Choosing the Right Crowd: Expert Finding in Social Networks

Alessandro Bozzon, Marco Brambilla, Stefano Ceri, Matteo Silvestri, Giuliano Vesci
Politecnico di Milano, Dipartimento di Elettronica e Informazione
Via Ponzio 34/5, Milano, Italy
{name.surname}@polimi.it

ABSTRACT

Expert selection is an important aspect of many Web applications, e.g., when they aim at matching contents, tasks or advertisement based on user profiles, possibly retrieved from social networks.

This paper focuses on selecting *experts* within the population of social networks, according to the information about the social activities of their users. We consider the following problem: given an expertise need (expressed for instance as a natural language query) and a set of social network members, who are the most knowledgeable people for addressing that need? We considers social networks both as a source of expertise information and as a route to reach expert users, and define models and methods for evaluating people's expertise by considering their profiles and by tracing their activities in social networks. For matching queries to social resources, we use both text analysis and semantic annotation. An extensive set of experiments shows that the analysis of social activities, social relationships, and socially shared contents helps improving the effectiveness of an expert finding system.

Categories and Subject Descriptors

H.2.1 [**Logical Design**]: Data models; H.2.5 [**Heterogeneous Databases**]: Data translation; H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing; H.3.4 [**Systems and Software**]: User profiles; H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Algorithms, Design, Performance.

Keywords

Expertise finding, Web data, information retrieval, social network, crowdsourcing, semantic analysis, expert finding, user profile, information extraction.

EDBT/ICDT '13 March 18 - 22 2013, Genoa, Italy

1. INTRODUCTION

Involving crowds in performing tasks is an important aspect of modern Web-based systems and applications [21][14]. A lot of emphasis has been given so far to addressing generic crowds for micro-task assignment on platforms such as Amazon Mechanical Turk.[1] More recently, a new trend has emerged, consisting in using social networks as crowd platforms and asking questions to their members (i.e., crowd–searching) [2][6].

For certain tasks, selecting random workers on traditional crowdsourcing platforms is a good choice. For instance, if we are interested in locating the best prices of specific goods, crowd workers can search for cheap offers on online marketplaces and provide a list of advantageous websites for a small economic incentive. However, in many cases, routing queries to our social networks is a better solution. The main reason is trust; answers from trusted circles bear greater validity than answers from unknown workers. Social platforms, such as Facebook, Twitter and LinkedIn, easily provide their members with several hundreds of known contacts, with variable expertise about the various questions. These contacts can be easily reached by exploiting the connections built on top of the social platform; however, they are typically moved by non-monetary incentives, and, although generally responsive, they are not available on a continuous and demanding basis. Therefore, a careful selection of the *small crowd of the top-k experts* whom to ask questions is very relevant. Expert selection is an important aspect of many Web applications that use social networks as a platform, thanks to the recent availability of stable APIs that support their development.

In this application paper, we consider the problem of ranking the members of a social group according to the level of knowledge that they have about a given topic; after such ranking, the *top-k* experts are chosen. These experts can match very different needs, spanning: responses to factual questions (crowd–searching queries); providing recommendations upon products, people or places; or performing generic tasks (as in traditional crowdsourcing platforms).

The classic approach to this problem consists in profiling the group members, matching textual queries against such profiles, and ranking members according to the matching. However, profile information in most social networks may be quite limited, as many of their members give the smallest amount of information which is required for registering, and do not explicitly state all their interests and skills.

Our solution departs from the classic approach and takes

[1]https://www.mturk.com/

into account, beyond profile information, also the behavioral trace that users leave behind them through their social activities. The main result of this paper is the empirical demonstration of the greater contribution of activities of social network members with respect to their profiles for assessing the user expertise. We also found that certain profiles and activities of closest social contacts may provide useful information, thus giving a positive contribution to the expert ranking. As the content of information objects is of textual nature, we adopted standard information retrieval techniques for analyzing resources, by matching them according to their relevance w.r.t. the expertise need. In addition, an entity recognition and disambiguation activity identifies the real world entities respectively contained in the expertise need and in the resources, and such entities are also used in the matching.

We have focused on the most popular social networks: Facebook, LinkedIn, and Twitter. We have used the cumulative information from the three social networks to produce the ranking of their members, and then compared it with the rankings obtained by using them separately. Of course, social platforms are different both in their organization and content, hence their contribution to detecting experts also depends on the specific domain of the expertise need; thus, we have also compared the ability of the various social networks to rank their members with respect to expertise domain.

Our approach is illustrated in Figure 1. Consider Anna's query requiring expertise on free style swimmers; candidate responders are Alice, Charlie, Bob and Chuck, who are members of her social network. Their social activities, represented on the right side of the figure, comprise Alice's tweet on Michael Phelps's medal, Charlie's post about his training in freestyle, Bob's profile showing swimming as a hobby, and Chuck's friendship to Bob. Based on this ir
the system ranks Alice first, then Charlie, then
Chuck, while Peggy is not considered because sl
ther direct knowledge of the domain, nor close c
showing the requested expertise. Anna will th
her question according to the ranking (e.g., just t
to Alice and then Charlie, or to both of them a
time, and so on). The system comprises a moc
source analysis which extracts and indexes social
then an analysis module which extracts expertise
such resources, and a matching and ranking mod
compare the expertise need with the analyzed r
extract a list of experts; these modules are exte
scribed in Section 2.

We have recruited a group of 40 people which
on the considered social platforms; we created
interaction scenario, using the three social networ
their public APIs and extracting their resources a
the privacy settings of the involved users and the
Our approach can be adapted to different cont
as domain-specific social communities or compa
business management software.

The paper is organized as follows. Section
our formalization of the problem, and includes
meta-model of social networks for resource des
a a platform-independent way; Section 3 describ
tensive experiments and provides a discussion of
comes. Finally, Section 4 describes related work a
5 presents our conclusions.

2. FINDING EXPERTS IN SOCIAL NETWORKS

The rest of the paper assumes the abstract organization of the social expert finding system as in Figure 1. In the following, we formalize the problem of expertise retrieval inside social networks.

2.1 Definition

We call *candidate experts* **CE** the whole set of users registered in our system and potentially available for being assigned a task. This set consists of users residing in one or more social platforms and interacting with their peers by creating, consuming, or sharing *resources* (i.e. posts, status updates, comments, etc.), which we assume containing relevant clues about the candidate's expertise.

An *expertise need* q is an information need that relates with specific skills or knowledge. The expertise need can be stated as a natural language question, an unstructured document, or a structured set of information; an expertise need q refers to at least one *domain* of expertise, i.e., a field of interest.

Social expert finding addresses the following questions: *Given an expertise need q, who are the candidate experts most suited to address q? And which is the best social platform to contact them?*. Answering these questions requires, for each candidate $ce \in \mathbf{CE}$, the calculation of a measure of expertise $\mathbf{score}(\mathbf{q}, \mathbf{ce})$ that expresses the likelihood for the candidate to be able to contribute to the need q. The system returns a list $\mathbf{EX} \subseteq \mathbf{CE}$ of expert candidates that are deemed as suitable, ordered according to $\mathbf{score}(\mathbf{q}, \mathbf{ce}) > 0$.

2.2 Model of Social Resources

The key aspect of social expert finding is to collect evi-

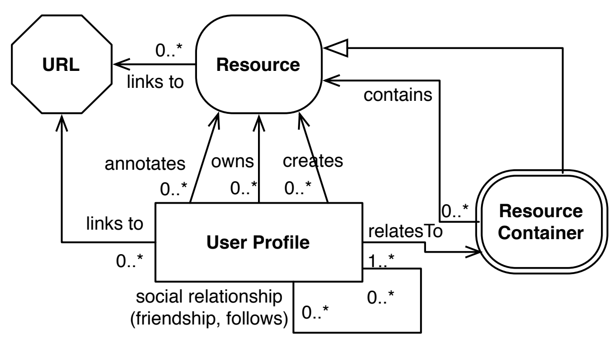


Figure 2: A simplified meta-model for social networks users, relationships, and resources.

Figure 2 depicts a simplified social graph meta-model for the description of expert candidates and resources in social networks. The content of the *User Profile* depends on the kind of social network. For example, the Twitter profile is less informative than the LinkedIn one, as the former contains only a short description of the user, while the latter may contain a detailed career description.

Each candidate is associated with a set of *Resources*, i.e. informative material that can be found inside a social plat-

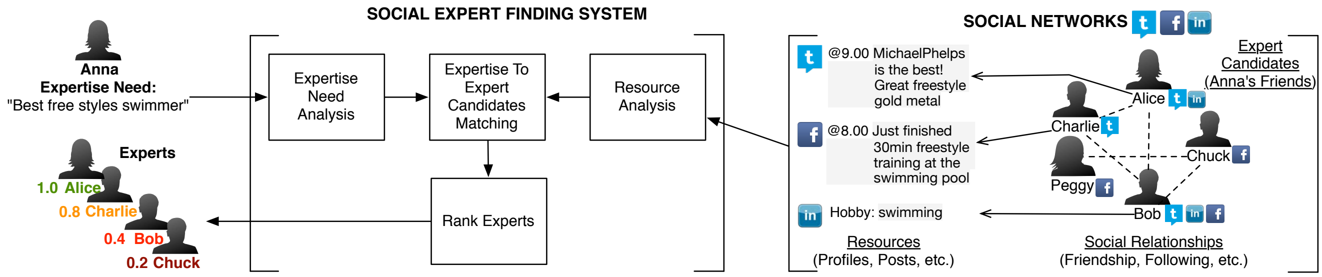


Figure 1: Approach to social expert finding

Distance	Resources
0	**Expert Candidate Profile**
1	**Expert Candidate** *owns/creates/annotates* `Resource`; **Expert Candidate** *relatedTo* `Resource Container`; **Expert Candidate** *follows* `User Profile`
2	**Expert Candidate** *relatedTo* `Resource Container` *contains* `Resource`; **Expert Candidate** *follows* `User Profile` *owns/creates/annotates* `Resource`; **Expert Candidate** *follows* `User Profile` *relatedTo* `Resource Container`; **Expert Candidate** *follows* `User Profile` *follows* `User Profile`

Table 1: The resources considered in this work, organized according to their distance with respect to an **Expert Candidate** `User Profile` in the social graph defined by the meta–model of Fig. 2.

form (e.g., Facebook status updates, Twitter tweets); resources can be organized in *Resource Container*s, i.e. logical aggregators of resources (e.g., Facebook or LinkedIn groups, Facebook pages) which are typically focused on a specific topic or real world entity (e.g., a group about Information Retrieval, or the Facebook page of Google); *Resource Container*s are typically described at least by a short textual description.

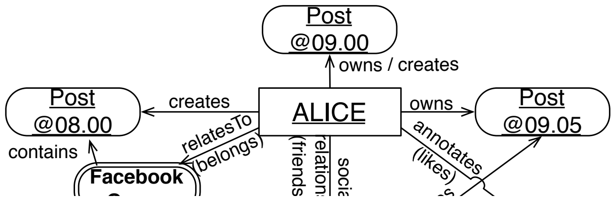
Profiles, resources and resource containers can include *URL*s to external Web pages, which we assume as containing information related to the profile, resource, or container. For instance, a candidate profile might contain a link to her Web page, which might be a good source of knowledge about the candidate's expertise; likewise, a link in a tweet will likely point to an external, related resource.

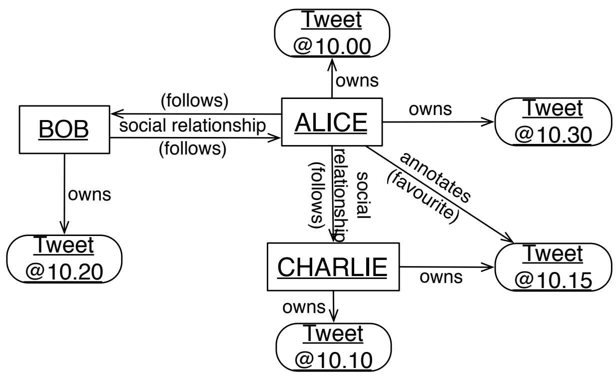
A resource can be directly or indirectly related to an expert candidate. *Directly related resources* are the ones *created* (e.g. a status update in Facebook, or a tweet in Twitter), or *annotated* (e.g. *Liked* in Facebook, or marked as Favorite in Twitter) by the candidate, regardless of their location. Examples of directly related resources are status updates in Facebook, tweets in Twitter, or posts in LinkedIn groups. We also consider directly related those resources created by other users, but *owned* by the candidate as they are published on the candidate's Facebook wall or Twitter/LinkedIn stream. *Indirectly related resources*, instead, are the ones included in *Resource Containers* related to the candidate, but created by other users (e.g. posts in a Facebook page "liked" by the candidate).

To characterize the relationship between the considered expert candidate and the related resources, we organize objects in the social graph according to their *distance* in the graph with respect to the candidate. Although the nature of the social graph would allow for its complete traversal (with upper bound equal to the longest shortest path between the candidate profile and any other node in the network), concrete issues of privacy, computational cost, and platform access constraints naturally limit the reach of the graph exploration. Considering a candidate expert *profile* as the initial node ($distance = 0$), we decided to take into account only resources up to $distance = 2$. Table 1 lists the resources considered in our work, organized according to their distance.

According to the targeted set of users and usage scenarios, each social network provides different set of features. While Facebook and LinkedIn are provided with groups and pages to allow people to express interest in specific domain of knowledge or expertise, Twitter lacks a similar tool. However, it is common practice in Twitter for users to *follow* individuals (e.g., Tim Berners Lee) and/or companies (e.g., Google) to receive updates on their activities and interests; therefore, *followed* users in Twitter can be assimilated to specialized, thematically focused external resources, such as Facebook Pages. We distinguish *followed* users from *friends* users by the absence (or presence) of a bidirectional social relationship between them. Figure 3 contains two examples of social graphs conforming to the meta-model of Figure 2, respectively for Facebook and Twitter: as in Facebook social relationship are bidirectional, Alice and Bob have a *friendship* relationship there; Alice and Bob are considered friends on Twitter too, since they mutually follow each other. On the other hand, in Twitter, Charlie is a *followed* user of Alice, while no relationship exists between them in Facebook.

We stress the difference between *followed* and *friend* users because bidirectional relationships typically reflects the existence of a real-world bond between individuals, which might not naturally imply shared interests or expertise. As we will see in Section 3, this is supported by empirical evidence too. Indeed, our experiments show that taking into account resources of *friend* users to characterize the expertise of an expert candidate does not increase the effectiveness of the expertise matching.





(b) Twitter

Figure 3: Examples of social network models for Facebook and Twitter, conforming to the meta-model of Fig. 2.

2.3 Expertise Need and Resource Processing

In order to perform the expert selection process, both the social resources related to expert candidates and the expertise need must be analyzed. The analysis is symmetrically performed on both needs and resources; therefore, in the following we describe only the analysis process for the resources (depicted in Figure 4).

The first step is the extraction of social data from the different platforms through their APIs:[2] we used the CrowdSearcher[3] [6] platform to collect users authentication tokens and privacy permissions; then, for each considered resource, the *Resource Extraction* module performed the analysis flow as described in the next paragraphs. According to the usage terms of the considered social platforms, no information has been stored on secondary storage systems.

The content of resources is normally composed by text, but they often include URLs linking newspaper articles, blog posts, or, more generally, external Web pages. As the information contained by such materials can be very useful to determine user expertise, the original textual content of the resources have been enriched with the content extracted from the linked Web pages.[4]

Then, as social network users can interact with textual resources of different languages, a *Language Identification* step allows the classification of resources according to their main language; such a classification is very important for the following *Text Processing* and *Entity Recognition and Disambiguation* steps, which are language dependent.

Text Processing deals with standard information retrieval preprocessing, such as sanitization, tokenization, stop word removal, and stemming. The *Entity Recognition and Disambiguation* step, deals with identifying concepts perceived by humans as a self-contained whole (e.g. people, organizations, places, etc.), and associating to them a unique interpretation in the context of the resource.

Entity recognition and disambiguation is typically performed by extracting named entities from text, and then enriching their descriptions with semantic information, such as the unique identifier of the entity (as taken, from instance, from Wikipedia), its type (e.g. Person, City, Sports Team, Athlete) and domain (e.g. tv, sports, education, business).

Several approaches to named entity extraction have exploited grammar-based techniques as well as statistical models (e.g., Stanford NER [11]). Recently emerging solutions rely on cross-linking text snippets to Wikipedia, FreeBase or similar Web Ontologies; given the nature of the analyzed resources, in this work we adopted the system for short text annotation described in [10], which identifies and disambiguates entities, returning a Wikipedia URI and a *disambiguation confidence* value for each entity in the text. This fine-grained analysis allows a more precise discrimination of entities that have a clear meaning in the context of the text in which they are contained, while penalizing ambiguous interpretations. The results will be used in the calculation of resource relevance in the following section.

2.4 Matching Expertise Needs to Candidate Experts

Our approach builds on the well-known vector space model, where resources, related entities, and expertise needs are represented in the same space. This simple model has a twofold advantage: one one hand, a uniform representation space enables the retrieval of relevant entities and related resources within the same expertise need, as resources are represented both as *bag-of-words* (according to the traditional model for textual search) and as set of entities. On the other hand, the vector space model can be easily extended so as to inject relevance evidence into resources and entities, at the purpose of influencing the retrieval process by including *a priori* knowledge.

Given an expertise need q and a set of resources $\mathcal{R}$, the set of relevant resources $\mathcal{RR}$ is produced by calculating the relevance of each resource $r \in \mathcal{R}$ as the weighted linear combination of the contribution of resource textual terms and entities. In more details, the score of a resource given a query q is:

$$
\begin{aligned}
score(q, r) =& \alpha \cdot \sum_{t \in q} \left(tf(t, r) \cdot irf(t)^2 \right) \qquad (1) \\
&+ (1-\alpha) \cdot \sum_{e \in E(q)} \left(ef(e, r) \cdot eirf(t)^2 \cdot w_e(e, r) \right)
\end{aligned}
$$

[2]Details and limitations regarding Facebook, Twitter and LinkedIn APIs are respectively available at `https://developers.facebook.com`, `https://dev.twitter.com`, and `https://developer.linkedin.com`.

[3] `http://crowdsearcher.search-computing.org/`.

[4]The task has been performed using the Alchemy Text Extraction API `http://www.alchemyapi.com/api/text/`.

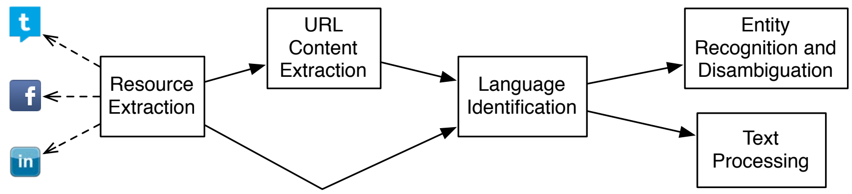
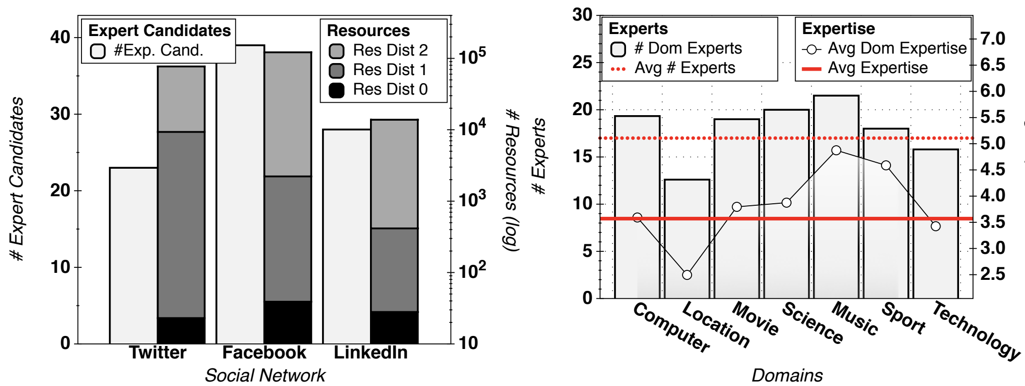


Figure 4: Schematic representation of the analysis process.



(a) (b)

Figure 5: Evaluation dataset: a) Distribution of resources and users among the considered social networks, and b) Distribution of experts and expertise in the considered domains.

where $\mathbf{t}$ are the terms in the expertise need $\mathbf{q}$, and $\mathbf{e} \in E(q)$ are the entities identified in the query by the *Entity Recognition and disambiguation* step. For each term $\mathbf{t}$, the functions $\mathbf{tf}(\mathbf{t}, \mathbf{r})$ and $\mathbf{irf}(\mathbf{t})$ respectively calculate the *term frequency* for the considered resource $\mathbf{r}$, and the *inverse resource frequency* of $\mathbf{t}$ in the whole resource collection. Likewise, for entity $\mathbf{e}$, the functions $\mathbf{ef}(\mathbf{e}, \mathbf{r})$ and $\mathbf{eirf}(\mathbf{e})$ respectively calculate the *entity frequency* for the considered resource, and the *inverse resource frequency* of $\mathbf{e}$ in the whole entity collection; $\mathbf{w_e}(\mathbf{e}, \mathbf{r})$ is weight that expresses the relevance of the entity $\mathbf{e}$ in a resource $\mathbf{r}$, and it is calculated as:

$$w_e(e, r) = \begin{cases} 1 + dScore(e,\, r) & \text{if } dScore(e,\, r) \geq 0 \\ 0 & \text{if } dScore(e,\, r) = 0 \end{cases} \qquad (2)$$

where $dScore(e,\, r)$ is a measure of disambiguation confidence for the entity $\mathbf{e}$ in a resource $\mathbf{r}$, as calculated in the *Entity Recognition and Disambiguation* step.

Finally, α is weighting factor that denotes the importance of the contribution provided by keyword matching and entity matching in the calculation of the resource score. Thanks to the parameter α, we can vary the importance given to textual term matching and entity matching, thus balancing their contribution in the evaluation of the relevance score of resources. In Section 3.4 we assess the importance of textual term matching w.r.t. entity matching through different experiments.

2.4.1 Ranking Experts

Given the set $\mathcal{RR}$ of resources retrieved by the *Social Resources Matching* step, we identify the set: $\mathbf{EX} = \{\mathbf{ce_1}, ..., \mathbf{ce_m}\}$ of candidate experts as the set of social network users related with the relevant resources. The `Ranking Experts` step orders experts in $\mathbf{EX} \subseteq \mathbf{CE}$ according the following expertise scoring function, which also considers the relevant resources associated to the candidate experts:

$$score(q,\, ex) = \sum_{r_i \in \mathcal{RR}} score(q,\, r_i) \cdot w_r(r_i,\, ex) \qquad (3)$$

where $\mathbf{ex} \in \mathbf{EX}$ is an expert, i.e., a candidate expert with associated relevant resources, and $\mathbf{w_r}(\mathbf{r_i}, \mathbf{ex})$ is a weighting term that quantifies how the expertise inferred from the resource can be associated to the expert. Differently from traditional works in expert finding [18], where the relationships between users and resources have to be extracted from the document text, in the social networks setting these connections are explicitly identifiable. Considering the social graph meta-model in Figure 2, resources are weighted according to their *distance* in the graph with respect to the expert $\mathbf{ex}$.

The number of relevant resources $\mathcal{RR}^*$ in in the set of retrieved resources $\mathcal{RR}$ depends both on the "popularity" of the requested expertise in the considered user base, and on the reach of the social graph exploration; such a number is expected to affect the performance of the expert finding system.

As we assume a direct correlation between the number of resources related to a query, and the potential expertise of the user, no normalization on the number of retrieved resources is included. However, as the number of matching resources can be high, a `window size` parameter defines the number of considered relevant resources. In Section 3.4 we also provide an analysis of the influence of this parameter.

3. EXPERIMENTAL EVALUATION

In this section we first describe the dataset, metrics, and parameter setting that has been used for experiments; then we present a large number of experiments and discuss their results.

3.1 Dataset

To the best of our knowledge, no dataset targeting the problem of expert finding in social networks exists; some datasets exist in the context of Enterprise Information Retrieval, but they address a significantly different content types. Therefore, we devised a set of 30 expertise needs formulated as textual queries, spanning over the following domains: *computer engineering*, *location*, *movies & tv*, *music*, *science*, *sport*, and *technology & videogames*. Examples of queries for each domain are: *a*) **Computer engineering**: *Which PHP function can I use in order to obtain the length of a string?*; *b*) **Location**: *Can you list some restaurants in Milan?*; *c*) **Movies & tv**: *Can you list some famous actors in how I met your mother?*; *d*) **Music**: *Can you list some famous songs of Michael Jackson?*; *e*) **Science**: *Why is copper a good conductor?*; *f*) **Sport**: *Can you list some famous European football teams?*; *g*) **Technology & videogames**: *I am looking for a graphic card to play Diablo 3 but I don't want to spend too much. What do you suggest?*.

We recruited volunteers by advertising our experiment on public social networks (Facebook, LinkedIn, and Twitter) and known groups. Forty people, active on the three social networks, accepted to participate in our experiment. Through these volunteers, we were able to collect around 330,000 information resources, among which 70% contained a URL pointing to an external Web page. We only considered resources containing English text, which summed up to 230,000 items.

Figure 5a represents the distribution of resources and expert candidates among the considered social networks. For each expert candidate, we retrieved the full set of available (i.e., accessible according to platform limitations and privacy permissions) profile information and the full set of created and annotated resources. For each resource container we retrieved the most recent resources. As expected, Facebook is the social network that features the highest number of resources (wall posts, group posts, etc.). Twitter provided less resources w.r.t. Facebook, a result accountable to the simpler structure of the social network, which offers only user profiles and tweets as processable objects; however, Twitter provides the highest number of resources at distance = 1, which include the tweets of the current user and the profiles of the followed users. The lower amount of LinkedIn resources is explained by the nature of the platform, which being job-related, provides less incentives to users for general-purpose interaction and content publishing. 95% of the LinkedIn resources were posts on groups (distance level 2), while only few users contributed with status updates (or did it through cross–social network posting, e.g., from Twitter; in our experiment we ignored this kind of updates, because they were already accounted for in the other social network).

We created the ground-truth for our experiments by asking the 40 expert candidates to perform a self-assessment questionnaire, where they were asked to rate their expertise in each of the 30 expertise needs in a 7-point likert scale. Since each need referred to a domain, based on the users' self–assessment, we derived the level of expertise in the 7 domains of interest. We considered *domain experts* only those having a level of expertise higher than the average expertise of that domain. Therefore, the expertise matching is a boolean function (0 for experts below average and 1 for experts above average). Figure 5b depicts the resulting distribution of experts and expertise in the considered domains: on average, each domain featured 17 experts, with an average expertise level of 3.57.

In all our experiments we compare the system performance with a random baseline. Random figures have been calculated by averaging, for each query, the results of 10 runs in which 20 users were randomly selected.

3.2 Metrics

Each experiment on a given expertise need (query) computes a list of candidate experts; such result is compared to a list of domain experts on the query domain, constructed on the basis of the self-assessment questionnaire; several standard retrieval metrics provide an evaluation of how the former list approximates the latter, taken as ground-truth. We used: i) **Mean Average Precision** (MAP), ii) **11-Point Interpolated Average Precision** (11-P curve), iii) **Mean Reciprocal Rank** (MRR), and iv) **(Normalized) Discounted Cumulative Gain**, (NDCG). We decided to adopt different metrics because each of them describes different aspects of the system behavior. While MAP and 11-P curve provide a compact measure of the precision of the retrieval capability, MRR and (N)DGC measure the ability of the system to retrieve highly relevant users at high positions in the result set. In particular, while MMR gives a clear intuition of the behavior of the system for the first retrieved items, NDCG@10 is very well suited for understanding the perceived quality of the first 10 retrieved results.

3.3 Influence of Parameters

As the model includes several parameters and is based upon critical assumptions, we first assessed their impact on the method effectiveness. We evaluated the contribution of: *a*) the `window size`, which defines the number n of relevant resources to consider for user ranking; *b*) the α value, which dictates the contribution of entities vs. text analysis in the calculation of the relevance score of a resource; and *c*) the inclusion of resources belonging to expert candidates' friends. In all the following experiments, we fixed the weighting terms $w_r(r_i, ce)$ in an interval $[0.5, 1]$, with value linearly decreasing w.r.t. the distance of the considered resource.

3.3.1 Window size

First, we assessed the effect of the window size parameter; Figure 6 depicts the variations of the MAP, MRR, NDCG and NDCG@10 metrics for increasing window sizes, considering up to 10% of the matching resources. The test has been conducted with resources at distance 1 and distance 2, and setting the value of the α parameter at 0.5, thus giving equal importance to textual terms and entities.

Increasing the number of resources, regardless of the resource distance, resulted in an increase in MAP and NDCG. For distance 2 resources, increasing the window size leads to increasing MAP and NDCG up to 30%, when considering in the expert ranking the 10% of matching resources. On the other hand, NDCG@10 and MRR curves have a different

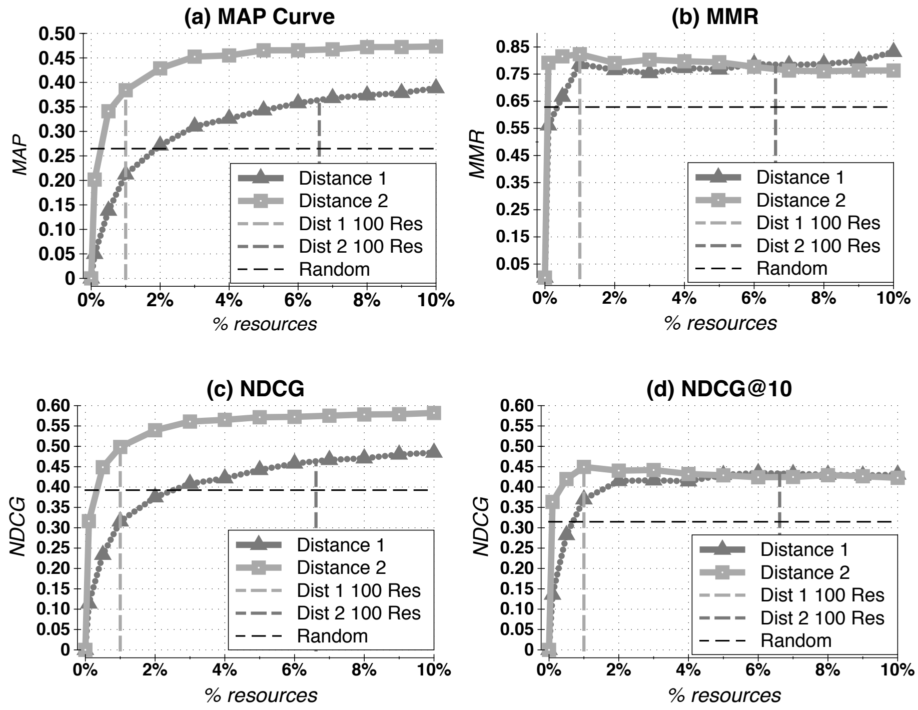


Figure 6: Evaluation metrics at different window sizes. (a) MAP, (b) MMR, (c) NDCG, (d) NDCG@10.

behavior, as both metrics do not appear to be significantly affected by the increased number of considered resources.

This behavior can be intuitively explained by the presence of a relatively small percentage of resources that determine the top experts in the candidate pool; the addition of new resources marginally improves the discovery of such top experts, but it increases the overall number of retrieved experts, thus improving the values of MAP and NDCG metrics. Based on these experiments, we set the *window size* to 100 resources (identified in Figure 6 with dashed vertical lines), roughly corresponding to the 6% of resources retrieved on average when querying resources at distance 1, and 1% of resources at distance 2.

3.3.2 Relevance of Textual Terms and Entities

Next, we performed a sensitivity analysis on the α parameter. Figure 7(a,b,c) respectively depicts the variation of the MAP, MMR, NDCG, and NDCG@10 metrics for increasing values of α. With resources at distance 0, the evaluation of resource relevance only with entities ($\alpha = 0$) greatly decrease the effectiveness of the system; such a result can be justified by the low amount of information that can be collected in expert candidate profiles which, in turn, leads to a low number of entities and to poor disambiguation performance. Higher values of alpha generally provide increased metrics, with better results with resources at distance 2. Based on this analysis, metrics appear to be stable for values of the α parameter in the $[0.3, 0.8]$ interval. For all the subsequent experiments we set $\alpha = 0.6$.

3.3.3 Relevance of Friendship Relations

In Section 2.2 we motivated the exclusion of friend's resources from our analysis (e.g., Twitter resources which are retrieved by means of symmetric *follows* relationships). In this section we assess the impact of friends' resources on Twitter, which is the most open platforms in terms of resource access. Considering, for instance, Facebook friends would be anyway impossible, as only 80 (0.6%) out of the 13K friends of the 40 Facebook account had a privacy setting allowing us to retrieve their profile and social activities information.[5]

Table 2 and Figure 8 assess the correctness of our choice; they report the results of the comparison for resources of Twitter friends at distance 1 and distance 2 (with *window size* = 100 and $\alpha = 0.6$). Although a considerable amount (60,000) of additional resources were analyzed, they did not produce significant improvement: at distance 1, introducing friends brought a modest 1% increase in all the considered measures, while at distance 2 it slightly worsened average precision and NDCG. The results suggest that the addition of Twitter friends would give no particular benefit.

3.4 Contribution of Resource Distance

In this section we evaluate how resources at the various distances contribute to the system performance. Table 3 (All) and Figure 9 summarize the obtained results. By taking into account only the expert candidate profiles (resources at distance 0) we obtain the worst measures, which are lower than random selection, suggesting that profiles alone are inadequate to the task of expert selection. The addition of resources at distance 1 significantly improves all the metrics, which reach their maximum by adding resources at distance

[5]We are aware of a Facebook "Subscription" functionality, but, as of August 7th 2012, the corresponding Facebook APIs are not documented, and the use of the functionality is limited.

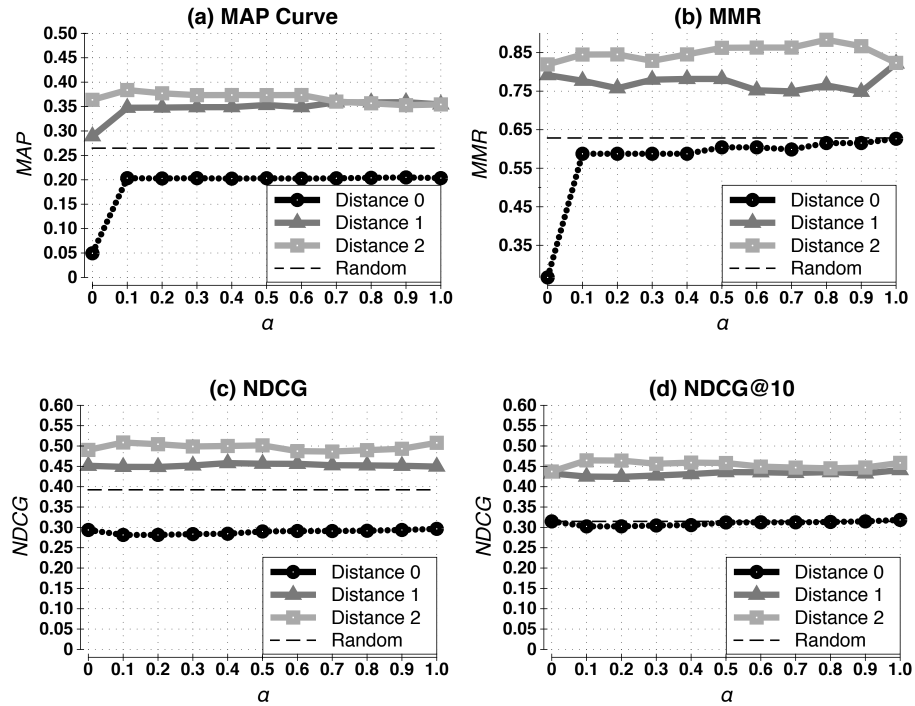


Figure 7: Analysis on the α parameter. (a) MAP, (b) MMR, (c) NDCG, (d) NDCG@10.

Model		**Metrics**			
Dist.	Friend	**MAP**	**MRR**	**NDCG**	**NDCG@10**
Random		.2648	.6285	.3924	.3147
1	N	.3742	.7716	.4318	.4405
	Y	**.3844**	**.7833**	**.4576**	**.4625**
2	N	**.4708**	.6744	**.5390**	.4630
	Y	.4390	**.7555**	.5249	**.4769**

Table 2: Comparison of the results obtained on Twitter considering *Friend* relationships

Model		**Metrics**			
SN	Dist.	**MAP**	**MRR**	**NDCG**	**NDCG@10**
Random		.2648	.6285	.3924	.3147
All	0	.2023	.5875	.2843	.3055
	1	.3488	.7816	.4580	.4310
	2	**.3736**	**.8453**	**.5001**	**.4592**
FB	0	.0478	.3444	.0733	.0893
	1	.3682	.8055	.5071	.4377
	2	.2877	**.8408**	.4245	.4607
TW	0	.0600	.5777	.1257	.1529
	1	.3742	.7716	.4318	.4405
	2	**.4708**	.6744	**.5390**	**.4630**
LI	0	.1623	.6638	.2519	.2787
	1	.2607	.7166	.3676	.3394
	2	.3051	.7205	.4408	.3501

Table 3: Comparison of the results obtained with *All* the social networks, or separately by ***FaceBook***, ***TWwitter***, and ***LinkedIn***. In bold the best results for each evaluation metric.

2, well above the random selection configuration. These results confirm that static profiles do not provide sufficient expertise information, and that adding social behavior helps to reach better results.

3.5 Contribution of Social Networks

Next, we assess the contribution of each social network separately considered, again measured according to the three distances of resources. As shown in Table 3, Twitter proved better suited to expertise extraction, although Facebook features the best MRR figure. Surprisingly, the use of Twitter alone at level 2 outperforms the use of all the social networks on three metrics out of four. LinkedIn proved worse than other social networks in all cases.

3.6 Domain Specific Experiments

By considering the effect of social networks on domain–specific results, further interesting considerations can be drawn. Table 4 presents the breakdown of the evaluation metrics for each considered domain, and for each social network. Also in these experiments, Twitter is associated with the highest figures in computer engineering, science, sport, and technology & games, and achieves good figures in all domains.

3.7 Trustworthiness of Social Information

To allow an expert finding technique based on resources to work, it's crucial that the considered resources must reflect correctly and completely the expertise of the related expert candidates. Unfortunately, in social networks, this hardly occurs in a perfect manner, as demonstrated by the poor performance (see Table 4) obtained by domains such as *Sports* or *Music*: a rather large set of expert candidates in our pool declared themselves experts, nonetheless the sys-

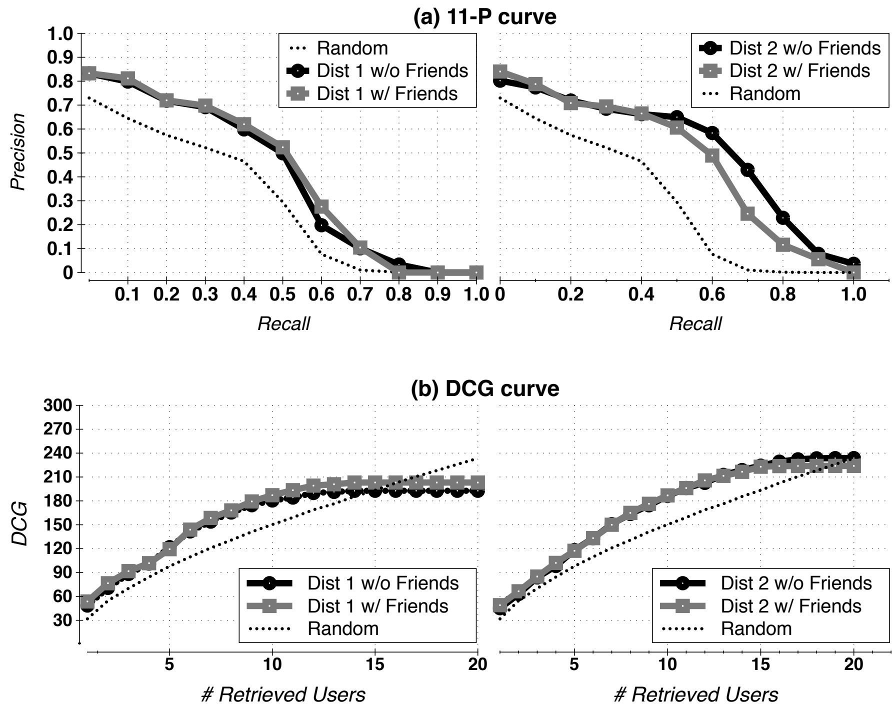


Figure 8: (a) MAP and (b) DCG for Dist = 1 and Dist = 2 with and without *Friend* user resources on Twitter.

tem effectiveness was limited. The reasons for this results are quite obvious: if a user claims to be super–expert in music but then neither her profile nor any of her social actions include music, then there is no possible method that will extract that user as a music expert.

Facebook is associated with high metric figures in location, music, sport, and movies & tv; it also produces good rankings, as the MMR and NDCG@10 measures are sometimes better than those obtained with Twitter. Facebook figures are much lower in domains such as computer engineering and science: this is consistent with the platform scope, as it is quite common on Facebook to write about entertainment-related topics, while it is less likely to read about medicine studies or electrical conductors.

LinkedIn has lower figures in all domains (including computer engineering) and overall; however, the metric figures for computer engineering at distance 0 are quite high. This is because LinkedIn profiles contain accurate descriptions of expert candidates skills and work experience, and thus are good sources for inferring expertise in work-related domains; LinkedIn also achieves a good precision in the science domain at resource distance 2.

Domain-specific results are influenced by the nature of dataset, and specifically with the expertise distribution depicted in Figure 5b. Domains like *Computer Engineering* and *Sport*, which have the best overall results, also feature a good number of experts; on the other hand, a domain like *science*, which is also supported by a good number of experts, has worse results (up to -22% MAP). Such a behavior can be explained by a lower amount of resources associated with the domain, i.e. people hardly write about biology or medicine in their social walls. A different problem occurred for the *Location* domain: although a considerable amount of resources (especially user profiles) contained geographically-related information, few expert candidates considered themselves sufficiently skilled in the domain; consequently, the list of expert candidates descending from the ground-truth was smaller. The low number of experts combined with the widespread presence of location information for all the candidates, made it harder for the system to pinpoint the right experts. This result calls for domain-specific solutions for location related expertise needs.

We therefore analyzed the performance of each expert candidate against the expert needs in the dataset, to assess how often their expertise was accurately estimated by the system. Figure 10 depicts, the F1-score for each user: 6 candidates obtained a value greater than 0.70; 8 candidates were deemed completely unreliable; and half of them had an F1-score above the average. Notice that there is a clear correlation between the number of available resources and the ability of the system to predict the user expertise. Although satisfactory, the results in Figure 10 are indicative of the fact that users do not completely expose their own interests and expertise on social networks. It is important to notice that such an omission can be explicit or implicit: while some users have social network account for flagship or promotional reasons (thus heavily limiting the number and scope of published data), other might apply strict rules to protect their privacy. However, notice that privacy policies are a limitation only for third-party applications (such as the ones we used to collect resources), while social-networks owners are able to access the full user information (and therefore would not be limited in case they want to apply expertise matching themselves). To provide some evidence to our hypothesis, Figure 10 contains also a regression on the number of resources published by each expert candidate: not sur-

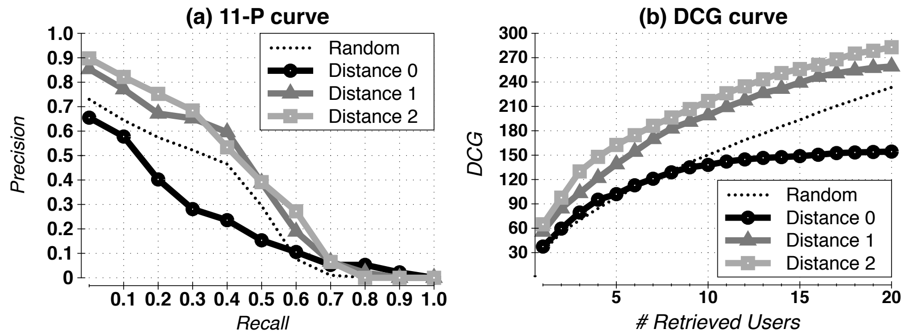


Figure 9: a) Interpolated 11-Point Precision/Recall and b) DCG curves considering resources of all the social networks.

Domain	**Dist.**	**MAP**				**MMR**				**NDCG@10**			
		All	FB	TW	LI	All	FB	TW	LI	All	FB	TW	LI
Computer engineering	0	**.5474**	.0671	.1351	.5564	**1**	.3333	**1**	.8333	**.6543**	.1319	.3370	.6335
	1	.3681	.3504	.4344	.5830	**1**	**1**	**1**	**1**	.4946	.4606	.5447	.5731
	2	.5052	.2356	**.7104**	.4992	**1**	**1**	**1**	.8333	.6387	.5031	**.7346**	.5048
Location	0	.2907	.1329	.1876	.3226	.5952	.4000	.7000	.7500	.4318	.2169	.3154	.4029
	1	.3733	.3594	.3860	.3300	.8666	**.9000**	.7066	.6666	.5223	**.4970**	.4099	.4811
	2	.2695	.1965	**.4234**	.2381	.7222	.6952	.5833	.6833	.4282	.4455	.4893	.2908
Movies & TV	0	.0796	0	.0683	.0947	.4900	0	.3000	5166	.1628	.0006	.1173	.1361
	1	.2882	.3239	.3092	.0968	.7666	.8000	**.8666**	.5666	.3848	.3890	**.4454**	.1823
	2	.3541	.3126	**.4014**	.1730	.8000	.8000	.6900	.3200	.4198	.3992	.4206	.3280
Music	0	.1109	.0250	.1527	.0714	**1**	.5000	**1**	**1**	.3649	.1100	.3669	.2536
	1	.2913	.3770	.4498	.0767	.4166	.5000	**1**	.625	.3010	.4152	.4879	.2803
	2	.3971	.3639	**.5008**	.3333	**1**	**1**	.6666	**1**	.4379	**.5117**	.4355	.4676
Science	0	.0513	.0185	.0208	.0521	.0833	.3333	.1666	.0833	.0506	.0161	.0256	.0516
	1	.2524	.1907	**.4229**	.1054	.7500	.3888	.7333	.4444	.3552	.2051	**.4192**	.0880
	2	.3201	.2437	.4192	.4126	.7500	**.7500**	.5000	.7777	.3609	.3399	.3780	.4592
Sport	0	.2249	.0862	.0799	.2280	.7222	.5555	.5833	.8055	3741	.2059	.1274	.3570
	1	.4608	.4695	.3660	.2923	**1**	**1**	.7416	.8333	.5847	**.6251**	.4261	.4271
	2	.3061	.3005	**.4934**	.2325	.9167	.8333	.6722	.8888	.5430	.5529	.4408	.4194
Technology & games	0	.1923	.0572	.0801	.1971	.4566	.4000	.5666	.6000	2700	.0773	.1087	.3456
	1	.3476	.3832	.3456	.2915	.5400	.0800	.0500	.7500	.3387	.3847	.3731	.3639
	2	.3670	.2727	**.4352**	.2395	.8000	**1**	.5655	.6866	.3571	**.4098**	.4032	.2336

Table 4: Evaluation metrics split for each domain, and for each domain/social network.

prisingly, a correlation may exist between the number of published resources and the corresponding expertise assessment quality. Figure 11 depicts the difference Δ between the number of experts retrieved by the system and the expected number of experts as defined in the ground-truth, for each query and for each resource distance. The graphs clearly show a correlation between the amount (depending on the distance) of considered resources, and the ability of the system to retrieve experts; notice that at distance 2, one third of questions are under-represented, while 5 questions are clearly over-represented, thus showing space for further improvement. A more complete analysis of such correlation is out of the scope of this paper, and it is therefore left to future work.

4. RELATED WORK

Finding a list of people which possess a given set of skills, or that are knowledgeable about a given topic is a widely studied problem, typically known as the *expert retrieval* problem. In 2005, TREC introduced the *Expert Finding* task, which involved the analysis of an enterprise data corpus (an email archive) and the retrieval of a set of people that are experts in a given topic. Major contributions to the field are due to [3], where the author proposes a document-centric method based on a probabilistic Bayesian approach. DeMartini et al. [9] introduces a model for retrieving and ranking entities and its application to expert finding. Although several works address user profiling on social networks [17][1], to the best of our knowledge, our work is the first that applies resource–based methods [3][9] to the context of social networks, and performs and extensive analysis of the performance on different online social platforms. Some works investigated the relation between expert finding and document retrieval [18], by applying metrics similar to ours. While the conclusions of [18] are in line with our findings, our work is characterized by the exploitation of social relations in addition to document relevance.

Another characterization of the expert finding task is the "Expert Team Formation Problem": [15] describes an approach which uses social relationships between individuals, and the total communication cost as the optimization term

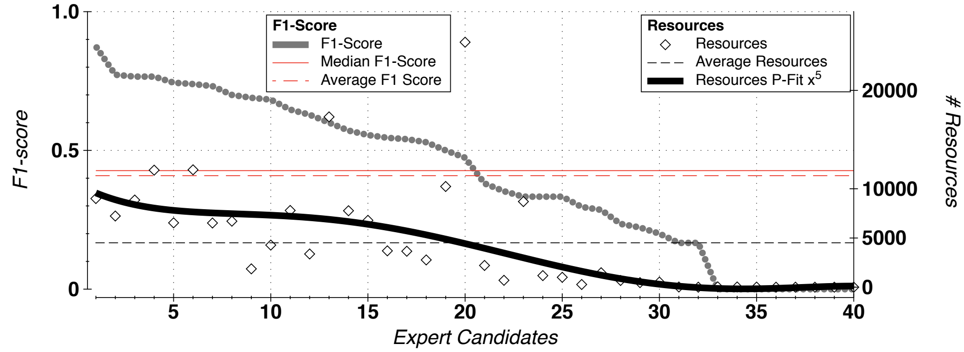


Figure 10: Relationship between users' expertise and available social information.

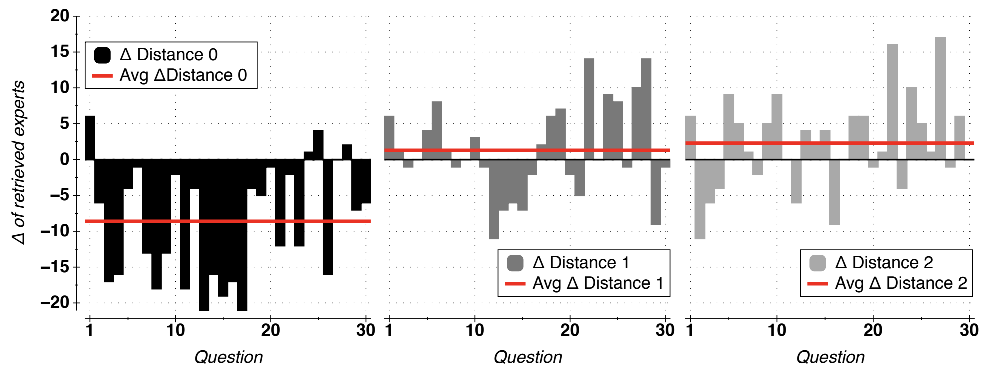


Figure 11: Differential number of retrieved experts.

of the objective function. [8] addresses the Jury Selection Problem by exploiting micro-blog services (e.g., Twitter) to solve decision-making tasks. The authors describe two models for selecting jury members that minimize the overall decision making error–rate; they estimate the error rate of each member by analyzing a Twitter graph of 690K nodes.

The problem of expert finding in online communities can also be targeted to blogs and forums: since the first 2006 TREC Blog track [4], researchers used social network and link analysis methods to identify experts according to their produced contents [7], interaction dynamics [23][22], authoritativeness [5][19], or question selection preferences [20]. [24] tackles the expert finding task in the context of an academic researcher network.

Several works focus on question–answering systems [12][5] [19][25][20], with the purpose of identifying the best community members able to answer a given question. The Aardvark social question answering system [13] used a statistical model to route questions to potential answerers. Similarity to our work, Aardvark exploited the profiles and activities on social networks to infer topic-related expertise of users; however, the analysis was confined only to the social information of the potential answerers, without considering their social relationships. [16] addresses the problem of routing questions in Yahoo! Answer by building a performance profile for each user based on his previous answers. While the problem definition is similar, this paper differs because it considers social activities as the source of expertise description, thus considering very different resources and context.

5. CONCLUSIONS

The selection of expert responders to query and recommendation tasks is increasingly relevant; this paper has explored how to effectively use the behavioral trace people leave when interacting on social platforms, in order to match and rank their expertise against given needs.

For assessing people expertise, we found that: **(1)** profile information is generally less effective than information about resources that they directly create, own or annotate; **(2)** resources which are produced by others (resources appearing on the person's Facebook wall or produced by people that she follows on Twitter) help increasing the assessment precision; **(3)** Twitter appears the most effective social network for expertise matching, as it very frequently outperforms all other social networks (either combined or alone); **(4)** Twitter appears as well very effective for matching expertise in domains such as computer engineering, science, sport, and technology & games, but Facebook is also very effective in fields such as locations, music, sport, and movies & tv; **(5)** surprisingly, LinkedIn appears less effective than other social networks in all domains (including computer science) and overall. These findings are the outcome of a very laborious method which combines text matching, entity extraction and disambiguation, and uses a variety of metrics; we also provided an analysis describing how we have set the various parameters required by the method.

This research is directly applicable to human and social computation systems, and recommendation systems upon social networks. The results are being integrated within the Crowdsearcher platform for further evaluations [6].

6. ACKNOWLEDGMENTS

This research is funded by the Search Computing (SeCo) ERC Advanced Grant. See more at: `http://www.search-computing.org`.

7. REFERENCES

[1] A. Alkouz, E. W. D. Luca, and S. Albayrak. Latent semantic social graph model for expert discovery in facebook. In *IICS*, pages 128–138, 2011.

[2] R. A. Baeza-Yates, S. Ceri, P. Fraternali, and F. Giunchiglia, editors. *Proceedings of the First International Workshop on Crowdsourcing Web Search, Lyon, France, April 17, 2012*, volume 842 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2012.

[3] K. Balog. *People Search in the Enterprise*. PhD thesis, University of Amsterdam, June 2008.

[4] K. Balog, M. de Rijke, and W. Weerkamp. Bloggers as experts: feed distillation using expert retrieval models. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '08, pages 753–754, New York, NY, USA, 2008. ACM.

[5] M. Bouguessa, B. Dumoulin, and S. Wang. Identifying authoritative actors in question-answering forums: the case of yahoo! answers. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '08, pages 866–874, New York, NY, USA, 2008. ACM.

[6] A. Bozzon, M. Brambilla, and S. Ceri. Answering search queries with crowdsearcher. In *Proceedings of the 21st World Wide Web Conference (WWW 2012)*, pages 1009–1018, 2012.

[7] C. S. Campbell, P. P. Maglio, A. Cozzi, and B. Dom. Expertise identification using email communications. In *In CIKM '03: Proceedings of the twelfth international conference on Information and knowledge management*, pages 528–531. ACM Press, 2003.

[8] C. C. Cao, J. She, Y. Tong, and L. Chen. Whom to ask? jury selection for decision making tasks on micro-blog services. *Proceedings of the VLDB Endowment*, 5(11):1495–1506, 2012.

[9] G. Demartini, J. Gaugaz, and W. Nejdl. A vector space model for ranking entities and its application to expert search. In M. Boughanem, C. Berrut, J. Mothe, and C. Soule-Dupuy, editors, *Advances in Information Retrieval*, volume 5478 of *Lecture Notes in Computer Science*, pages 189–201. Springer Berlin / Heidelberg, 2009.

[10] P. Ferragina and U. Scaiella. Tagme: on-the-fly annotation of short text fragments (by wikipedia entities). In *Proceedings of the 19th ACM international conference on Information and knowledge management*, CIKM '10, pages 1625–1628, New York, NY, USA, 2010. ACM.

[11] J. R. Finkel, T. Grenager, and C. Manning. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 363–370, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics.

[12] J. Guo, S. Xu, S. Bao, and Y. Yu. Tapping on the potential of q&a community by recommending answer providers. In *CIKM*, pages 921–930, 2008.

[13] D. Horowitz and S. D. Kamvar. The anatomy of a large-scale social search engine. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 431–440, New York, NY, USA, 2010. ACM.

[14] P. G. Ipeirotis and P. K. Paritosh. Managing crowdsourced human computation: a tutorial. In *Proceedings of the 20th international conference companion on World wide web*, pages 287–288, 2011.

[15] T. Lappas, K. Liu, and E. Terzi. Finding a team of experts in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '09, pages 467–476, New York, NY, USA, 2009. ACM.

[16] B. Li, I. King, and M. R. Lyu. Question routing in community question answering: putting category in its place. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, CIKM '11, pages 2041–2044, New York, NY, USA, 2011. ACM.

[17] H. Liu. Social network profiles as taste performances. *J. Computer-Mediated Communication*, 13(1):252–275, 2007.

[18] C. Macdonald and I. Ounis. The influence of the document ranking in expert search. In *18th ACM conference on Information and knowledge management*, CIKM '09, pages 1983–1986, New York, NY, USA, 2009. ACM.

[19] A. Pal and S. Counts. Identifying topical authorities in microblogs. In *WSDM*, pages 45–54, 2011.

[20] A. Pal, F. M. Harper, and J. A. Konstan. Exploring question selection bias to identify experts and potential experts in community question answering. *ACM Trans. Inf. Syst.*, 30(2):10, 2012.

[21] A. J. Quinn and B. B. Bederson. Human computation: a survey and taxonomy of a growing field. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 1403–1412, New York, NY, USA, 2011. ACM.

[22] J. Seo and W. B. Croft. Thread-based expert finding. In *Proceedings of the SIGIR'09 SSM Workshop*, Boston, Massachusetts, USA, July 2009.

[23] J. Zhang, M. S. Ackerman, and L. Adamic. Expertise networks in online communities: structure and algorithms. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 221–230, New York, NY, USA, 2007. ACM.

[24] J. Zhang, J. Tang, and J. Li. Expert finding in a social network. In R. Kotagiri, P. Krishna, M. Mohania, and E. Nantajeewarawat, editors, *Advances in Databases: Concepts, Systems and Applications*, volume 4443 of *Lecture Notes in Computer Science*, pages 1066–1069. Springer Berlin / Heidelberg, 2007.

[25] H. Zhu, H. Cao, H. Xiong, E. Chen, and J. Tian. Towards expert finding by leveraging relevant categories in authority ranking. In *CIKM*, pages 2221–2224, 2011.