1-30-2019

Relational social recommendation: Application to the academic domain

Saeed Amal

Chun-Hua Tsai

Peter Brusilovsky

Tsvi Kuflik

Einat Minkov

Relational social recommendation: Application to the academic domain

Saeed Amal [a],Chun-Hua Tsai [b],Peter Brusilovsky [b],Tsvi Kuflik [a],Einat Minkov [a,*]

[a] *Department of Information Systems, University of Haifa, Israel*

[b] *School of Computing and Information, University of Pittsburgh, United States*

[*] Corresponding author

Keywords*:* Graph-based recommendation, Social recommendation, Recommendation explainability, Information extraction

ABSTRACT

This paper outlines RSR, a relational social recommendation approach applied to a social graph comprised of relational entity profiles. RSR uses information extraction and learning methods to obtain relational facts about persons of interest from the Web, and generates an associative entity-relation social network from their extracted personal profiles. As a case study, we consider the task of peer recommendation at scientific conferences. Given a social graph of scholars, RSR employs graph similarity measures to rank conference participants by their relatedness to a user. Unlike other recommender systems that perform social rankings, RSR provides the user with detailed supporting explanations in the form of relational connecting paths. In a set of user studies, we collected feedbacks from participants onsite of scientific conferences, pertaining to RSR quality of recommendations and explanations. The feedbacks indicate that users appreciate and benefit from RSR *explainability* features. The feedbacks further indicate on recommendation *serendipity* using RSR, having it recommend persons of interest who are not apriori known to the user, oftentimes exposing surprising inter-personal associations. Finally, we outline and assess potential gains in recommendation relevance and serendipity using path-based relational learning within RSR.

1. Introduction

The queries submitted to commercial search engines often concern *persons* and other named entities. For such queries, it is desired to present to users a meaningful summary about the entity of interest, as opposed to a ranked list of webpages (Herzig, Mika, Blanco, & Tran, 2013). As of today, Wikipedia[1] and search engines like Google[2] display factual information about entities in the form of structured *infoboxes* (Bota, Zhou, & Jose, 2015; Wu & Weld, 2008), which frequently include relations with other entities. For example, the infobox for *Albert Einstein* on Wikipedia[3] contains factual details about Einstein in relational form, concerning *birth place, attended universities, awards*, and other attributes that characterize world-class scientists. It has been shown that users find infoboxes to be informative and engaging (Bota et al., 2015). Moreover, relational entity representation supports question answering applications, making it possible to address relational queries such as ‘what are the *prizes* awarded to *Einstein*?’,(e.g., Abujabal, Roy, Yahya, & Weikum, 2017) *exploratory search* applications (e.g., Yogev, Roitman, Carmel, & Zwerdling, 2012), and more.

In this work, we describe an *entity recommendation* approach that makes use of such entity profiles. We focus our attention on *person* entities, and specifically, *scholars*. Unlike the Noble-prize winner Einstein, most scholars, some of whom are highly influential in their communities, belong to the long tail of entities that do not have a Wikipedia page nor appear in other public knowledge base (Vexler & Minkov, 2016). Our relational social recommendation method, named RSR, therefore elicits factual details about scholars of interest from relevant Web text. We describe fact extraction methodology that obtains meaningful factual personal profiles with minimal

[1] www.wikipedia.org.

[2] www.google.com.

[3] See https://en.wikipedia.org/wiki/Albert_Einstein.

human intervention. Having constructed a ‘knowledge base’ of scholar profiles, we build a joint entity-relation graph from extracted relational profiles and use it to offer relational recommendation, which is based on multiple connections between entities in the graph.

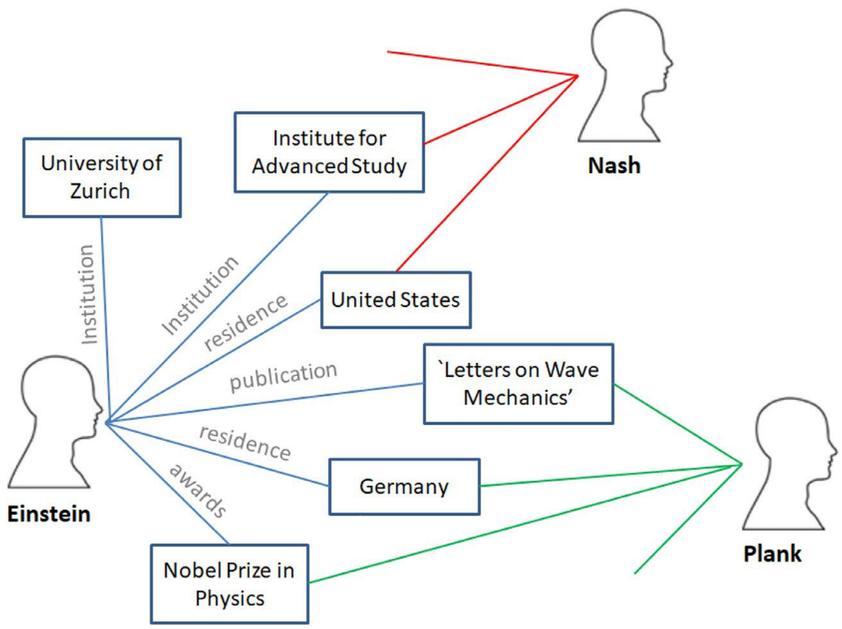


Fig. 1. A partial network of entities and their inter-relations linking famous scholars.

To illustrate our approach, we again turn to *Albert Einstein* as example. Fig. 1 displays some factual information about Einstein in a graph form. As shown, Einstein is represented as a graph node, which is linked (over edged colored in blue) to other nodes denoting related entities and concepts, such as the ‘University of Zurich’, ‘Institute for Advanced Studied’, ‘Germany’, and ‘Letters on Wave Mechanics’. The graph edges have a relation type specified, e.g., ‘Einstein’ was *an author of the publication* ‘Letters on Wave Mechanics’. Fig. 1 further illustrates how multiple personal profiles are unified to form a joint graph that comprises a heterogeneous entity-relation social network. In this graph, the association be- tween an entity pair corresponds to their *connecting paths* via related entities. As shown in the mock-up graph in Fig. 1, the scientists *Albert Einstein* and *Max Planck* connect via entities and concepts like ‘Germany’, ‘Nobel Prize in Physics’ and ‘Letters on Wave Mechanics’. The affinity between Einstein and Plank can be revealed and *explained* based on the discovered relational paths, e.g., *Einstein* $\xrightarrow{publication}$ *‘Letters on Wave Mechanics’* $\xrightarrow{publication-}$*Planck.*

RSR treats social recommendation as a query: “Who are the persons most related to person *p* in the graph?”. To address this query, it employs a well-studied random walk similarity measure, namely Personalized PageRank (Page, Brin, Motwani, & Winograd, 1999), assessing and ranking *persons* represented as graph nodes by their relatedness to *p*.

There are several advantages of RSR compared with other recommendation approaches. The first is *explainability*–in addition to relevance scores, RSR presents the relational connecting paths between the user and the recommended person, providing the user with a detailed explanation to the question,“*how* am I and the recommended entity (person) related?”. We also believe that RSR promotes *serendipity* in recommendation, in the sense that it can reveal inter-personal connections that are not yet known to the user. Finally, relational learning is applied within RSR to improve recommendation *quality* using dedicated path-based features. Some of the proposed features are serendipity-oriented, and favor the recommendation of persons whom the user is not directly related with in the graph.

In this paper, we focus on the application of RSR to recommend scholar peers to participants of academic conferences. In addition to describing the general RSR framework, we present in detail its application to social recommendation in the conference context. Recommendation is performed in this case over a social graph that unifies the graph profiles of all of the conference attendees–having the profiles of these scholars extracted by RSR for this purpose. RSR aims to reveal relevant conference participants with whom the user would be interested to meet, exchange expertise, or establish professional collaboration.

In order to assess RSR performance we presented the recommendations and relational explanations generated by RSR at several academic conferences and collected user feedbacks on site. Users generally found RSR to be engaging, sometimes surprising, and demonstrated interest in exploring the supporting evidence in the form of relational paths. Quantitative analyses of the user feedbacks indicate that RSR recommends relevant persons who are not tracked by a state-of-the-art multi-facet recommendation system designed for the scholarly domain, such as Conference Navigator (CN) (Brusilovsky, Oh, López, Parra, & Jeng, 2017).

The remaining part of the paper starts with a review of related research (Section 2). Section 3 formalizes the graph-based social recommendation approach, and discusses its advantages with respect to explainability and serendipity. We then describe in de- tail the application of RSR to the academic domain, including the automatic extraction of relational scholar profiles (Section 5). The empirical evaluation and results are described and discussed in Section 6. The paper concludes with a summary and a discussion of future research directions.

2. Related Work

There are several research areas that are closely related to our work. In this section, we review previous research on graph- based social recommendation, *explainability* and *serendipity* in social recommendation, and previous efforts targeted at social recommendation in the scholarly domain.

2.1. *Graph-based social recommendation*

2.1.1. *Relational entity representation*

Several alternatives to traditional search have been proposed in recent years in the context of *entity search*. Researchers have considered concrete entity search tasks, such as the ranking of entities in response to a topical query, having entities represented by their Wikipedia pages or homepages (Balog et al., 2012). Inspired by ‘entity cards’, or ‘infoboxes’ (Bota et al., 2015; Miliaraki, Blanco, & Lalmas, 2015; Yogev et al., 2012), we represent an entity as a set of relational facts. These facts correspond to a graph comprised of typed entities and relations. Since most entities do not have a public relational profile available (Kang et al., 2011; Miliaraki et al., 2015), we apply information extraction and learning techniques to elicit relevant facts in relational form from their personal home- pages (Chang, Kayed, & Girgis, 2006; Minkov, 2016). In particular, we describe a fact extraction procedure for scholars that involves minimal manual intervention. Having extracted the personal relational profiles, we explore *social recommendation* over a relational network in which the individual profiles are interconnected.

2.1.2. *Social graphs*

Much research exists on making personalized recommendations within existing social networks (Lee & Brusilovsky, 2018), e.g., suggesting new friendships (Backstrom & Leskovec, 2011; Leskovec, Huttenlocher, & Kleinberg, 2010), finding relevant colleagues in the enterprise (Guy, Zwerdling, Ronen, Carmel, & Uziel, 2010), or suggesting items of interest based on social proximity (Seo, Kim, Lee, & Baik, 2017). In this work, we generate social recommendations, i.e., aim at recommending people using a social graph built for this purpose (See Guy, 2018a for a recent survey on the goals and methods of people recommendation on social media.) The underlying graph consists of general facts in relational forms about persons of interest, which may be obtained from available sources or extracted from the Web. Our work is closely related to Adamic and Adar (2003) who mined student homepages to create a social network. They too extracted named entities from one’s homepage, and connected individuals based on com- mon named entities–as well as based on additional information sources like mailing lists. Our works differ however in terms of goals and methods. Adamic and Adar predicted whether a person was a friend of another based on the number of linked items that they had in common. According to them, student home- pages provide a glimpse into the social structure of university communities, and can give context to these relationships. Our focus is on recommendation as opposed to community detection. Arnet-Miner (Tang et al., 2008)[4] is another related system for extracting and mining academic social networks. They too process homepages to extract researchers' profiles. However they restrict the extracted information to pre-specified fields. ArnetMiner offers expert finding functionality, but does not incorporate personal recommendation capabilities. Unlike the abovementioned works, we construct a semantically richer graph that is comprised of associative mined relational facts and includes directed and type- inferred labeled edges, aiming at promoting explainability and serendipity in social search. We further employ relational learning methods for improving inter-personal relatedness assessment.

2.1.3. *Graph-based recommendation*

The recommendation framework applied in this work follows closely on previous works (Minkov, Kahanov, & Kuflik, 2017; Pritsker, Kuflik, & Minkov, 2017). Please refer to Minkov et al. (2017) for an elaborate review of literature on related graph-based recommendation approaches. In brief, graph-based recommendation has been shown to be advantageous in modeling and incorporating relational knowledge into the recommendation process using a heterogeneous multi-type graph scheme. We have previously shown that graph-based recommendation is competitive with classical and state-of-the-art approaches due to the modeling of background knowledge when rating history is limited (Minkov et al., 2017; Pritsker et al., 2017). For example, we have shown that in the recommendation of

[4] https://aminer.org/.

museum exhibits to visitors, the representation of exhibits' *themes* and *locations* as graph nodes served to model physical and semantic proximity, leading to performance gains (Minkov et al., 2017). In another work, the modeling of personality traits and user interests in the graph improved recommendation in cold start conditions (Pritsker et al., 2017). Another recent work of interest (Chaudhari, Azaria, & Mitchell, 2017) successfully applied a similar knowledge-rich graph-based recommendation framework, linking users and items to related entities in a large knowledge graph. Compared with previous works on graph-based recommendation, we believe that this work is first to construct and use a graph that represent associative facts about users for social recommendation purposes. Our main emphasis is on improving *explainability* and *serendipity* (Tsurel, Pelleg, Guy, & Shahaf, 2017) of recommendation, by re- covering and presenting connecting relational paths as supporting evidence to users. We further apply relational learning based on user feedbacks (Lao, Minkov, & Cohen, 2016; Minkov & Cohen, 2010) to improve these aspects, proposing a set of task-specific learning features, some of which are serendipity-oriented.

2.2. *Explainability for Social Recommendations*

Providing explanations has been proved useful in improving transparency and trust in a recommender system, contributing to the user experience (Pu & Chen, 2007; Tintarev & Masthoff, 2015). In fact, it has been shown that when a recommender system has low interpretability, the user tends to select a narrow set of the top recommended items leading to a possible lack of diversity in information exposure and selection (Graells-Garrido, Lalmas, & Baeza- Yates, 2016; Pariser, 2011). In a social recommender system with low explainability, users may “trap” themselves in a social circle they are already familiar with, missing the prospect of making new social connections.

Diverse relational data may be fused to generate high-quality user-item relevance scores using matrix factorization techniques (Yu et al., 2014; Zhao, Yao, Li, Song, & Lee, 2017). But, these methods fail to provide clear explanations for the generated recommendations. A few works attempted to address the explainability issue using a graph-based recommendation approach. For ex- ample, He, Chen, Kan, and Chen (2015) introduced a tripartite graph encoding user-item-aspect relationships for a review-aware recommendation. They explained recommendations by the "specificity" of the aspects (key phases in users' reviews), and let users match their personal aspect preference along this dimension. Ji and Shen (2016) adopted a somewhat similar tag-based explaining approach in graph-based recommender. These works revealed little on the relational association between the user and the recommended items. Another study (Heckel, Vlachos, Parnell, & Dünner, 2017) proposed a co-clustering approach to gain explainability in a user-item bipartite network, letting the user inspect the recommended items through a short description of the user-item purchase history. Their approach cannot accommodate diverse aspects in a heterogeneous graph.

In our work, we gain explainability using graph-based recommendation in the form of weighted and labeled relational connecting paths. A somewhat similar solution has been recently introduced by Wang et al. (2018). They present user preferences on a knowledge graph, and assess user-item relatedness probability scores based on indirect connecting paths. In their framework, a recommended movie can be supported by a set pre-defined explanations such as *your friend also watched this movie* or "*this movie was directed by your favorite director*". Some other works have displayed such pre-defined explanations to users using textual templates (e.g., Sánchez, Sauer, Recio-García, & Díaz-Agudo, 2017).

We are rather interested in revealing relational associations to the user in a free and associative manner.

Here, we extend graph-based explanations to the full and weighted set of relational paths that connect the user and recommended item. We incorporated these relational explanations in a user interface of a social recommendation system, and report user comments collected in two user studies. This work therefore makes additional contribution in addressing the challenge of gaining explainability in graph-based recommender systems.

2.3. *Serendipity in social recommendation*

As recommender systems strive for accurate recommendations, this has a side effect of recommending “more of the same”. Hence the need for going “beyond accuracy” attracts a lot of research attention recently, as reviewed by Kaminskas and Bridge (2016). In their review, the authors point to the most discussed beyond- accuracy objectives in recommender systems research, including diversity, serendipity, novelty, and coverage. They present a measure for serendipity and several works in recommender systems that focused on this aspect specifically. The authors further note that serendipity is a bit neglected, as it is hard to explain to users and since it is somewhat similar to novelty. While novelty pertains to the recommendations of items that are yet unknown to the user, serendipity implies that the recommendations involve some element of surprise. In this work, we evaluate the extent to which RSR recommends researchers who are relevant and still not personally known to the user. We aim at serendipity as opposed to novelty, as it is reasonable to assume that certain acquaintance exists between members of the community. We also show that the display of relational connecting paths is sometimes surprising, possibly indicating on non-trivial or unknown inter-personal associations. The importance of serendipity is indicated by a study cited by Kaminskas and Bridge which found that “interestingly, despite providing less enjoyable recommendations, the serendipity-enhancing system version was preferred over the baseline system as the users were willing to sacrifice recommendation accuracy for the sake of discovering new interesting artists”. Serendipity was investigated by McCay-Peet and Toms (2015) who defined a single model of the process of serendipity, consisting of: Trigger, Connection, Follow-up, and Valuable Outcome, and an Unexpected Thread.

Considering the setting of our study, as we see later, these elements exist within RSR and provide the potential to provide an interesting service to the user.

2.4. *Recommendations in the academic domain*

There exist abundant literature on recommendation in the scholarly domain, aiming to facilitate and accelerate the process of information seeking in digital libraries (Fuhr et al., 2007). It has been shown that information seeking in digital libraries may be enhanced by incorporating the Semantic Web and social net- working technologies, e.g., Kruk, Kruk, and Stankiewicz (2009). Indeed, multiple academic social networking sites (ASNSs) have evolved over the recent years. Mendeley,[5] Zotero,[6] and CiteULike[7] are citation management products, which also have social networking functionality, allowing users to find and follow each other. ResearchGate[8] and Academia.edu[9] are primarily social networking sites (Willinsky, 2006). Several works review these platforms and the functionalities that they provide (Bhardwaj, 2017; Gasparyan et al., 2017). As of today, however, no single bibliographic database or scholarly networking platform provides perfect cover- age of scholarly information (Bhardwaj, 2017; Ortega, 2015).

Several researchers aimed to perform social recommendation using information that is available in ASNSs and other resources. It was pointed out that the recommendation of resources that are complementary to the researcher's information needs in digital libraries may aid in discovering multdisciplinar collaboration possibilities (Porcel, Moreno, & Herrera-Viedma, 2009). Serrano-Guerrero, Herrera-Viedma, Olivas, Cerezo, and

---

[5] https://www.mendeley.com.

[6] https://www.zotero.org.

[7] http://www.citeulike.org.

[8] www.researchgate.com.

[9] www.academia.edu.

Romero (2011) proposed a recommender system adapted to Google Wave,[10] an application for real-time collaboration, in- tended to encourage and facilitate possible collaborations between multdisciplinar researchers, and recommend complementary resources useful for the interaction. Chen, Tang, Li, Mao, and Xiao (2013) recommended researchers within ASNSs based on the similarity of their research fields, and Amini, Ibrahim, Othman, and Selamat (2014) inferred scholars' research interests from mediated profiles reproduced by multiple bibliographic databases, and textual analysis of their academic homepages. Heck (2013) ad- dressed the task of researcher recommendation based on author co-citation analysis and bibliographic coupling of authors collected from ASNSs, combined with bookmarks and tags information available on the social bookmarking service CiteULike. Neither of these works provided explanations to the users. In this work, we chose to use researchers' personal webpages as a main information as we empirically found that this yields excellent coverage, with webpages providing diverse personal information. Nevertheless, the outlined graph scheme may represent and integrate relational information from multiple sources, including the various ASNSs, and general social media sites.

In this study, we consider social recommendation onsite of academic conferences as a case study of interest – a contextual recommendation scenario subject to a concrete environment con- text (Champiri, Shahamiri, & Salim, 2015). Academic conferences often include hundreds of authors and attendees, where this causes social information overload (Guy, 2015; 2018b). A useful social recommender system should assist the user by facilitating social interactions in such venues. We compare social recommendation using RSR with Conference Navigator, an online

[10] https://sites.google.com/a/pressatgoogle.com/googlewave/; this product by Google is no longer available.

‘conference support’ system, which has been developed to a third version (CN3) (Lee & Brusilovsky, 2014). The Conference navigator system is an academic tool developed at University of Pittsburgh. It is a testbed for experiments for social recommendation algorithms and interfaces (Parra & Brusilovsky, 2015; Tsai & Brusilovsky, 2017; Verbert, Parra, Brusilovsky, & Duval, 2013), and as such has been implemented and evaluated in over 30 academic conferences so far.[11] The CN system consists of a set of tools intended to help conference attendees in browsing and exploring the conference programs, publications, authors and attendees. It provides social recommendations to users based on the conference proceeding, academic publications data and social network information (avail- able from arnetMiner and citeULike), including paper bookmarks and user-to-user links. Concretely, the CN system uses four separate recommender engines, described as *Academic, Social, Interest* and *Distance* features (Tsai & Brusilovsky, 2018). The *Academic feature* computes text relevance between a pair of scholars, measured based on the content of their academic publications. The *Social feature* considers network proximity (distance and number of shared neighbors) between the two scholars in a co-authorship network. The *Interest feature* is determined by the number of this random walk procedure is defined recursively as: common papers bookmarked, as well as the common number of users followed within CN by the two scholars. Finally, the *Distance*

$$V_{d+1} = (1 - \alpha)V_q + \alpha \mathbf{M} V_d \quad (1)$$

feature corresponds to geographic distance, calculated based on the geographic coordinates of the institutions that the scholars are affiliated with. CN generates a ranking of recommended scholars ac- cording to the summation of the relevance scores produced by the four measures, allowing the user to manually set the weighting of each feature according to her

[11] http://halley.exp.sis.pitt.edu/cn3. The tool is available by request.

preferences. We compare the recommendation by RSR to those generated by CN3, and show that RSR brings to the user's attention many new and relevant connections that are not revealed by any of the recommender engines used by CN3. Unlike RSR, CN does not provides explanations to the user.

3. Relational social recommendation (RSR)

We assume that *relational profiles* for persons of interest are available, or can be obtained. The profile of a *person p* corresponds to a set of facts in the form of triplets {*p, r, e*}, having a triplet denote a relation of type *r* between *p* and a related entity *e*. The entity *e* may be assigned a semantic class (e.g., *person, organization*), following some set of semantic definitions of choice *R* (Ling & Weld, 2012). In particular, *p* may correspond to some subtype of *person*, e.g., *student, artist* etc. We will later focus our attention on generating factual profiles for *scholars*.

The set of relational facts is represented as a star-shaped graph. Concretely, the focus person *p* and each unique entity name *e* are denoted by distinct graph nodes,[12] where an outgoing edge links node *p* with each related entity *e* over edge labeled with the respective relation $r \in R$.

Having obtained multiple personal profiles, they are joined into a relational *social graph*, as illustrated in Fig. 1. This social graph is heterogeneous and typed. In order to make the graph well- connected, for every linked node pair, we add another edge in the inverse direction, typed with an 'inverse' relation label.[13]

The social recommendation task is defined as follows. Given some person node *p*, which is represented in the graph, the goal is to rank the other *person* nodes by their graph-based similarity to *p*. Importantly, we further address the complimentary question, "*how* are *p* and a

---

[12] Ideally, coreferent named entity mentions should be unified (Minkov, 2016).

recommended *person q* related?". In the rest of this section, we describe the RSR framework in detail, and discuss its advantages with respect to explainability and serendipity.

3.1. *Graph-based entity similarity*

There exist various graph-based measures that one may apply to assess similarity, or relatedness between graph nodes (Libennowell & Kleinberg, 2007). Following previous work on graph- based relational recommendation (Minkov et al., 2017; Pritsker et al., 2017), we apply the Personalized PageRank (PPR) random walk based measure (Page et al., 1999; Richardson & Domingos, 2002) for this purpose.

A detailed description of PPR is available elsewhere, e.g., Minkov and Cohen (2010). In brief, this measure applies a Markovian random walk process which consists of two operations: at each time step, the random walker either chooses with probability α to move from the current node i over an outgoing link to a neighboring node j, or the walker chooses with probability $1 - \alpha$ to 'jump' to some random graph node. The probability distribution of finding the walker at each of the graph nodes at time d, V_d, using where the transition matrix **M** encodes the probability that the walker move to any neighboring node j from node i following an outgoing edge. As default, **M** assumes uniform distribution over the all of the node's direct neighbors.

Rather than assume that the random walker resets to any graph node uniformly at random, the *personalized PageRank* random walk scheme limits the reset operation to those graph nodes which are known to be of interest to the walker, V_q. The PPR random walk process is guaranteed to

[13] We simply add a suffix 'inv' to the original relation name; e.g. if node x links to node y over *employment* relation, we add an inverse edge from y to x labeled as *employment.inv*.

converge to a unique stationary distribution, V^*, where the PPR score of node j, p_j, denotes its probability in the stationary state distribution V^*. Due to an exponential decay over walk length, the infinite graph walk process can be approximated by graph walks for a finite number of steps k (Fogaras, Rácz, Csalogány, & Sarlós, 2005; Minkov & Cohen, 2010; Toutanova, Manning, & Ng, 2004).

Importantly, PPR preserves an association between the computed node scores and V_q. In general, graph nodes that are connected over short paths to the query nodes are considered more relevant by the PPR method; similarly, nodes that are reached over multiple paths from the query are also considered more relevant. In summary, the computed PPR node scores are query-specific, reflecting structural similarity, or relevancy, with respect to a query V_q.

3.2. *Ranking generation and path recovery*

In our setting, the query distribution consists of a single node, representing the user. Having computed V^*, we rank all of the graph nodes typed as *person* by their PPR scores. In addition to these relevance scores, for each recommended *person*, supporting explanation is generated in the form of the *relational paths* linking the user to that person.

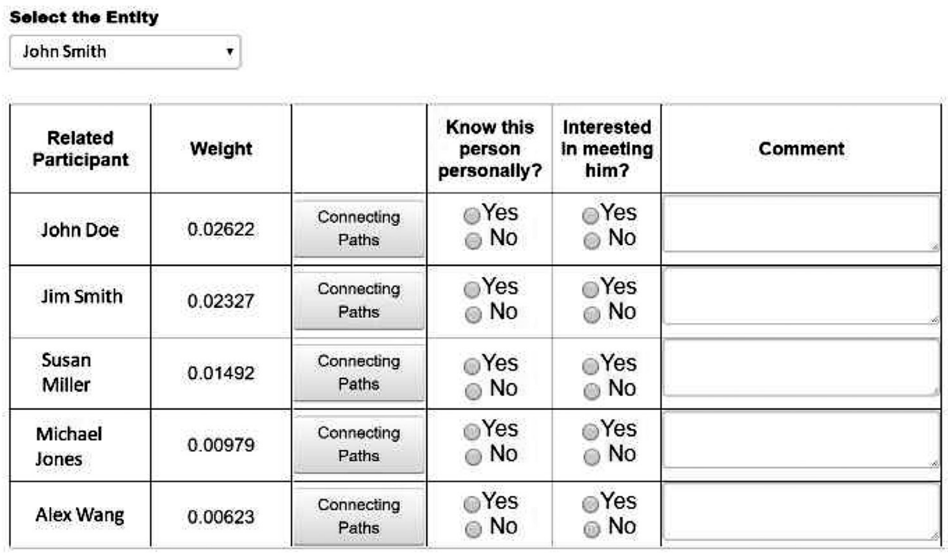
Formally, let us establish that the association between a query and target node pair corresponds to the paths over which the entities connect in the graph, having a *path* denote a sequence of labeled *entities* and *relations*. It has been shown that the PPR score for a target node t and a query node q equals a summation over all the paths leading from q to t (including cyclic paths), weighted by path traversal probabilities (Fogaras et al., 2005; Jeh & Widom, 2003). In other words, the PPR relatedness score between a node pair $<q,t>$ equals the summation of the weights of the individual paths that connect q with t over the random walk process. As we approximate PPR with finite graph walk of k steps, we un- cover the set of connecting paths between nodes $<q,t>$ up to length k. A procedure for efficiently extracting these connecting paths along with their probabilities using bi-directional search is described elsewhere (Lao et al., 2016).

Notably, given user feedbacks about the usefulness of recommended items (persons), it is possible to tune the personalized random walk process using learning (Lao et al., 2016; Minkov & Cohen, 2010). We propose a set of features designed for the social recommendation task that describe the target node *t* based on the set of connecting paths leading to it from the query node *q* and describe preliminary results of re-ranking the graph walk results using these features in Section 8.

4. Enhancing user experience with explanations

Having retrieved multiple inter-personal relationships from the graph, we wish to communicate this information to the user effectively, so as to gain *explainability* in social recommendation. There are generally multiple objectives in explaining a social recommender system. An explainable recommender system can enhance the user perception of transparency, trust and satisfaction (Tintarev & Masthoff, 2015). Different design principles have been suggested to facilitate new user interaction patterns for the purposes of breaking the 'filter bubble' (Liao & Fu, 2013), diversifying social recommendations (Tsai & Brusilovsky, 2018), or conducting a "serendipitous discovery" (Zhang, Séaghdha, Quercia, & Jambor, 2012). Here, we argue that gaining explainability in the graph-based recommender will also increase *serendipity* – the probability of finding valuable social connections which are not only "novel" but also "surprising" (Zhang et al., 2012). That is, the explainable interface should encourage the user to explore social recommendations outside of her existing social network.

The presentation of full path information for all of the recommended ranked items may be overwhelming. We therefore allow the user to view relational evidence a selected ranked item of interest, on demand. Fig. 2(a)



(a) A ranked list of relational social recommendation

(b) Path-based explanation

Fig. 2. An illustration of the user interface implemented for the user studies: the general screen displays a ranked and weighted recommendation list, and allows the user to view the paths that connect her to every recommended person on-demand using the 'Connecting Path' button. This screen also accommodates structured and free form user feedbacks (top). The path-based explanation displayed for the scholar 'Michael Jones' are authentic and have been anonymized (bottom).

displays an example of a ranked list generated by RSR for a sample conference participant, having the user and recommended scholar names anonymized and replaced with common names. Using this interface, the user can inspect the recommended persons' names alongside the weight of the recommendations (the computed PPR scores). The user can ask for an explanation for one recommended person of choice at a time by clicking on the 'Connecting Paths' button. The two 'yes/no' functions are intended for collecting user feedbacks.[14]

The path-based explanation is illustrated in Fig. 2(b). In this case, the social recommendation was explained by the set of 2-hop connecting paths between the user 'John Smith' and a person on his recommendation list, whom we name as 'Michael Jones'. The paths are listed in the order of their contribution to total PPR score. As shown, there are five paths connecting the two persons in the graph. According to this evidence, both persons are associated with the *University of Udine* (first path), as well as with the *University of Torino*. While no semantic type was assigned to the relation by the classifier in the first case, the relation with the *University of Torino* is labeled as *education*. The second and fourth paths traverse *per- son* entities – denoting researchers that both persons are linked to as part of their *publications* or *employment* history. Either of these facts may not be known apriori to the user, and ignite interest and a mutual conversation.

The lowest-weighting path shown traverses the general term 'IOS Press'. This fact seems trivial in the context of a scientific conference, as many participants are likely to have interacted with this publisher. Indeed, the PPR random walk scheme assigned low probability score to this path, due to the popular publisher node being linked to a large number of

[14] We collected such feedbacks, and used them for system evaluation and learning; a real-world system may similarly benefit from collecting feedbacks.

researcher nodes in the graph (see Section 3.1). In contrast, having only a few researchers connected with *University of Udine*, this atypical and perhaps surprising fact was assigned the highest weighting evidence score by PPR for this person pair. Thus, the ranking of the relational paths by path weight often aligns with serendipity.

In our study, we chose to present all of the connecting paths to the user. However, it is sensible to present a narrower set of significant paths by applying a threshold over path weight. In addition, it is straightforward to expand path information with avail- able context, e.g., a paragraph from which the relational information was extracted from. These design choices may be tuned so that the user gains useful information, find the explanation convincing and gain trust in the system.

5. The application of RSR to the academic domain

So far, we assumed that relational entity profiles were avail- able. It is possible that relevant personal information is available in structured form in social media profiles, but these lack coverage, and are typically include limited personal information. The application of RSR to the academic domain therefore required the *elicitation* of researcher profiles – given a target person *p* represented by her name, we first automatically extract relational facts about that individual and then proceed to construct her graph profile G_p. We chose in this research to extract relevant facts about scholars from their personal webpages; ideally, these pages would include non-trivial and potentially ‘surprising’ facts.

This section describes in detail the use of learning and information extraction methodologies towards the automatic identification of researcher homepages, and the extraction of relational facts from this source. While is it assumed that researchers’ home- pages are available in the English language, the approach is generally language-independent.

5.1. *Homepage identification*

Ideally, a researcher would be assigned a unique identifier and associated with a detailed researcher profile, including her personal webpage. While endeavors like ResearcherId and the Open Researcher and Contributor Identification (ORCID) aim to achieve just that, much like ASNSs, they lack coverage (Gasparyan et al., 2017; Martín-Martín, Ordunaalea, Thelwall, & López-Cózar, 2018). We therefore tackle the task of identifying researchers' homepages on the Web.

Some previous works used information from personal web- pages, provided that these pages or the respective Web directory were manually specified (Adamic & Adar, 2003; Das, Giles, Mitra, & Caragea, 2011; Das Gollapalli, Caragea, Mitra, & Lee Giles, 2015). We are interested in a scalable approach that involves minimal hu- man intervention, and therefore identify the relevant homepages mostly automatically.

Similarly to arnetMiner (Tang et al., 2008), we take a retrieval- based approach to homepage finding. Given a scholar name, we submit a query with the name string to a search engine.[15] The re- searcher homepage, if one exists, may not be the top result, but is typically included among the top *k* retrieved webpages. This narrows the homepage finding task to binary classification, namely, determine for each of the top *k* retrieved results whether it is in- deed a homepage of the specific person.

We addressed this task using supervised learning. Having annotated the top *k*=7 webpages retrieved for 100 random researcher names as relevant homepages or not, we employed the Weka learning suite (Hall et al., 2009) to train a classifier using these (overall, 700) labeled examples. We used lexical features to represent a retrieved webpage, including the bag of words included in the webpage's title, URL, and the captions of images on the page, as well as binary features indicating whether the specified research name appeared in any of these sections.

[15] We used Bing search engine: https://www.bing.com.

Cross validation evaluation using a decision tree (J48) classifier yielded precision of 0.70 and recall of 0.94. In general, classification performance can be improved using additional annotated data, alternative classifiers and enhanced features. It is further recommended to add the researcher's affiliation to the query as well as features in order to avoid ambiguity issues. This was not the focus of our research however, and we leave this for future work. In our study, we rather chose to use this high-recall classifier in a semi-supervised fashion, manually validating the relevance of those webpages predicted as homepages, thus achieving nearly perfect homepage detection performance at affordable annotation cost.

5.2. *Fact extraction*

Having identified the homepage(s) of person *p*, we wish to ex- tract meaningful facts from this source in the form of relational triplets {*p, r, e*}. In this step we diverge from past work (Tang et al., 2008), which aimed to extract a pre-specified set of researcher attributes. The task of general fact extraction involves main two steps, namely, named entity recognition and relation extraction.

5.2.1. *Named entity recognition*

We use two strategies for identifying a set of named entities {*e*} that are mentioned on the researcher's homepage. Since entity name mentions are often tagged with hyperlinks (Blanco et al., 2011), we consider all hyperlinked texts as related entity names. In addition, we apply the Stanford named entity tag- ger (Finkel, Grenager, & Manning, 2005)[16] to identify *person, location* and *organization* entity names mentioned in free form on the homepage. Alternatively, other named entity taggers that use a wider set of named entity types may be preferred, e.g., Ling and Weld (2012). We note that these named entity tagging tools were pre-trained on general English text and may be adapted to the scholarly domain as well as across languages (Al-Rfou, Kulkarni, Perozzi, & Skiena, 2015). Such extension is out of the scope of this work however.

5.2.2. *Relation extraction*

The type of relation that holds between the target person *p* and each entity *e* must be inferred based on the context in which *e* is mentioned on the personal webpage. Consider, for example, the sentence "graduated from *university of Zurich* in 2008", where the 'University of Zurich' has been identified as a named entity of type *organization*. In this case, the phrase 'graduated from' indicates on the relation types between the two entities. In order to simplify the representation scheme, as well as support relational learning (see Section 8), we chose to constrain the assigned relations to a small set of pre-defined types (Minkov, 2016), R = {*education, employment, publications, other*}. Again, we addressed relation prediction as a supervised classification task: given a named entity mention *e* and local context represented as the surrounding words, the respective relation has to be predicted out of the target semantic types, $r \in R$.

In order to avoid costly manual annotation of a large number of relation instances, we devised a domain-specific *distant supervision* (Mintz, Bills, Snow, & Jurafsky, 2009) labeling scheme for this purpose. Leveraging the fact that scholars often make their Curriculum Vitae (CV) available on their webpage, we sought such matching CV document and homepage pairs. The relation type linking to each named entity mention *e* found on the homepage was then automatically determined based on the section in which it was mentioned on the CV document. For example, let us assume that the name "University of Zurich" was detected on the scholar's CV in a section titled as "Academic education". Using a set of manually-crafted mapping rules, the relation was mapped based on this title to the generic relation type *education*. Likewise, co-author names and venues, which appeared in CV sections de- scribing 'publication history' were assigned the

[16] http://nlp.stanford.edu/software/CRF-NER.shtml.

relation type *publications*. Overall, we identified about 300 matching homepage and CV pairs of scholars across academic institutions and countries. The resulting automatically-annotated dataset included 605 entity names assigned the relation type *employment*; 581 entities annotated as *publications*; and 188 entity names–as *education*. For each relation type, we used a similar number of entities that were as- signed with another type as negative examples.

In relation classification, each named entity e was represented by the terms that comprised the entity name, as well as a context window of five tokens before and after the name mention. The results of 10-fold evaluation were evaluated for several classifiers (including J48 decision tree, logistic regression, SVM and Naive Bayes) using the Weka learning suite. The best results were achieved using a Naive Bayes classifier, measuring 0.87 and 0.82 in precision and recall, respectively. This classifier was then applied to predict the relation types applying to the extracted named entities based on the available context, with no additional manual intervention.

6. Onsite evaluation

Henceforth, we describe the results of social recommendation experiments conducted onsite of two International conferences. Conference participants who took part in the study were presented with personalized social recommendations generated by RSR, and provided us with their detailed feedbacks and comments. We evaluate RSR performance based on these user feedbacks. In addition to relevancy, we pay attention to the perceived serendipity of the recommendations, comparing RSR with Conference Navigator, a multi-facet scholarly recommendation system.

6.1. *Experimental setup*

In our user study, we presented personalized recommendations to participants at two International conferences, namely the Conference on User Modeling, Adaptation and Personalization (UMAP) and the ACM Conference on Hypertext and Social Media (Hyper- text), both taking place in 2017. Given the names and affiliation information of registered participants,[17] the personal webpages of these scholars were identified using the procedure described in Section 5.1. Overall, we retrieved the homepages of 159 and 77 scholars, respectively. A manual examination for a sample of 55 scholars out of this list revealed that while ResearchGate profiles existed for all of them, none of those profiles had the researcher's personal website specified; and, Only 24 of these scholars had OR- CID profile. Nevertheless, using the semi-automatic procedure we successfully tracked the relevant homepages, with only a handful of exceptions. In those cases, a highly common scholar name resulted in ambiguous and noisy search results. Adding the person's affiliation to the query resolved this issue. Consequently, fact ex- traction and generation of personal relational profiles took place. If more than one personal website was identified for a scholar, we used all of them in profile construction.

It is generally desired that the social graph be dense, encapsulating rich inter-entity associations. We therefore incorporated into the social graph roughly 800 additional profiles of scholars who took part in the International Conference on Intelligent User Inter- faces (IUI) in 2015 and 2016 as either authors or committee members. We believe that the considered conferences represent inter- leaved research communities. The resultant social graph includes about 1000 personal profiles of scholars who belong to at least one of these communities.

Table 1 includes detailed statistics of the experimental graph. This graph is considered compact and sparse, and can be fully stored and manipulated in memory of standard PC. For each conference participant, a ranked list of recommended persons was created and stored in advance, with average ranking computation time measuring several seconds per

[17] We received this information by courtesy of the conference organizers.

participant. We note there exist a plethora of efficient and scalable implementations of PPR, which can support fast node ranking computation at scale, and in ‘real time’.

Table 1

Statistics of the experimental social graph: number of nodes and edges by type.

	Type	Count
Nodes	Person	17K
	Organization	14K
	Location	3K
Edges	Publication	78K
	Employment	9K
	Education	1K
	Other	5K

6.2. *Feedback elicitation*

We have implemented a prototype interface of RSR which, given the name of a conference participant, displays a personalized ranked list of other conference participants who might be of interest to her. As shown in Fig. 2, this interface allows the user to browse the relational connecting paths between her and each recommended person, upon request. The participants in our study were requested to examine the relevancy and novelty of the dis- played relational evidence and determine whether they were interested in meeting the recommended person based on this information.

It is likely that some of the proposed contacts correspond to existing acquaintances of the user, where ideally the system would bring to her attention new potential social ties. In order to distinguish between these two possibilities, we collected detailed feed- back from the users, (Fig. 2) indicating whether a given recommended person is: (a) relevant, in the sense that the user finds interest in meeting that person, and (b) already known to the user. Finally, the interface gives opportunity to enter comments in free form per recommendation. We note that the collected user feedbacks were based solely on the information provided within RSR, and the user's own knowledge. That is, the participants in the study did not query the Web or other sources for additional in- formation.

The participation in the study was on a voluntary basis, with no award offered. Due to privacy considerations, the user feedbacks have been anonymized and aggregated for the purposes of this study.[18]

6.3. Feedback statistics

Overall, 94 distinct users enrolled in the study and provided us with a total of 1227 feedbacks. Table 2 presents summary statistics of the collected feedbacks. As indicated in the table, the recommendation lists varied in size: for about 20% of the users participating in our study, the recommendation list included between 1 and 5 recommendations. In contrast, in some of the cases (3%), the list included more than 100 persons. In general, the number of recommendations generated depends on the user's connectivity in the graph, which is in turn affected by the wealth of facts extracted from the user's homepage and their affinity to other scholar profiles. The median list length was in the range of 16–20. Table 2 further shows that users were generally engaged, providing us with a large number of feedbacks each. Specifically, for short lists of up to 5 persons, we received feedbacks for all of them; users introduced with longer lists (15–50 persons) provided us with roughly 15 feedbacks each. Otherwise, users who were presented with very long lists (> 50 persons) delivered more than 20 feedbacks on average.

[18] This study was conducted before the General Data Protection Regulation took effect, yet it is compliant with the GDPR requirements.

Table 2

Statistics of the collected feedbacks the average number and ratio (out of all recommendations) of feedbacks obtained for recommendation lists of various lengths.

List size	Users		Feedbacks	
	Count	(Ratio %)	Avg. count	Avg. ratio
1-5	18	(19)	2.7	1.00
6-10	12	(13)	6.8	0.97
11-15	12	(13)	12.2	0.90
16-20	14	(15)	14.1	0.80
21-30	9	(10)	15.2	0.63
31-40	9	(10)	15.7	0.45
41-50	4	(4)	16.8	0.39
51-70	8	(9)	24.2	0.31
71-100	5	(5)	19.2	0.22
101-145	3	(3)	22.7	0.18
All	94	(100)		

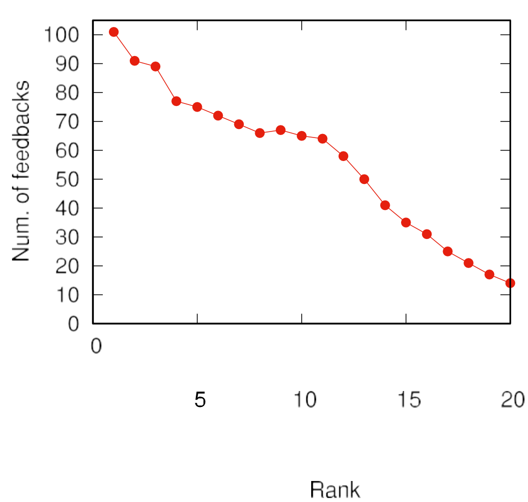


Fig. 3. Number of feedback obtained by rank: users tend to focus on the top of the ranked lists.

Fig. 3 shows the total number of feedbacks obtained per rank of the recommendation lists. Due to bias of user attention towards the top ranked items (see also Minkov et al., 2017) and the variance of list length, the total number of elicited feedbacks decreases for lower-ranked recommended persons. Accordingly, in our evaluation of RSR below, we focus on assessing recommendation performance for the top ranked scholars.

7. Results

There are several aspects of performance that we consider in evaluating RSR. The first aspect is *relevance*. It is generally desired that persons who are relevant to the user is assigned higher ranks than those judged as irrelevant. Another aspect of ranking quality is *serendipity*.

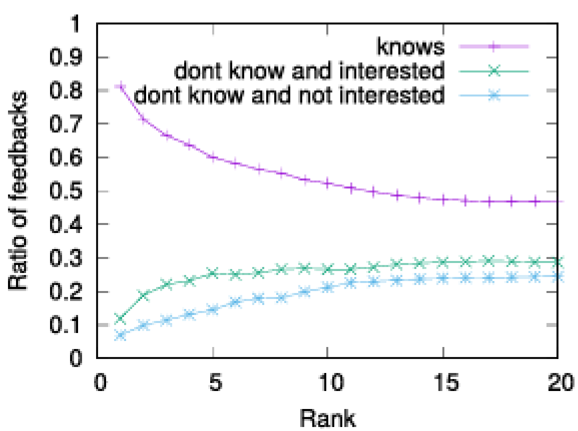


Fig. 4. The distribution of collected feedback types by rank: number of persons whom the user was previously familiar with ('knows'), or otherwise, interested – or not interested – in meeting with.

According to comments received in a preliminary study (Amal, Kuflik, & Minkov, 2017), users expect known acquaintances to be ranked high in the recommendation list, where this serves to build user trust in the system. Yet, the main goal of a social recommendation system is to point out new, and possibly surprising, social ties. We measure the extent to which RSR reveals new relevant contacts. We further discuss the ways in which the *explainability* of RSR contributes to user engagement and helps to reveal interesting inter-personal relations. In addition to quantitative and qualitative analysis of RSR based on the feedbacks received in the user study, we compare RSR with several recommendation approaches implemented in the Conference Navigator (Brusilovsky et al., 2017) and make the case that RSR is complimentary to approaches such as CN.

7.1. *The relevance of RSR recommendations*

We distinguish between the following types of user feedbacks: 'know' – a recommended persons whom the user is already acquainted with; 'don't know and interested' – yet-unknown contacts whom the users found relevant; and 'don't know and not interested' – i.e., irrelevant recommendations. Fig. 4 details the distribution of feedback types per rank. As shown, a large proportion of the listed persons are already familiar to the user, and the respective curve dominates the other feedback types. The ratio of persons known to the user is as high as $\sim$ 80% at the topmost rank, decreasing gradually to roughly 50% at rank 10 and below. This result indicates that existing collaborators and friends are well-connected to the user in the graph, and is positive in the sense that it serves to acquire user trust in the recommendation system. As discussed below, users found interest in exploring the connecting paths also for known contacts.

As for the persons who were not yet known to them, users indicated the recommendations to be either interesting, or irrelevant. Ideally, the recommendations should include a large proportion of the first kind, and a small proportion of the latter. As shown in Fig. 4, this is the case for RSR. The curve that represents the ratio of unknown and interesting recommendations dominates the curve describing the ratio of irrelevant suggestions across all ranks, where dominance is most pronounced at the top of the list. For example, about 25% of the recommendations at rank 5 correspond to interesting and yet-unknown persons, where the ratio of irrelevant recommendations at this rank is less than 15%.

In summary, RSR demonstrates high ranking quality in the sense that the top ranks are populated with recommendations perceived as relevant by users. The chances of observing irrelevant recommendations is as low as 15% at rank 5, reaching 20% at rank 20. This means that a striking majority of the suggestions made by RSR are relevant. Next we turn to take a closer look at the question, to what extent, and how, does RSR detect novel and possibly surprising social ties, compared with other recommendation systems?

7.2. *Serendipity within RSR*

7.2.1. *How unique are RSR recommendations?*

We compare RSR with recommender approaches implemented in CN3, the latest version of the Conference Navigator system (Brusilovsky et al., 2017), which provides personalized social recommendations to conference participants. As detailed in Section 2, CN3 generates multiple rankings using different information sources and criteria, including semantic similarity of past publications (CN3:academic), co-authorship history (CN3:social), similarity of interests measured as overlap in bookmarked papers (CN3:interest), and geographical distance (CN3:distance).

Table 3

The average number of positive (relevant) feedbacks obtained at different ranks of RSR recommendation lists, compared with their distribution in CN3 rankings using different information sources (‘CN3:*’), or any of these sources (‘CN3:union’).

	count@rank5	count@rank10	count@rank15	count-full
RSR	4.10	6.88	9.07	12.17
CN3:academic	1.03	1.51	1.75	3.80
CN3:social	1.00	1.37	1.66	3.20
CN3:interest	0.36	0.41	0.44	0.53
CN3:distance	0.00	0.15	0.25	1.64
CN3 union	1.81	2.36	2.64	4.09

For comparison purposes, we obtained the ranked lists produced by the CN3 framework for the participants of the UMAP'17 conference. Notably, user feedbacks was elicited in our study per the recommendation lists generated by RSR. This means that there may exist relevant and novel recommendations by CN3, for which we did not receive feedbacks. For this reason, we cannot compare the two systems in terms of absolute recommendation quality. In- stead, we perform a directed comparison, observing the extent to which RSR suggested new and potentially surprising recommendations which were not discovered by CN3.

Table 3 details the average number of persons indicated as relevant (either known or yet unknown) per user at ranks 5,10,15, and in the full recommendation list by RSR. In addition, the table reports the number of persons known to be relevant among the top ranks, and the full recommendation lists generated by CN3 variants, with each variant considering a single information source. The table shows that a large number of persons who were recommended by RSR and judged to be relevant by users, do not ap- pear in CN3 lists. For example, there are 4.10 relevant persons on average at rank 5 of RSR recommendation lists, where only 1.03 persons known to be relevant (or less) are included in the top 5 recommendations by a CN3 variant. Similar trend is observed at ranks 10 and 15 of the ranked lists: RSR lists includes 6.88 and 9.07 relevant recommendations at these ranks, compared with up to 1.51 and 1.75 using a CN3 variant. The rightmost column of Table 3 shows the total number of persons indicated as relevant per user in the full recommendation lists. Overall, RSR lists include 12.17 relevant recommendations on average, compared with up to

3.80 by CN3. This comparison shows that *RSR presents a substantial number of relevant contacts at high ranks, which are not included among the top recommendations, or totally missed, by CN3 recommenders*.

Table 3 further compared RSR with the *union* of CN3 lists ('CN3 union'), considering the set of *distinct* relevant persons found by any CN3 variant at rank *k*. As shown, overall there are overall 1.81 distinct relevant persons found at rank 5 by any CN3 variant. Considering the full lists, we observe that 'CN3 union' tracks only 4.09 persons indicated as relevant vs. 12.17 in total by RSR. We there- fore conclude that *RSR can reach many relevant persons who are not found using any CN3 variant.*

Table 4 repeats this analysis, considering only recommended persons indicated by users as yet unknown and relevant. As this reference set of feedbacks is a subset of all relevant feedbacks, the figures in this analysis are strictly lower than in Table 3. However, similar trends are observed. Overall, RSR makes roughly 3.7 new and relevant recommendations per list, out of which only 0.95 are also found by any CN3 variant ('CN3 union'). This means that *RSR recommends yet unknown and relevant persons who are not found by alternative recommender approached implemented in CN3.*

Table 4

The average number of "relevant and yet unknown" feedbacks obtained at different ranks of RSR recommendation lists, compared with their distribution in CN3 rankings using different information sources ('CN3:*'), or any of these sources ('CN3 union').

	count@rank5	count@rank10	count@rank15	count-full
RSR	0.93	1.71	2.78	3.70
CN3:academic	0.14	0.19	0.22	0.88
CN3:social	0.07	0.09	0.19	0.73
CN3:interest	0.02	0.03	0.03	0.12
CN3:distance	0.00	0.02	0.05	0.56
CN3 union	0.22	0.29	0.42	0.95

7.2.2. *A path-based exploration of serendipity*

Interestingly, according to Table 3, CN3 rankings using coauthorship or publication similarity captured the largest number of persons indicated as relevant by the users of RSR. We indeed expect publication-related information to appear on personal scholar webpages. The gap between the number of relevant persons using the approaches systems indicates however that there is additional information on the personal webpages that do not appear in structured dedicated sources.

The question is, what are the sources of serendipity in the RSR framework, compared with methods like CN3? To address this question, we closely examined the relational paths between users and recommended scholars, indicated by the users as yet unknown and of interest to them. We found that many of the connecting paths concerned joint publication history, however, they encoded associative rather than structured links. For example, con- sider the connecting paths displayed in Fig. 2 between the re- searchers named as ‘John Smith’ and ‘Michael Jones’. The highest- weight path traverses the entity ‘University of Udine’. Indeed, the website of ‘John Smith’ reveals that he has been involved in a project coordinated by a Professor from University of Udine, and that ‘Michael Jones’ had a joint publication with another researcher affiliated with the University of Udine. Since the University of Udine has only few connections in the graph, this path was as- signed high weight by the PPR algorithm. Some of the other paths displayed in Fig. 2 correspond to direct co-authorship relations. In other cases, researchers got connected based on common publication venues, e.g. a journal or workshop name, etc.

In summary, RSR tracks non-trivial social connections because of several reasons. First, the graph accommodates multiple types of entities and inter-entity links, encoding diverse inter-personal associations. In addition, the PPR graph walk similarity measure assigns high importance to rare (and therefore, surprising) facts. Finally, RSR builds on facts extracted from personal homepages, and is not limited with respect to the types of the information it models. This means that in addition to information about academic publications, inter-personal relatedness may be established based on similar education history, hobbies or any other personal information available. Showing academic relations next to personal links should provide rich, holistic, and more intriguing, user experience.

7.3. *Explainability and user experience*

We found that the participants in this study took advantage of the system as intended, and tended to spend a relatively long time (8 minutes on average, while bounded by time constraints of the conference) browsing the recommendations and exploring the connecting paths that associated them with the suggested contacts. Some of the users provided us with comments in free form, either using the GUI, or by means of conversation. We hereby summarize these comments.[19]

Users indicated that they liked finding their academic super- visors or co-authors at the top part of the recommendation list, where this served to build their confidence in the system. Reasonably enough, senior researchers found interest in tracking the names of junior researchers participating in the conference who might fit Ph.D or postdoc positions, and juniors were interested in exploring their connections to senior figures, with the goal of expanding their professional social network. According to the users' comments, they found the display of detailed connecting paths to be helpful, and sometimes used the word 'surprising' in this con- text. For example, "I am surprised to see that the algorithm seems to make a connection through the creative technology programme in U.T, where I am on the advisory board (maybe N. also? But I have not been to the meetings recently so I never met her there).", or, "I was surprised to see A. at the top but when I checked out their research profile it makes some sense." And, "I certainly found it an interesting activity to go through the list for 10 minutes and check out the home pages of some of these people for which I was unfamiliar (or in some cases had forgotten)." Some participants mentioned that being shown the full set of connecting paths with multiple shared entities had raised their interest and motivation to get to know the person. Based on the above, we conclude that somewhat similarly to 'entity cards' which have

[19] Some comments were received also in a preliminary study using an earlier version of RSR (Amal et al., 2017).

proven to in- crease user engagement in search (Bota et al., 2015), presenting users with supporting evidence in the form of associative relational paths increases user engagement in the recommendation process.

Some of the participants in the study raised points for improvement. Several users wished to view a more elaborate description of the recommended person. This may be achieved by presenting the person's relational graph profile, and including a link to her home- page. Other comments referred to issues related to entity extraction, e.g., the extracted span "Seventeenth International Conference on User Modeling" omits the suffix, "Adaptation, and Personalization". Such errors in named entity recognition may be alleviated by using NER models trained on in-domain text and dictionaries.

A couple of other challenges indicated by users concern ambiguity and temporal aspects of fact extraction. Indeed, short named entities (e.g., abbreviations like 'ISI', 'SCE') or common names in general (e.g., 'Wang', 'Smith') are highly ambiguous. Entity disambiguation may be non-trivial however (for example, see Li et al., 2013) and is beyond the scope of this work. Following early feedbacks, we chose to ignore short entity names to avoid ambiguity.[20] Several participants have suggested to incorporate temporal relation modeling, and distinguish between recent and old facts. There is active research on temporal information ex- traction (Ling & Weld, 2010), which could be incorporated into frameworks like RSR in the future.

8. Can we do better with learning?

We have shown that the graph walk similarity measure yields high-quality rankings. Nevertheless, Markovian random walk schemes like PPR model local associations in the graph. Ideally, relational information such as

[20] Specifically, we required person names to include more than one token, or be otherwise longer than *n* characters.

edge sequences would be also taken into account in computing inter-node relatedness (Lao et al., 2016; Minkov & Cohen, 2010). Consider for example the path *education*(p, z)∧*education.inv*(z, q), indicating that persons p and q attended the same institution as students. Ideally, the similarity score of person q with respect to p would reflect the assumed importance of this evidence.

We apply a learning-to-rank approach, aiming to improve the initial ranking generated by PPR based on additional high-level evidence (Minkov & Cohen, 2010). Learning is supervised, aiming to rank the persons judged as relevant by the users above the irrelevant ones. In what follows, we propose several types of relational features designed to improve recommendation performance, and describe a set of preliminary experimental results.

8.1. *High-level relational features for social recommendation*

Given a pair of source and target nodes $< p,q >$, representing the user p and a person recommended for her q, we encode that pair using the following feature types, some of which are novel and adapted to the task of social recommendation.

- *PPR score.* We incorporate the node scores generated by the Markovian PPR as a standalone feature in the learning framework.
- *Path diversity.* A dedicated feature encodes the number of connecting paths between $< p,q >$, assuming that a larger number of paths indicate on stronger affinity, and vice versa. We further model the maximal proportion of probability mass contributed by a single connecting path to the PPR score of q. Presumably, a node reached over multiple high- scoring meaningful paths, each making a moderate contribution to the final similarity score, is more relevant than a node reached over a single dominant path (in which case the encoded proportion will be high).
- *Indirect vs. direct connecting paths.* We model the proportion of indirect paths among the set of connecting paths leading from p to q.

Presumably, bias towards indirect connecting paths would yield higher serendipity, as contacts who are linked to the user over a single-hope path (i.e., explicitly mentioned on her homepage) are likely to be already known to the user.

Related entities. Finally, we encode several features that de- scribe the entities included in the connecting paths. Concretely, these features indicate whether the connecting paths traverse any entity of type *person, organization*, or *location*. Using these information, the classifier may learn to prefer diversity with respect to the types of related entities that link the person pair. For a similar reason, another feature encodes the number of unique entities included in the connecting paths. Finally, we model the ratio of named entities traversed that are shorter than *k* characters, as we believe that these entities are likely to be common words or ambiguous names, and would like to minimize their impact on the final ranking.

Table 5

Ranking performance using RSR before and post learning.

	PPR	Reranked_1	Reranked_2
MAP	0.911	0.924	**0.927**
precision@rank1	0.968	0.957	**0.979**
precision@rank5	0.894	**0.912**	0.902
precision@rank10	0.844	**0.860**	0.854

8.2. *Experimental setup*

Our experimental dataset consists of the feedbacks collected from the participants of UMAP'17 and HYPERTEXT'17, including roughly 1,000 person pairs labeled with the user feedbacks. We evaluate learning results based on 5-fold cross validation experiments conducted using SVMRank, a popular learning-to-rank algorithm (Joachims, 2006). In order to mimic real-world settings–in which recommendations are generated for *new* users–the examples were strictly separated by user across data splits. This means that in each single experiment, a classifier is trained based on the feed-backs obtained from 80% of the participants, and evaluated on the reranked recommendation lists of the remaining participants.

8.3. *Learning results*

Table 5 details the performances of the initial rankings by PPR, and the re-ranked recommendation lists, in terms of mean aver- age precision (MAP) and precision-at-rank-*k* (the ratio of relevant persons by rank *k*). In this evaluation, all examples judged as relevant by the users (including "known" persons) were considered as positive responses. We attempted two training modes in reranking. The first mode used all training examples labeled as relevant ('$Rerank_1$'). In order to focus the learning process on non- trivially "yet-unknown" relevant contacts, we eliminated in the second mode all of examples labeled as 'known' from the training set while training another model, while maintaining the same test set ('$Rerank_2$').

As shown in the table, reranking generally improved recommendation quality with respect to all measures. The best MAP results (marked in boldface) were obtained using the novelty- focused learner ('$Rearnk_2$'). The respective improvement in MAP (from 0.911 to 0.927) is substantial considering the initially-high performance level. Importantly, our experimental dataset is limited in its size, and we therefore find these results to be highly encouraging. Should additional data be available for learning purposes, higher performance gains are to be expected.

9. **Conclusion**

We described RSR, a social relational recommendation approach that computes and ranks inter-personal affinity in a social graph comprised of personal relational profiles. Unlike other approaches, RSR presents explanations to the user in the form of detailed and labeled connecting relational paths. Such transparency is necessary for raising the user's engagement and interest in recommendations out of her social circle. As a case study, we described in detail the application of RSR to the task of social recommendation at scientific conferences. We used learning and information extraction techniques to automatically detect and process scholars' homepages into factual relational profiles for this purpose. User feedbacks elicited onsite of two international conferences indicated that the rankings and relational explanations provided by RSR were sensible, engaging, and surprising at times. We also showed that RSR made a substantial number of relevant and novel recommendations, and is therefore complementary to alternative scholarly recommendation tools. Finally, we outlined a relational learning schema, designed to further improve recommendation relevance and serendipity within RSR, with preliminary experiments yielding encouraging results.

The RSR framework is general, suggesting multiple applications and venues of future research. In the scholarly domain, associative personal facts extracted from semi-structured homepages may be combined with available structured sources, e.g., academic social networks, by means of post-processing or data integration. We expect this to improve the coverage, diversity, and explainability of recommendations. Task-wise, the graph-based framework is scalable, and may be applied to general social recommendation tasks in the scholarly domain, across disciplines and within ASNSs. Ideally, path and entity importance should be tuned per application and task. As discussed throughout this paper, explainability and serendipity are desired properties in recommender systems in general. Following some previous work in this area, we believe that the RSR approach can enhance also recommender systems of products and services, e.g., by extracting relational profiles of movies or businesses from their textual descriptions, tuning and displaying to the user detailed connectivity paths to these items. Finally, orthogonal areas of future research concern information extraction. For example, it is desired to incorporate temporal aspects into fact extraction and relational learning.

Credit authorship contribution statement

Saeed Amal: Conceptualization, Methodology, Software, Data curation, Investigation, Visualization, Writing - original draft, Writing - review & editing.

Chun-Hua Tsai: Data curation, Investigation, Writing - original draft, Writing - review & editing.

Peter Brusilovsky: Investigation, Supervision, Writing - review & editing.

Tsvi Kuflik: Conceptualization, Methodology, Investigation, Supervision, Writing - review & editing.

Einat Minkov: Conceptualization, Methodology, Investigation, Supervision, Writing - original draft, Writing - review & editing.

References

Abujabal, A., Roy, R. S., Yahya, M., & Weikum, G. (2017). Quint: Interpretable question answering over knowledge bases. In *Proceedings of the 2017 EMNLP system demonstrations* (pp. 61–66).

Adamic, L. A., & Adar, E. (2003). Friends and neighbors on the web. *Social Networks, 25*(3), 211–230.

Al-Rfou, R., Kulkarni, V., Perozzi, B., & Skiena, S. (2015). POLYGLOT-NER: Massive multilingual named entity recognition. In *Proceedings of the SIAM international conference on data mining (ICDM)* (pp. 586–594).

Amal, S., Kuflik, T., & Minkov, E. (2017). Harvesting entity-relation social networks from the web. In *Proceedings of the 2017 conference on user modeling, adaptation, and personalization, short papers (UMAP)* (pp. 351–352).

Amini, B., Ibrahim, R., Othman, M. S., & Selamat, A. (2014). Capturing scholars knowledge from heterogeneous resources for profiling in recommender systems. *Expert Systems with Applications, 41*, 7945–7957.

Backstrom, L., & Leskovec, J. (2011). Supervised random walks: Predicting and recommending links in social networks. In *Proceedings of the ACM international conference on web search and data mining (WSDM)* (pp. 635–644).

Balog, K., Carmel, D., Arjen, de Vries, P., Herzig, D. M., Mika, P., et al. (2012). The first joint international workshop on entity-oriented and semantic search (JIWES). *SIGIR Forum, 46*(2), 87–94.

Bhardwaj, R. K. (2017). Academic social networking sites: Comparative analysis of ResearchGate, Academia.edu, Mendeley and Zotero. *Information and Learning Science, 118*(5/6), 298–316.

Blanco, R., Halpin, H., Herzig, D. M., Mika, P., Pound, J., & Thompson, H. S. (2011). Entity search evaluation over structured web data. In *Proceedings of the first international workshop on entity-oriented search (EOS-SIGIR)*.

Bota, H., Zhou, K., & Jose, J. M. (2015). Playing your cards right: The effect of entity cards on search behaviour and workload. In *Proceedings of the ACM on conference on human information interaction and retrieval (CHIIR)* (pp. 131–140).

Brusilovsky, P., Oh, J. S., López, C., Parra, D., & Jeng, W. (2017). Linking information and people in a social system for academic conferences. *New Review of Hyper- media and Multimedia, 23*(2), 1–31.

Champiri, Z. D., Shahamiri, S. R., & Salim, S. S. B. (2015). A systematic review of scholar context-aware recommender systems. *Expert Systems with Applications, 42*, 1743–1758.

Chang, C.-H., Kayed, M., & Girgis, M. R. (2006). A survey of web information ex- traction systems. *IEEE Transactions on Knowledge and Data Engineering (TKDE), 18*(10), 1411–1428.

Chaudhari, S., Azaria, A., & Mitchell, T. M. (2017). An entity graph based recommender system. *AI Communications, 30*(2), 141–149.

Chen, J., Tang, Y., Li, J., Mao, C., & Xiao, J. (2013). Community-based scholar recommendation modeling in academic social network sites. In

Proceedings of the international conference on web information systems engineering (pp. 325–334). Springer.

Das, S., Giles, C. L., Mitra, P., & Caragea, C. (2011). On identifying academic home- pages for digital libraries. In *Proceedings of the eleventh annual international ACM/IEEE joint conference on digital libraries (JDCL)* (pp. 123–132).

Das Gollapalli, S., Caragea, C., Mitra, P., & Lee Giles, C. (2015). Improving researcher homepage classification with unlabeled data. *ACM Transactions on the Web (TWEB), 9*(4), 1–32.

Finkel, J. R., Grenager, T., & Manning, C. (2005). Incorporating non-local information into information extraction systems by Gibbs sampling. In *Proceedings of the annual meeting on association for computational linguistics (ACL)* (pp. 363–370).

Fogaras, D., Rácz, B., Csalogány, K., & Sarlós, T. (2005). Towards scaling fully personalized PageRank: Algorithms, lower bounds, and experiments.. *Internet Mathematics, 2*(3), 333–358.

Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., et al. (2007). Evaluation of digital libraries. *International Journal on Digital Libraries, 8*, 2138.

Gasparyan, A. Y., Nurmashev, B., Yessirkepov, M., Endovitskiy, D. A., Voronov, A. A., & Kitas, G. D. (2017). Researcher and author profiles: Opportunities, advantages, and limitations. *Journal of Korean Medical Science, 32*(11), 1749–1756.

Graells-Garrido, E., Lalmas, M., & Baeza-Yates, R. (2016). Data portraits and intermediary topics: Encouraging exploration of politically diverse profiles. In *Proceedings of the twenty-first international conference on intelligent user interfaces* (pp. 228–240). ACM.

Guy, I. (2015). Social recommender systems. In *Recommender systems handbook* (pp. 511–543). Springer.

Guy, I. (2018a). People recommendation on social media. In *Social information access* (pp. 570–623). Springer.

Guy, I. (2018b). People recommendation on social media. In P. Brusilovsky, & D. He (Eds.), *Social information access: Systems and technologies* (pp. 570–623)). Springer International Publishing.

Guy, I., Zwerdling, N., Ronen, I., Carmel, D., & Uziel, E. (2010). Social media recommendation based on people and tags. In *Proceedings of the international ACM SIGIR conference on research and development in information retrieval (SIGIR)* (pp. 194–201).

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The Weka data mining software: An update. *SIGKDD Explorations, 11*(1), 10–18.

He, X., Chen, T., Kan, M.-Y., & Chen, X. (2015). TriRank: Review-aware explain- able recommendation by modeling aspects. In *Proceedings of the twenty-fourth ACM international on conference on information and knowledge management* (pp. 1661–1670). ACM.

Heck, T. (2013). Combining social information for academic networking. In *Proceedings of the conference on computer supported cooperative work (CSCW)* (pp. 1387–1398).

Heckel, R., Vlachos, M., Parnell, T., & Dünner, C. (2017). Scalable and inter- pretable product recommendations via overlapping co-clustering. In *Proceedings of the thirty-third IEEE international conference on data engineering (ICDE)* (pp. 1033–1044). IEEE.

Herzig, D. M., Mika, P., Blanco, R., & Tran, T. (2013). Federated entity search using on-the-fly consolidation. In *Proceedings of the international semantic web conference* (pp. 167–183). Springer.

Jeh, G., & Widom, J. (2003). Scaling personalized web search. In *Proceedings of the international conference on world wide web (WWW)* (pp. 271–279).

Ji, K., & Shen, H. (2016). Jointly modeling content, social network and ratings for explainable and cold-start recommendation. *Neurocomputing, 218*, 1–12.

Joachims, T. (2006). Training linear SVMs in linear time. In *Proceedings of* *the ACM conference on knowledge discovery and data mining (KDD)* (pp. 217–226).

Kaminskas, M., & Bridge, D. (2016). Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Transactions on Interactive Intelligent Systems, 7*(1), 2:1–2:42.

Kang, C., Vadrevu, S., Zhang, R., van Zwol, R., Pueyo, L. G., Torzec, N., et al. (2011). Ranking related entities for web search queries. In *Proceedings of the international conference companion on world wide web (WWW)* (pp. 67–68).

Kruk, S. R., Kruk, E., & Stankiewicz, K. (2009). Evaluation of semantic and social technologies for digital libraries. In *Semantic digital libraries* (pp. 203–214). Springer.

Lao, N., Minkov, E., & Cohen, W. W. (2016). Learning relational features with back- ward random walks. In *Proceedings of the annual meeting of the association for computational linguistics (ACL)* (pp. 666–675).

Lee, D., & Brusilovsky, P. (2014). Recommending talks at research conferences using users' social networks. *International Journal of Cooperative Information Systems, 23*, 1441003.

Lee, D., & Brusilovsky, P. (2018). Recommendations based on social links. In P. Brusilovsky, & D. He (Eds.), *Social information access: Systems and technologies*. In *Lecture notes in computer science* (pp. 391–440). Cham: Springer.

Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2010). Predicting positive and negative links in online social networks. In *Proceedings of the international conference on world wide web (WWW)*.

Li, Y., Wang, C., Han, F., Han, J., Roth, D., & Yan, X. (2013). Mining evidences for named entity disambiguation. In *Proceedings of the nineteenth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1070–1078). ACM.

Liao, Q. V., & Fu, W.-T. (2013). Beyond the filter bubble: Interactive effects of per- ceived threat and topic involvement on selective exposure to information. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2359–2368). ACM.

Liben-nowell, D., & Kleinberg, J. (2007). The link-prediction problem for social net- works. *Journal of American Society for Information Science and Technology (JASIST), 58*(7), 1019–1031.

Ling, X., & Weld, D. S. (2010). Temporal information extraction. In *Proceedings of the twenty-fourth AAAI conference on artificial intelligence (AAAI)* (pp. 1385–1390).

Ling, X., & Weld, D. S. (2012). Fine-grained entity recognition. In *Proceedings of the conference on artificial intelligence (AAAI)* (pp. 94–100).

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & López-Cózar, E. D. (2018). Google scholar, web of science, and scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics, 12*(4), 1160–1177.

McCay-Peet, L., & Toms, E. G. (2015). Investigating serendipity: How it unfolds and what may influence it. *Journal of the Association for Information Science and Technology (JASIST), 66*(7), 1463–1476.

Miliaraki, I., Blanco, R., & Lalmas, M. (2015). From “Selena Gomez” to “Marlon Brando”: Understanding explorative entity search. In *Proceedings of the international world wide web conference (WWW)* (pp. 765–775).

Minkov, E. (2016). Event extraction using structured learning and rich domain knowledge: Application across domains and data sources. *ACM Transactions on Intelligent Systems and Technology (TIST), 7*(2), 16:1–16:34.

Minkov, E., & Cohen, W. W. (2010). Improving graph-walk-based similarity with reranking: Case studies for personal information management. *ACM Transactions on Information Systems (TOIS), 29*(1), 4:1–4:52.

Minkov, E., Kahanov, K., & Kuflik, T. (2017). Graph-based recommendation integrating rating history and domain knowledge: Application to on-site guidance of museum visitors. *Journal of the Association for Information Science and Technology (JASIST), 68*(8), 1911–1924.

Mintz, M., Bills, S., Snow, R., & Jurafsky, D. (2009). Distant supervision for relation extraction without labeled data. In *Proceedings of the annual meeting of the ACL and the IJCNLP of the AFNLP (ACL and AFNLP)* (pp. 1003–1011).

Ortega, J. L. (2015). Disciplinary differences in the use of academic social networking sites. *Online Information Review, 39*(4), 520–536.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web. *Technical report*.

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. UK: Penguin. Parra, D., & Brusilovsky, P. (2015). User-controllable personalization: A case study with setfusion. *International Journal of Human-Computer Studies, 78*, 43–67.

Porcel, C., Moreno, J. M., & Herrera-Viedma, E. (2009). A multi-disciplinar recommender system to advice research resources in university digital libraries. *Expert Systems with Applications, 36*, 12520–12528.

Pritsker, E. W., Kuflik, T., & Minkov, E. (2017). Assessing the contribution of Twitter's textual information to graph-based recommendation. In *Proceedings of the twenty-second international conference on intelligent user interfaces (IUI)* (pp. 511–516).

Pu, P., & Chen, L. (2007). Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems, 20*(6), 542–556.

Richardson, M., & Domingos, P. (2002). The intelligent surfer: Probabilistic combi- nation of link and content information in PageRank. In *Advances in neural information processing systems conference (NIPS)* (pp. 1441–1448).

Sánchez, L. Q., Sauer, C., Recio-García, J. A., & Díaz-Agudo, B. (2017). Make it personal: A social explanation system applied to group recommendations. *Expert Systems with Applications, 76*, 36–48.

Seo, Y., Kim, Y., Lee, E., & Baik, D. (2017). Personalized recommender system based on friendship strength in social network services. *Expert Systems with Applications, 69*, 135–148.

Serrano-Guerrero, J., Herrera-Viedma, E., Olivas, J. A., Cerezo, A., & Romero, F. P. (2011). A Google wave-based fuzzy recommender system to disseminate information in university digital libraries 2.0. *Information Sciences, 181*, 1503–1516.

Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). ArnetMiner: Extraction and mining of academic social networks. In *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining (KDD)* (pp. 990–998).

Tintarev, N., & Masthoff, J. (2015). Explaining recommendations: Design and evaluation. In *Recommender systems handbook* (pp. 353–382). Springer.

Toutanova, K., Manning, C. D., & Ng, A. Y. (2004). Learning random walk models for inducing word dependency distributions. In *Proceedings of the international conference on machine learning (ICML)* (pp. 103–111).

Tsai, C.-H., & Brusilovsky, P. (2017). Providing control and transparency in a social recommender system for academic conferences. In *Proceedings of the twenty– fifth conference on user modeling, adaptation and personalization* (pp. 313–317).

Tsai, C.-H., & Brusilovsky, P. (2018). Beyond the ranked list: User-driven exploration and diversification of social recommendation. In *Proceedings of the twenty-third international conference on intelligent user interfaces (IUI)* (pp. 239–250).

Tsurel, D., Pelleg, D., Guy, I., & Shahaf, D. (2017). Fun facts: Automatic trivia fact ex- traction from Wikipedia. In *Proceedings of the tenth ACM*

international conference on web search and data mining (WSDM).
Verbert, K., Parra, D., Brusilovsky, P., & Duval, E. (2013). Visualizing recommendations to support exploration, transparency and controllability. In *Proceedings of the 2013 international conference on intelligent user interfaces* (pp. 351–362). ACM.
Vexler, R., & Minkov, E. (2016). Multi-source named entity typing for social me- dia. In *Proceedings of the sixth named entities workshop, collocated with ACL (News-ACL'16)* (pp. 11–20).
Wang, H., Zhang, F., Wang, J., Zhao, M., Li, W., Xie, X., & Guo, M. (2018). RippleNet: Propagating User Preferences on the Knowledge Graph for Recommender Systems. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)* (pp. 417–426).
Willinsky, J. (2006). *The access principle: The case for open access to research and scholarship*. Cambridge, Mass.: MIT Press.
Wu, F., & Weld, D. S. (2008). Automatically refining the Wikipedia infobox ontology. In *Proceedings of the international conference on world wide web (WWW)* (pp. 635–644).
Yogev, S., Roitman, H., Carmel, D., & Zwerdling, N. (2012). Towards expressive exploratory search over entity-relationship data. In *Proceedings of the twenty-first international conference on world wide web* (pp. 83–92).
Yu, X., Ren, X., Sun, Y., Gu, Q., Sturt, B., Khandelwal, U., et al. (2014). Personalized entity recommendation: A heterogeneous information network approach. In *Proceedings of the seventh ACM international conference on web search and data mining* (pp. 283–292). ACM.
Zhang, Y. C., Séaghdha, D. Ó., Quercia, D., & Jambor, T. (2012). Auralist: Introducing serendipity into music recommendation. In *Proceedings of the fifth ACM international conference on web search and data mining* (pp. 13–22). ACM.
Zhao, H., Yao, Q., Li, J., Song, Y., & Lee, D. L. (2017). Meta-graph based

recommendation fusion over heterogeneous information networks. In *Proceedings of the twenty-third ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 635–644). ACM.