLOWS IN ACADEMIA

In this section, we discuss the instantiation of our conceptual framework of *reputation flows* in the academic context to model the transference of reputation between authors, papers, research groups and publication venues. The relations between these scientific entities may be captured through distinct metrics and, as far as we know, the most important ones (including citation-based metrics) fit well in our conceptual framework. In particular, let us start by defining the relations between authors and papers. It is easy to identify *reputation flows* from authors to authors, from authors to papers, from papers to authors, and from papers to papers.

Each one of these *reputation flows* is associated with a specific quadrant of an Author-Paper × Author-Paper relation matrix, as illustrated in Figure 2.

$$\begin{array}{cc} & \begin{array}{cc} \scriptstyle Author & \scriptstyle Paper \end{array} \\ \begin{array}{c} \scriptstyle Author \\ \scriptstyle Paper \end{array} & \left[\begin{array}{cc} Author \rightarrow Author & Author \rightarrow Paper \\ Paper \rightarrow Author & Paper \rightarrow Paper \end{array} \right] \end{array}$$

Figure 2: Reputation flows between authors and papers.

In the first quadrant, the framework represents the reputation flow from authors to authors, which can be expressed in terms of co-authorship relations or citations from an author to another. In the second and third quadrants, the framework represents author-paper and paper-author relations, respectively. An author who publishes a paper somehow transfers its own reputation to that paper or the converse, a paper may transfer its reputation or acceptance by the community to the authors who published it. In the fourth quadrant, the framework represents the reputation flow between papers. When a paper cites another, it is somehow transferring part of its reputation to the cited paper. This last quadrant is the focus of much more attention than the other ones by the academic community. The raw number of citations among papers, as well as well known citation-based metrics such as H-Index and Impact Factor can be represented in this quadrant. Additionally, there are further indicators such as the number of downloads of a paper. It is an indicator intrinsically related to the papers and has nothing to do with the reputation flow from authors to papers. In other words, the number of downloads is a reputation flow from the audience of paper readers to the papers. These external indicators can be expressed as bias variables.

The idea of reputation flows is broad and encompasses a large amount of indicators. Here, we define a more specific concept called *publication flows* to refer to the study of reputation flows where the transference of reputation is made by using only publication volume and without using citation data. In our experiments, we study how the reputation of a reference set of research groups is propagated to the venues they publish in and to other individual researchers by applying the concept of publication flows. In this conceptual framework, publication venues are aggregations of papers and research groups are aggregations of authors, as shown in Figure 3. These aggregations are sufficient to establish core relations that allow ranking these entities.

$$\begin{array}{cc} & \begin{array}{cc} \scriptstyle Group & \scriptstyle Venue \end{array} \\ \begin{array}{c} \scriptstyle Group \\ \scriptstyle Venue \end{array} & \left[\begin{array}{cc} Group \rightarrow Group & Group \rightarrow Venue \\ Venue \rightarrow Group & Venue \rightarrow Venue \end{array} \right] \end{array}$$

Figure 3: Reputation flows between groups and venues.

4.1 Overview and Assumptions

The basic idea of the P-score metric is to associate a reputation with publication venues based on the publication patterns of a *reference set* of research groups in a given area or sub-area of knowledge. Given a pre-selected set of reference research groups, P-score associates weights with the publication venues the researchers in the reference groups publish in. Further, these weights can be used to rank other research groups or authors.

The reputation of a research group is strongly influenced by the reputation of its members, which is largely dependent on their publication records. We assume that:

1. A research group conveys reputation to a publication venue proportionally to its own reputation.
2. A publication venue conveys reputation to a research group proportionally to its own reputation.

Once a reference group is selected, the reputation of its members is transferred to the venues. Recursively, since the reputation of research groups is correlated with the reputation of the venues in which they published, the venues transfer reputation to the groups. A score for venues can then be computed by solving a system of linear equations relating publication venues and research groups in the reputation graph, as exemplified next.

4.2 A Small Example

Figure 4 shows an example with two research groups used as reputation sources, Group 1 and Group 2, and three venues used as reputation targets, venues v_1, v_2 and v_3.

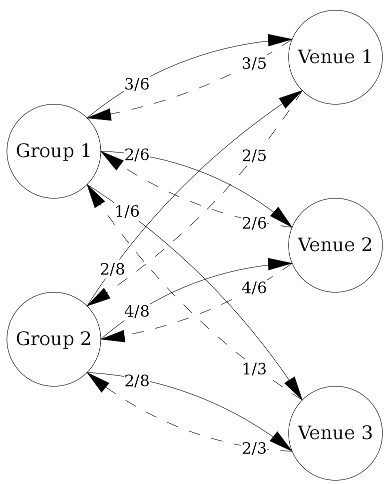

Figure 4: Markov chain for an example with 2 research groups and 3 publication venues.

From Figure 4, Group 1 published 3 papers in venue v_1, 2 papers in venue v_2, and 1 paper in venue v_3. The number of publications of Group 1 is 6. Venue v_1 receives 3 papers of Group 1, and 2 papers of Group 2. The fractions of publications from groups to venues and from venues to groups are the edge weights. We have:

$$\mathbf{P} = \left[\begin{array}{cc|ccc} 0 & 0 & 3/6 & 2/6 & 1/6 \\ 0 & 0 & 2/8 & 4/8 & 2/8 \\ \hline 3/5 & 2/5 & 0 & 0 & 0 \\ 2/6 & 4/6 & 0 & 0 & 0 \\ 1/3 & 2/3 & 0 & 0 & 0 \end{array} \right].$$

This stochastic matrix corresponds to the Markov chain displayed in Figure 4, which can be immediately aggregated to a two-state Markov chain, yielding:

$$\mathbf{P}' = \left[\begin{array}{cc} 0.467 & 0.533 \\ 0.400 & 0.600 \end{array} \right],$$

which is the stochastic matrix we use in the solution of Equation (6). (Recall that the dimension of $\mathbf{P}'$ is $T \times T$ and, as such, much smaller than that of $\mathbf{P}$ for real size problems.) Solving Equation (6) and applying Equation (7), we obtain the ranking for the three venues:

$$\boldsymbol{\nu} = \langle 0.36, 0.43, 0.21 \rangle. \tag{9}$$

Venue v_2 has the highest rank, followed by v_1, and then by v_3. We remark that the individual values give the *relative reputation* of each publication venue.

5. EXPERIMENTAL SETUP

In this section, we describe the setup that supports the empirical evaluation of our proposed model for reputation-oriented ranking, as described in Section 3. In particular, we aim to answer the following research questions:

Q1. How effective is our proposed random walk model for reputation-oriented ranking?

Q2. How robust is our model with respect to perturbations in the chosen reputation sources?

Q3. Can we alleviate the cost of manually selecting effective reputation sources within our model?

In order to evaluate our model, we consider its instantiation in an academic search setting, as described in Section 4. In the remainder of this section, we describe the academic search dataset built for our experiments, our evaluation procedure, and the baselines used in our investigation.

5.1 Academic Search Dataset

In order to evaluate our proposed model for reputation-oriented ranking, we built a test collection for two distinct academic search tasks. To this end, we focused on the tasks of ranking publication venues and individual researchers in the broad area of computer science, for which we could obtain appropriate ground-truths, as described next.

5.1.1 Ground-Truths

For the venue ranking task, we considered as ground-truth the set of venues classified by the Qualis system maintained by CAPES,[1] a government agency linked to the Brazilian Ministry of Education. CAPES assigns a committee of experts to each area of knowledge and these experts are responsible for evaluating all information acquired about the venues and produce a classification. This classification is updated annually and follows a set of criteria, such as: the number of publications in each venue, the number of repositories in which it is indexed, the amount of institutions publishing in it, citation information whenever available, among others. According to the Qualis system, the venues in each area of knowledge are classified (in decreasing order of importance) as A1, A2, B1, B2, B3, B4, B5 or C.

For the researcher ranking task, we considered as ground-truth the set of researchers with an active (as of 2014) productivity grant awarded by CNPq,[2] the Brazilian National Council for Scientific and Technological Development, as a means to stimulate excellence in research. In order to apply for a productivity grant, researchers working in Brazil must submit detailed information about their academic career to CNPq, including a research project to be conducted over the coming years. To award the grants, CNPq evaluates a set of productivity indicators including academic output, contribution to the formation of human resources, academic leadership, among others, and classifies researchers in five different levels of productivity in descending order of prestige: 1A, 1B, 1C, 1D, and 2. The starting point for any newly awarded researcher is the productivity level 2.

5.1.2 Reputation Sources, Targets, and Collaterals

In order to instantiate the reputation graph in our evaluation, as a starting point, we chose as candidate reputation sources the top 126 US computer science graduate programs (which represent research groups in our instantiation) evaluated in the 2011 assessment conducted by the US National Research Council (NRC).[3] In particular, for each of these groups, we retrieved the list of group members, which were then manually reconciled against the DBLP repository.[4]

As reputation targets, we considered all publication venues in DBLP with at least one publication by any of the aforementioned 126 candidate reputation sources and with at least one citation in Google Scholar,[5] as required by the baselines described in Section 5.3. Of these, we retained a total of 704 publication venues included in the venue ground-truth presented in Section 5.1.1, which comprise the venues to be ranked by our model and the baselines. For the researcher ranking task, we modeled individual researchers as reputation collaterals, so as to assess the quality of their ranking as induced by the selected reputation targets. To this end, we considered all computer science researchers with at least one publication in any of the previously selected publication venues and at least one citation in Google Scholar. Of these, we retained a total of 274 researchers included in the researcher ground-truth described in Section 5.1.1.

Table 1 summarizes salient statistics about our produced dataset, including the total number of reputation sources, targets, and collaterals. In addition, we describe the number of reputation targets and collaterals in each of the classes defined by our ground-truths. Table 2 shows the number of edges running across these nodes in the reputation graph.

Table 1: Salient statistics about the academic search dataset used in our experimental evaluation.

	Total	Nodes per relevance level							
		8	7	6	5	4	3	2	1
Sources	126								
Targets	704	102	115	186	110	86	91	14	0
Collaterals	274				15	17	22	52	168

Table 2: Total number of edges running across the different node types in the reputation graph.

	Sources	Collaterals
Targets	203415	15596

[1] http://www.capes.gov.br/
[2] http://www.cnpq.br/
[3] http://www.nap.edu/rdp/
[4] http://dblp.uni-trier.de/
[5] http://scholar.google.com/

5.2 Evaluation Procedure

To compare the rankings produced by P-score and the baselines described in Section 5.3 for the venue and researcher ranking tasks, we use the discounted cumulative gain (DCG) metric [15]. DCG adopts a non-binary notion of relevance, by assessing a given ranking based upon a graded scale, from less relevant to more relevant. This metric also uses a log-based discount factor that reduces the impact of the gain as we move lower in the ranking. Let l_i be the non-binary relevance level associated with the item ranked at the i-th position. The DCG at a rank position k is defined as:

$$\mathrm{DCG}@k = \sum_{i=1}^{k} \frac{2^{l_i} - 1}{\log_2(i+1)}. \tag{10}$$

To bind the results within the interval [0,1], we use the normalized version of DCG, denoted nDCG, which is obtained by dividing the DCG@k value by the maximum possible value at the same ranking cutoff k. As relevance levels, we consider a linear mapping from the classes defined by each of our ground-truths, as described in Table 3.

Table 3: Mapping from Qualis and CNPq classes (discussed in Section 5.1.1) to the relevance levels used by nDCG.

nDCG level	8	7	6	5	4	3	2	1
Qualis class	A1	A2	B1	B2	B3	B4	B5	C
CNPq class				1A	1B	1C	1D	2

5.3 Ranking Baselines

We compare P-score with two citation-based baselines, namely, a raw citation count and the well known H-Index. Our choice of these baselines is motivated by the wide adoption of citation-based metrics for assessing productivity in academia. Indeed, nowadays citation counts and H-Index are both considered standard indicators to compare publication venues or individual researchers. As discussed in Section 5.1, we collected all citation data from Google Scholar.

6. EXPERIMENTAL RESULTS

In this section, we discuss the results of the empirical validation of our random walk model for reputation-based ranking. In the following, Sections 6.1, 6.2, and 6.3 address the three research questions stated in the previous section, regarding the effectiveness and robustness of our model, as well as the selection of suitable reputation sources.

6.1 Ranking Effectiveness

In order to address research question Q1, we assess the effectiveness of P-score in contrast to existing citation-based metrics used as baselines, as discussed in Section 5.3. To this end, we analyze rankings of publication venues and individual researchers in terms of nDCG at multiple ranking cutoffs. To instantiate P-score, we consider two simple alternatives for selecting reputation sources. The first alternative (denoted P-score 10) considers only the top 10 research groups ranked by the NRC whereas the second (denoted P-score 126) considers all 126 ranked groups as reputation sources, as described in Section 5.1.2. A further alternative for semi-automatically choosing an effective reference set of reputation sources is later investigated in Section 6.3.

6.1.1 Ranking Publication Venues

To assess the effectiveness of P-score, we first consider its application for the task of ranking publication venues, which represent reputation targets in the instantiation of our model. In particular, Figure 5 shows the effectiveness of the two aforementioned P-score variants and the H-Index baseline[6] in terms of nDCG at multiple ranking cutoffs k up to the number of venues to be ranked, namely, 578.

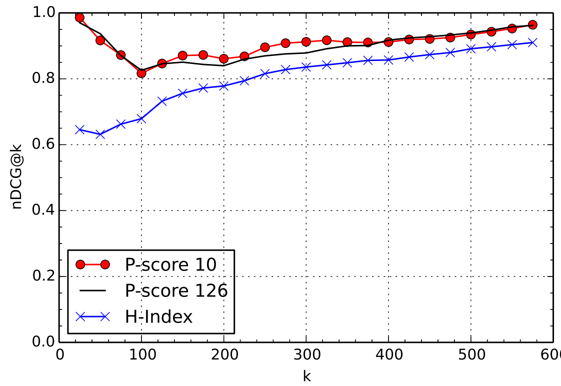

Figure 5: Venue ranking effectiveness.

From Figure 5, we first observe that the two variants of P-score, using either the top 10 or top 126 NRC groups as reputation sources, perform similarly across the entire range of nDCG@k values. This suggests that a few highly reputable sources are enough to transfer reputation to the target set of publication venues. More importantly, both variants of P-score consistently and substantially outperform the H-Index baseline for all ranking cutoffs. This is a remarkable result, given that citation information is a core element of manual academic assessments, such as the ones conducted to produce the ground-truth used in this investigation, as discussed in Section 5.1.1. In contrast, the current instantiation of our model, which solely exploits publishing behavior as a reputation signal, delivers the best ranking performance.

6.1.2 Ranking Individual Researchers

The results in Section 6.1.1 demonstrated the effectiveness of P-score for ranking reputation targets, represented by publication venues. In this section, we further assess the effectiveness of our model at transferring reputation from target nodes to collateral nodes, represented by individual researchers.[7] This evaluation is motivated by the possibility of reusing the immediate results of the random walk performed on the reputation graph to rank entities outside the initial set of reputation targets. To this end, Figure 6 contrasts the effectiveness of P-score with the Citations and H-Index baselines for the task of ranking researchers. Once again, effectiveness figures are given in terms of nDCG at multiple ranking cutoffs k, up to the total number of researchers to be ranked, namely, 274.

From Figure 6, we first note that P-score 10 once again outperforms both the Citations as well as the H-Index base-

[6]The citations baseline could not be assessed since raw venue citation information is not available from Google Scholar.

[7]A preliminary version of the experiment discussed in Section 6.1.2 appeared as a short workshop paper [1].

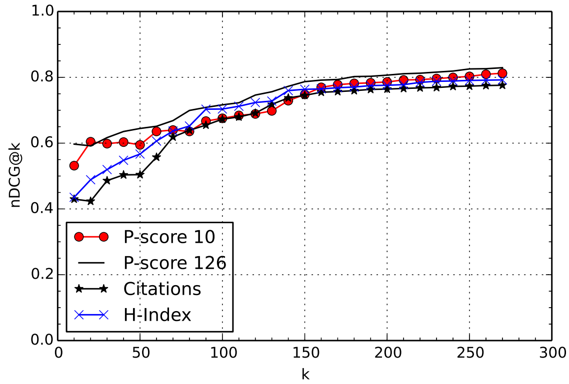

Figure 6: Researcher ranking effectiveness.

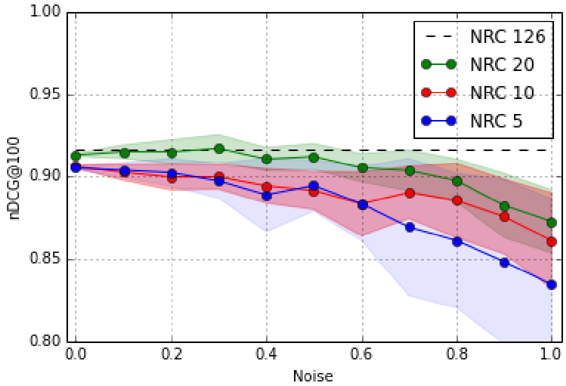

Figure 7: Venue ranking robustness with respect to random perturbations of the selected reputation sources.

lines in terms of nDCG@k for ranking cutoffs up to $k = 70$. In addition, P-score 126 is consistently the most effective of all tested approaches throughout the full range of nDCG@k values. Recalling research question Q1, the results in this and the previous section attest the effectiveness of P-score as a ranking approach for two different academic search tasks. Moreover, they show that the model can successfully transfer reputation from selected sources to both immediate targets as well as to collateral nodes in a post-hoc fashion. Lastly, the relative effectiveness between P-score 10 and P-score 126 raises an interesting observation. In particular, while a few reputation sources may perform effectively (as was the case with the results in Figure 5 for the ranking of publication venues), these sources must be carefully selected. In the next section, we will further assess the robustness of our model to perturbations in the selected reputation sources.

6.2 Ranking Robustness

The results in Section 6.1 attested the effectiveness of our proposed model for ranking publication venues and individual researchers when a careful selection of research groups are used as reputation sources. Nevertheless, this selection may eventually include noisy reputation sources, making it sub-optimal. To address research question Q2, we assess the robustness of the rankings produced by P-score with respect to random perturbations in the selected reputation sources. Figure 7 shows the results of this investigation for venue rankings. In particular, the x-axis denotes the amount of noise randomly injected into a reference set of reputation sources—in our case, the top k research groups ranked by the NRC, for $k \in \{5, 10, 20\}$. For instance, $x = 0.2$ indicates that 20% of the reputation sources are replaced by research groups randomly chosen from outside the reference set. Accordingly, $x = 0.0$ indicates no noise (i.e., the untouched top k NRC groups), whereas $x = 1.0$ indicates maximum noise (i.e., a random set of k research groups). On the y-axis, we show mean nDCG@100 figures averaged across 30 repetitions of this perturbation process, with shaded areas denoting the observed standard deviation from the mean. An additional curve including all 126 NRC groups as reputation sources is shown as a reference for comparison.

From Figure 7, we observe that larger sets of reputation sources are generally more robust to noise, as demonstrated by the green curve (NRC 20). Indeed, this setting delivers nearly the same ranking effectiveness as the one achieved when using all 126 NRC groups as reputation sources. More importantly, all venue rankings produced by our model are relatively stable up to a noise level around 0.3 (i.e., when 30% of the reputation sources are randomly chosen). Recalling question Q2, these results attest the robustness of the rankings produced by our model with respect to random perturbations in the selected reputation sources. Moreover, they open up an interesting direction towards automatically identifying a robust set of reputation sources.

6.3 Selecting Reputation Sources

The results in the previous sections demonstrated the effectiveness of our model and its robustness to random perturbations in the reference set of reputation sources. For both experiments, as discussed in Section 5.1.2, a careful selection of manually chosen reference reputation sources—the top 126 research groups ranked by the NRC—was used. In practice, such a manual selection of reputation sources can be costly. While a completely automatic alternative is beyond the scope of our current investigation, we performed an experiment to demonstrate the feasibility of semi-automatically choosing suitable reputation sources.

To address question Q3, we experiment with our proposed model itself as a means to identify effective reputation sources. Specifically, starting with all 126 groups evaluated by the NRC, we randomly choose a subset of 10 groups to use as reputation sources, with the remaining 116 groups used as reputation collaterals. After convergence, we choose the top 10 ranked collateral nodes as the new set of reputation sources. We repeat this procedure until the set of top 10 groups no longer changes. We applied this procedure 100 times to the aforementioned set of 126 groups. At each run, we repeated the selection of a new reference set of reputation sources (starting from a random selection) until the set of top 10 groups stabilized. Table 4 lists the 12 groups that appeared among the top 10 at least once in a given run, after the process stabilized. Moreover, the first 8 listed groups appeared among the top 10 at every single run.

From Table 4, we observe that all 12 listed groups are among the top 5th percentile in the official ranking produced by the NRC. Moreover, when used as reference reputation sources for ranking publication venues and individual re-

Table 4: Research groups that appeared at least once among the top 10 selected reputation sources after 100 selections that started from a random choice.

1	Carnegie Mellon University
2	Georgia Institute of Technology
3	Massachusetts Institute of Technology
4	Stanford University
5	University of California-Berkeley
6	University of California-Los Angeles
7	University of California-San Diego
8	University of Illinois at Urbana-Champaign
9	University of Maryland College Park
10	University of Southern California
11	University of Michigan-Ann Arbor
12	Cornell University

searchers, these semi-automatically selected groups yield a comparable retrieval effectiveness to the ones attained when using the top 10 groups ranked by the NRC itself as reputation sources, as previously shown in Sections 6.1.1 and 6.1.2. Recalling question Q3, these results show that it is possible to alleviate the manual effort incurred by the selection of reputation sources through a semi-automatic procedure with no penalty in effectiveness. Moreover, these results provide encouragement for further investigating fully automatic mechanisms to identify effective reputation sources.

7. CONCLUSIONS

In this paper, we have proposed a novel random walk model to identify the most reputable entities of a domain, based on a conceptual framework of reputation flows. Our model overcomes the challenges of quantifying reputation (arguably, a subjective and multi-faceted concept) by instead focusing on the transference of reputation among different entities. We instantiated our model in an academic search setting and empirically validated its effectiveness and robustness for two academic search tasks in the broad area of computer science, namely, publication venue and individual researcher ranking. Specifically, we demonstrated the effectiveness of our model in contrast to standard citation-based approaches for identifying reputable venues and researchers as well as its robustness to perturbations in the selection of reputation sources. Furthermore, we showed that effective reputation sources can be chosen in a semi-automatic fashion using our proposed random walk model itself.

Both the conceptual framework and its instantiation in an academic context open opportunities for future work. At the model level, we intend to further verify the generality of the concept of reputation flows when applied to other domains, such as enterprise search. At the instantiation level, we intend to further explore fully automatic mechanisms for identifying suitable reputation sources for academic search. In addition, we plan to test our model for academic search tasks in areas other than computer science.

8. REFERENCES

[1] S. Author. Anonymized for peer-reviewing. 2015.

[2] L. Backstrom and J. Leskovec. Supervised random walks: Predicting and recommending links in social networks. In *Proc. of WSDM*, pages 635–644, 2011.

[3] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic. The role of social networks in information diffusion. In *Proc. of WWW*, pages 519–528, 2012.

[4] K. Balog. Expertise retrieval. *Found. Trends Inf. Retr.*, 6(2-3):127–256, 2012.

[5] J. Bollen, H. van de Sompel, J. Smith, and R. Luce. Toward alternative metrics of journal impact: A comparison of download and citation data. *Information Processing & Management*, 41(6):1419–1440, 2005.

[6] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw.*, 30(1-7):107–117, 1998.

[7] G. Cormode, S. Muthukrishnan, and J. Yan. People like us: mining scholarly data for comparable researchers. In *Proc. of WWW*, pages 1227–1232, 2014.

[8] H. Deng, J. Han, M. R. Lyu, and I. King. Modeling and exploiting heterogeneous bibliographic networks for expertise ranking. In *Proc. of JCDL*, pages 71–80, 2012.

[9] E. Garfield. Citation indexes for science. *Science*, 122(3159):108–111, 1955.

[10] S. D. Gollapalli, P. Mitra, and C. L. Giles. Ranking authors in digital libraries. In *Proc. of JCDL*, pages 251–254, 2011.

[11] G. D. Gonçalves, F. Figueiredo, J. M. Almeida, and M. A. Gonçalves. Characterizing scholar popularity: A case study in the computer science research community. In *Proc. of JCDL*, pages 57–66, 2014.

[12] J. Hirsch. An index to quantify an individual's scientific research output. *Proc. Nat. Acad. Sciences*, pages 16569–16572, 2005.

[13] J. G. Hutton, M. B. Goodman, J. B. Alexander, and C. M. Genest. Reputation management: the new face of corporate public relations? *Pub. Rel. Rev.*, 27(3):247–261, 2001.

[14] M. Jamali and M. Ester. Trustwalker: A random walk model for combining trust-based and item-based recommendation. In *Proc. of SIGKDD*, pages 397–406, 2009.

[15] K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst.*, 20(4):422–446, 2002.

[16] P. Katerattanakul, B. Han, and S. Hong. Objective quality rankings of computing journals. *Commun. ACM*, 45, 2003.

[17] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632, 1999.

[18] A. LangVille and C. Meyer. *Google's PageRank and Beyond: The Science of Search Engine Rankings.* Princeton University Press, 2006.

[19] B. Larsen and P. Ingwersen. Using citations for ranking in digital libraries. In *Proc. of JCDL*, pages 370–370, 2006.

[20] L. Leydesdorff. How are new citation-based journal indicators adding to the bibliometric toolbox? *J. Am. Soc. Inf. Sci. Technol.*, 60(7):1327–1336, 2009.

[21] H. Lima, T. H. P. Silva, M. M. Moro, R. L. T. Santos, W. M. Jr., and A. H. F. Laender. Aggregating productivity indices for ranking researchers across multiple areas. In *Proc. of JCDL*, pages 97–106, 2013.

[22] G. Mann, D. Mimno, and A. McCallum. Bibliometric impact measures leveraging topic analysis. In *Proc. of JCDL*, pages 65–74, 2006.

[23] C. Meyer. Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems. *SIAM Review*, 31(2):240–272, 1989.

[24] S. Nelakuditi, C. Gray, and R. R. Choudhury. Snap judgement of publication quality: how to convince a dean that you are a good researcher. *Mobile Computing and Commun. Review*, 15(2):20–23, 2011.

[25] S. Nerur, R. Sikora, G. Mangalaraj, and V. Balijepally. Assessing the relative influence of journals in a citation network. *Commun. ACM*, 48(11):71–74, 2005.

[26] M. Newman. *Networks: An Introduction*. Oxford University Press, 2010.

[27] Z. Nie, Y. Zhang, J.-R. Wen, and W.-Y. Ma. Object-level ranking: Bringing order to web objects. In *Proc. of WWW*, pages 567–574, 2005.

[28] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. In *Proc. of WWW*, pages 161–172, 1998.

[29] H. Piwowar. Altmetrics: Value all research products. *Nature*, 493(7431):159–159, 2013.

[30] E. Rahm and A. Thor. Citation analysis of database publications. *ACM Sigmod Record*, 34(4):48–53, 2005.

[31] S. Saha, S. Saint, and D. Christakis. Impact factor: a valid measure of journal quality? *J. Med. Lib. Assoc.*, 91(1):42–46, 2003.

[32] Y. Sun and C. L. Giles. *Popularity weighted ranking for academic digital libraries*. Springer, 2007.

[33] J. Tang, R. Jin, and J. Zhang. A topic modeling approach and its integration into the random walk framework for academic search. In *Proc. of ICDM*, pages 1055–1060, 2008.

[34] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: Finding topic-sensitive influential twitterers. In *Proc. of WSDM*, pages 261–270, 2010.

[35] H. Wu, Y. Pei, and J. Yu. Detecting academic experts by topic-sensitive link analysis. *Front. Comp. Science in China*, 3(4):445–456, 2009.

[36] W. Xi, B. Zhang, Z. Chen, Y. Lu, S. Yan, W.-Y. Ma, and E. A. Fox. Link fusion: A unified link analysis framework for multi-type interrelated data objects. In *Proc. of WWW*, pages 319–327, 2004.

[37] E. Yan, Y. Ding, and C. R. Sugimoto. P-rank: An indicator measuring prestige in heterogeneous scholarly networks. *J. Am. Soc. Inf. Sci. Technol.*, 62(3):467–477, 2011.

[38] S. Yan and D. Lee. Toward alternative measures for ranking venues: a case of database research community. In *Proc. of JCDL*, pages 235–244, 2007.

[39] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proc. of SIGIR*, pages 325–334, 2011.

[40] D. Zhou, S. A. Orshanskiy, H. Zha, and C. L. Giles. Co-ranking authors and documents in a heterogeneous network. In *Proc. of ICDM*, pages 739–744, 2007.

[41] Z. Zhuang, E. Elmacioglu, D. Lee, and C. Giles. Measuring conference quality by mining program committee characteristics. In *Proc. of JCDL*, pages 225–234, 2007.