Towards Reputation-Aware Expert Finding with Linked Open Data

Sebastian Heil
Technische Universität Chemnitz
Chemnitz, Germany
sebastian.heil@informatik.tu-chemnitz.de

Stefan Wild
Technische Universität Chemnitz
Chemnitz, Germany
stefan.wild@informatik.tu-chemnitz.de

Michael Krug
Technische Universität Chemnitz
Chemnitz, Germany
michael.krug@informatik.tu-chemnitz.de

Martin Gaedke
Technische Universität Chemnitz
Chemnitz, Germany
martin.gaedke@informatik.tu-chemnitz.de

ABSTRACT

Distributed social networks allow creating new work patterns, addressing the workforce of a company as crowd. Here, finding suitable workers for specific functions is important for work quality, but largely relies on human assessment. In web-scale environments this assessment exceeds human capability. Linked open data has proven to be successful in providing semantic descriptions and discoverability of distributed resources. Hence, we leverage linked open data, so that each worker can have a semantic profile based on WebID and reference co-workers, skills, projects, etc. To recommend suitable experts for a given task, supporting systems are required, which use this profile data. In this paper, we extend our previous work on CRAWL towards reputation-aware expert finding in distributed social networks. We outline three major aspects – *Endorsements*, *Achievements* and *Meta Reputation* – to achieve reputation awareness and report on our progress, showcase open challenges and present a roadmap for future work.

CCS Concepts

•**Information systems** → *Web and social media search; Social recommendation; Resource Description Framework (RDF);*

Keywords

Expert Finding, Linked Open Data, Reputation, WebID

1. INTRODUCTION

In large enterprises with a diverse workforce distributed around the world, conventionally assigning work is difficult. Therefore, the workforce can be considered an *in-house crowd*: individual workers or ad hoc teams for problem-solving are built from "an undefined (and generally large) network of people" [8]. These workers have already passed pre-selection (assessment centers etc.) and possess a common understanding of the business objectives, company structure and values.

SEMANTICS 2016: Posters and Demos Track
September 13-14, 2016, Leipzig, Germany

With the low risk of fake/low quality contributions known in traditional crowdsourcing [1], professional selection criteria like expertise and reputation become more important. Finding the right people to solve a task or problem [8] is crucial for this way of working [3]. Expert Finding allows for retrieving a list of potential experts based on a query with target or constraint criteria and evaluation of potential candidates against these criteria [2, 3].

While traditionally employee data is managed by the Human Resources (HR) department, it is beneficial to make parts of it available to other employees, project partners and customers. Distributed social networks (DSNs) based on semantic technologies allow employees to do so beyond the scope of traditional HR. They enable employees to own, self-manage and control their personal data, organize social connections across companies, and support personal profile data beyond requirements/limitations of a particular company. Privacy is then controlled by the employees since they have control on what information to include in their profiles. This facilitates dynamic team building processes of in-house crowds and integration of subcontractors, because employees can bring all their personal data and professional history with them. So, companies do not need to re-create and maintain silos of an employee's personal data.

Linked open data (LOD) has proven to be successful in providing semantic descriptions and discoverability of distributed resources. Therefore, in [7] we introduced CRAWL. Extending WebID[1] profiles with semantic statements about skills, it allows expert finding in DSNs. The work in [7] is the base from where we try to move towards reputation-aware expert finding (RaEF). The consideration of skill endorsements can be regarded as first step in this direction.

A person's reputation is a community-wide judgment resulting from a social evaluation by community members who know the potential of the person in question [1]. It is part of a person's identity and an ongoing subject to change, e.g., by achievements, endorsements and negative impacts. A person with a high reputation is well-suited to provide a sound validation of another person's qualities. It is also suited as selection criterion and differentiator between members of the crowd and therefore of special relevance for expert finding.

[1]https://www.w3.org/2005/Incubator/webid/spec/

In this paper, we explore RaEF in DSNs. Following an overview of the use of semantic technologies in distributed expert finding in section 2, section 3 details main aspects, 4 discusses related work and 5 outlines a roadmap for next steps.

2. EXPERT FINDING IN DISTRIBUTED SOCIAL NETWORKS USING LOD

In this section we give an overview on expert finding in DSNs with CRAWL [7]. We employ semantic technologies and traverse LOD-based graphs. To find a suitable expert, the social graph of the requester, established by `foaf:knows` properties in WebID profiles, has to be traversed. WebID profiles provide a machine-readable description of an identity owner's personal data. Leveraging LOD, WebID relies on several RDF-vocabularies such as FOAF. Vocabularies such as Organization Ontology[2] or Human Resources Management Ontology[3] can be used to describe the employee data.

The algorithm traversing the social graph formed by connected WebID profiles is a depth-limited breadth-first search. It takes a WebID URI from the queue, fetches the corresponding WebID profile, calculates the rating function R (cf. Equation (1)), marks this WebID URI as visited and adds all WebID URIs referenced via `foaf:knows` and not yet visited along with their depth value to the queue. The queue is initialized with the WebID URIs of the requester and persons already involved with the given task. An upper boundary for the depth restricts the algorithm's runtime due to the exponentially rising number of nodes in a social graph with increasing depth [5].

To calculate the rating of the traversed WebID profiles, the set of concepts stating the required skills $\{r_1, \ldots, r_m\}$ is compared to the set of concepts describing the existing skills of each candidate $E_C = \{e_1, \ldots, e_n\}$. Leveraging LOD, we employ DBpedia URIs to represent the skills in these sets. This allows us to calculate the similarity $s(r, e)$ value between two concepts distinguishing different types of concept matches based on the information provided by the DBpedia resource:

1. Exact Concept Match - URIs are identical: $e = r$
2. Same Concept As Match - URIs are connected via `owl:sameAs`: r `owl:sameAs` e
3. Related Concept Match - URIs connected via `dbprop:paradigm`, `dcterms:subject`, `skos:narrower` etc.

For each pair (r_i, e_j), the similarity function s in Equation (1) gives different weight according to these concept matches. The candidate rating R is calculated considering only the maximum similarity per required skill:

$$R(C) = \sum_{i=0}^{m} \max_{0 \leq j \leq n} s(r_i, e_j) \quad e_j \in E_C \tag{1}$$

Then, the candidate WebID profile triples are added to a triplestore, asserting an additional triple for the calculated rating. The SPARQL query shown in Listing 1 yields the final ordered list of rated candidates. Due to the distributedness of profiles - each profile has to be fetched via HTTP - sequential traversal of WebID profiles has a huge impact on performance. We addressed this issue by concurrency and caching of user profiles and DBpedia-provided skill descriptions [7].

[2]https://www.w3.org/TR/vocab-org/

[3]http://mayor2.dia.fi.upm.es/oeg/oeg-upm/index.php/en/ontologies/99-hrmontology/

```
SELECT ?candidate ?rating
WHERE { ?candidate a foaf:Person .
  ?candidate vsrcm:rating ?rating.
  FILTER( ?rating > [MIN_RATING] )}
ORDER BY DESC( ?rating )
```

Listing 1: SPARQL query for candidates.

CRAWL demonstrates the basic concept of expert finding in distributed social networks leveraging knowledge from profiles, case descriptions and Linked Open Data [7]. Yet, it does not consider reputation aspects, representing knowledge of people about people, which is addressed in the following.

3. TOWARDS REPUTATION-AWARENESS WITH CRAWL·E

Expert finding can be considered a retrieval task: candidates are matched against the initial query. Each candidate is ranked according to the suitability for this query. A variety of aspects can be considered. The taxonomy presented in [1] distinguishes between *expertise* and *reputation*. Current expert finders chiefly rely on expertise i.e., they consider *credentials* or explicit *skill* lists of candidates made by themselves. However, reputation is mainly built on feedback [1] i.e., it is formed by statements made by other persons. So, to achieve RaEF, we consider the following three important aspects:

Endorsements: candidate reputation in terms of skill endorsements

Achievements: candidate reputation in terms of past projects & ratings

Meta Reputation: endorser reputation used to weight endorsements

While the first two aspects consider *candidate reputation* i.e., the reputation of the candidate to be ranked by the expert finder, the latter aspect additionally considers *endorser reputation* i.e., the influence of the reputation of those who contribute to the reputation of the candidate. In the following subsections we briefly outline these three aspects, indicate our current progress and propose ideas for further research.

3.1 Endorsements

Allowing others to endorse workers for their expertise partially enables consideration of candidate reputation for expert finding. In [6] we formally introduced the concept of *endorsements* for expressing arbitrary skill statements. An endorsement e is a tuple $e = (p1, p2, s)$: endorser $p1$ states that endorsee $p2$ possesses skill s. In CRAWL·E, $p1$ and $p2$ are WebID URIs identifying endorser and endorsee and referencing their RDF profile data, s is a DBpedia resource URI identifying the skill and referencing triples describing that skill. Reification is required to express the basic statement $p2$ has skill s and that $p1$ states that, using RDF. While current expert finders mainly consider skill self-claims i.e., endorsements $e = (p1, p2, s)$ with $p1 = p2$, we argue that considering general endorsements with $p1 \neq p2$ is important for reputation-awareness (RA). Adapting expert finding algorithms to consider endorsements is a step towards RA. We demonstrated this in [6]. Endorsement quantity, co-relating with candidate reputation, for a specific skill is represented as a factor in the ranking function R. The more endorsements, the higher the influence of the endorsed skill on the candidate rank. Therefore, we consider *Endorsements* as solved. In the following, we detail the two remaining open issues.

3.2 Achievements

An important part of reputation is explicit feedback "on a worker's quality or contributions" by rating "content the worker has created" [1]. RaEF should therefore consider feedback on past projects. We propose to describe achievements in terms of projects and goals with associated required skills. Collaborators can rate a worker's contribution to a goal or to the entire project. For instance, an extension of RDF vocabularies like *PDO*[4] or *DOAP*[5] could be used. For modeling goals and their relation to projects the *BMM*[6] specification should be considered. To express ratings of a worker's contributions to past projects, schema.org rating[7] or the RDF Review Vocabulary[8] could be re-used.

According to [1], the "worker profile" should contain this feedback information. As an extension of [6], we suggest adding it to workers' WebID profiles. The FOAF[9] (Friend of a Friend) vocabulary already employed in WebID profiles allows for easy integration via `foaf:currentProject` and `foaf:pastProject` properties and the `foaf:Project` class.

With the proposed extensions of worker profiles, general and skill-based candidate reputation would be available for expert finding algorithms. Then, successfully integrating this information in the candidate ranking metrics for expert finding is an open challenge. The basic idea is to introduce an achievement factor, similar to the endorsement factor ε in [6]. Alternatively, a combination of both factors forming a new reputation factor can simplify the metrics.

3.3 Meta Reputation

Even with the above improvements, an expert finder like [6] is not entirely reputation-aware, as it uses endorsement quantity, but not quality: Only the amount of skill endorsements influences the candidate reputation, but not the reputation of the endorsers, i.e., meta reputation. However, the endorsement of a renowned expert should have more weight compared to endorsements of less renowned persons. This requires considering endorsements of the endorsers in addition to the endorsements of the candidate. Figure 1 illustrates this idea. *Alice* is endorsed for skill s by *Bob* and *Davis*. $E1$ and $E2$ define her candidate reputation regarding s. $E1$, however, should have higher impact, because Bob is a renowned expert for s himself, as seen by *Charlie* and many *others* endorsing Bob for s in E^*1 to E^*n. The meta reputation from E^* influences Alice' reputation E.

We propose to use endorser reputation as a weight for endorsements. This way, $E1$ in the above example will have more impact on Alice' rating for s. There are two levels of endorsements, rating a candidate, first-level endorsements should be considered qualitatively i.e., taking meta reputation into account. To assess endorser reputation, however, second-level endorsements should be considered only quantitatively. This is a consequence of the increasing complexity

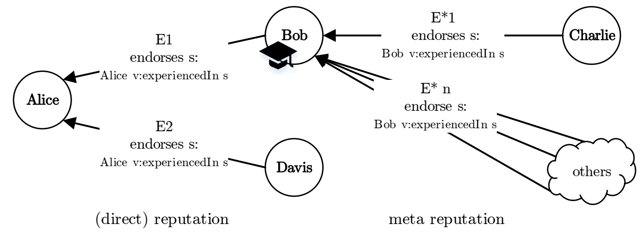

Figure 1: Two levels of endorsements/reputation in a social graph

of social graphs with increasing depth. In distributed social networks, each node potentially requires fetching via Internet, thus multilevel qualitative endorsements pose a significant performance issue. Loop avoidance is another point to consider for recursive multilevel qualitative endorsements.

4. RELATED WORK

Expert Finding has long been a research interest. For example, [2] provides an overview on Web-based expert finding approaches, [12] surveys expert finding systems with a focus on social context including communication and blogs.

The approach described by Xu et al. in [14] is similar to [7] in that it unites social graphs with skill relationship semantics. Network analysis on interlinked concept (expertise) and research (social) layer is employed. While it also considers hierarchical and correlation relationships in the expertise layer, tacit knowledge is used. [7] uses explicit knowledge from profiles and Linked Open Data, whereas [14] extracts information from unstructured text sources supported by WordNet. None of the above expert finding approaches works in DSNs, nor do they explicitly consider reputation. Only domain-specific measures, e.g., co-authorship on a document about a certain topic, could be considered an implicit endorsement within the group of authors.

[3] presents an expert finding method based on user's activities in centralized social networks like Facebook, Twitter etc. It analyzes social resources directly related (e.g. tweets, likes) or indirectly related (e.g. posts of liked pages) to persons. This approach employs text analysis: entity recognition for skills is performed on the resources, they are identified with Wikipedia URIs. Our proposed approach, by contrast, targets distributed social networks, uses explicit expertise information, leverages Linked Data [6] and aims at RA.

In spite of providing benefits such as a streamlined version of a resume, skill endorsements, which we consider an important aspect of RA, have not gained much attention in research so far. Centralized platforms like LinkedIn[10] already have successfully included them and addressed problems like unwanted endorsements, but do not consider the quality aspects yet [4] which we addressed in section 3.3. Reputation in more general terms has been considered in various domains like e-commerce, Q&A portals, collaborative filtering and in social network analysis [9, 13]. However, to the best of our knowledge there has not yet been a systematic consideration of reputation for expert finder systems in DSNs.

Closest to present work, Pérez-Rosés et al. propose an approach which combines social graphs with skill endorsements in [11]. This is an important step towards RA in expert

[4] http://vocab.deri.ie/pdo
[5] http://usefulinc.com/ns/doap
[6] http://www.omg.org/spec/BMM/1.2/
[7] http://schema.org/Rating
[8] http://vocab.org/review/terms.html#rating
[9] http://xmlns.com/foaf/spec/
[10] https://www.linkedin.com/

finding. The PageRank algorithm is applied to a deduced graph. Its deduction matrix is similar to the similarity matrix in [7]. However, the authors state that the definition of the deduction matrix is an open problem whereas the values of the matrix are defined leveraging Linked Open Data in [7]. Pérez-Rosés et al. assume bilateral relationships between social network members [11], while the `foaf:knows` semantics in [7] allow for unilateral relationships. Reputation in the sense of section 3.2 or section 3.3 is not considered.

CRAWL·E is similar to PageRank, HITS and SALSA [10], which traverse and rank web resources based on hyperlinks. PageRank increases ranking with the number of incoming links, equivalent to section 3.1, HITS Hubs and Authorities can be compared to the concept of Meta Reputation in section 3.3. However, CRAWL·E employs explicit semantics of links and profiles compared to generic HTML anchors and keyword-based content analysis. Also, since DSNs are by far not as large and connected as the web, we employ a combined traversal and ranking starting from the requester profile instead of pre-crawling resources for indexing and storage.

5. ROADMAP

In this paper we motivated the expert finders in distributed social networks, detailed three important aspects towards reputation-aware expert finding in DSNs and reported on our progress. While previous work presented the integration of endorsements, achievements and meta reputation are still to be investigated thoroughly. The following presents a roadmap for future work in the area of evaluation and optimization.

Future work has to focus on the implementation and evaluation of our proposal. We seek to address the challenges outlined in section 3 as indicated. In order to evaluate the approach we plan to create a set of sample social graphs containing additional expertise and reputation information. Crowdsourcing or expert interviews can then be employed to perform candidate selection manually on these samples and create reference results. Running the implementation of our proposed approach will then yield precision, recall and performance data, the three measures typically found in the evaluation of such retrieval systems. The data gained from this evaluation can be used both to compare the proposed approach to conventional candidate selection approaches and to optimize it. Adequate consideration of endorsements, achievements and meta reputation in the expert finding process by determining appropriate weights is an important objective here. These will be adapted with supervised machine learning methods aiming at increasing precision and recall.

The characteristics of distributed social networks and linked open data have a huge influence on expert finding in this context as indicated in [7]. We therefore seek to optimize our approach in terms of performance, investigating, for example, improvements by caching and concurrency. Further work is needed to inquire the possibility to convert the algorithm into a MapReduce variant which would allow running in the Hadoop environments of the major cloud providers.

References

[1] M. Allahbakhsh, B. Benatallah, A. Ignjatovic, H. R. Motahari-Nezhad, E. Bertino, and S. Dustdar. Quality control in crowdsourcing systems: Issues and directions. *IEEE Internet Computing*, 17(2):76–81, 2013. .

[2] I. Becerra-Fernandez. Searching for experts on the Web: A review of contemporary expertise locator systems. *ACM TOIT*, 6(4):333–355, 2006.

[3] A. Bozzon, M. Brambilla, S. Ceri, M. Silvestri, and G. Vesci. Choosing the Right Crowd : Expert Finding in Social Networks Categories and Subject Descriptors. In *Proc. of EDBT*, 2013. ISBN 9781450315975.

[4] A. Doyle. How To Use LinkedIn Endorsements, 2012. URL http://jobsearch.about.com/od/linkedin/qt/linkedin-endorsements.htm.

[5] S. Goel, R. Muhamad, and D. Watts. Social search in "small-world" experiments. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 701–710, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-487-4. .

[6] S. Heil, S. Wild, and M. Gaedke. CRAWLE: Distributed Skill Endorsements in Expert Finding. In S. Casteleyn, G. Rossie, and M. Winckler, editors, *To appear in Proceedings of ICWE2014*, Toulouse, France, 2014. Springer Berlin Heidelberg.

[7] S. Heil, S. Wild, and M. Gaedke. Collaborative Adaptive Case Management with Linked Data. In *Proceedings of WWW2014 companion*, pages 99–102, 2014.

[8] J. Howe. Crowdsourcing: A Definition. *Crowdsourcing Weblog*, 2006.

[9] S. Kitisin and C. Neuman. Reputation-based Trust-Aware Recommender System. In *2006 Securecomm and Workshops*, pages 1–7, Baltimore, MD, USA, 2006. IEEE. ISBN 1-4244-0422-3. .

[10] A. N. Langville and C. D. Meyer. A Survey of Eigenvector Methods for Web Information Retrieval. *SIAM Review*, 47(1):135–161, Jan 2005. ISSN 0036-1445. .

[11] H. Pérez-Rosés, F. Sebé, and J. M. Ribó. Endorsement Deduction and Ranking in Social Networks. In *7th GraphMasters Workshop*, Lleida, Spain, 2013.

[12] S. Perugini, M. Goncalves, and E. Fox. A connection-centric survey of recommender systems research. *Journal of Intelligent Information Systems*, 23(2):107–143, 2004.

[13] J. Sabater and C. Sierra. Reputation and social network analysis in multi-agent systems. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems part 1 - AAMAS '02*, pages 475–482, New York, New York, USA, 2002. ACM Press. ISBN 1581134800. .

[14] Y. Xu, X. Guo, J. Hao, J. Ma, R. Y. Lau, and W. Xu. Combining social network and semantic concept analysis for personalized academic researcher recommendation. *Decision Support Systems*, 54(1):564–573, 2012. ISSN 01679236. .