Social Networks and Information Retrieval, How Are They Converging? A Survey, a Taxonomy and an Analysis of Social Information Retrieval Approaches and Platforms

Mohamed Reda Bouadjenek, Hakim Hacid, Mokrane Bouzeghoub

▶ **To cite this version:**

Mohamed Reda Bouadjenek, Hakim Hacid, Mokrane Bouzeghoub. Social Networks and Information Retrieval, How Are They Converging? A Survey, a Taxonomy and an Analysis of Social Information Retrieval Approaches and Platforms. Information Systems, 2016, 56, pp.1-18. 10.1016/j.is.2015.07.008 . lirmm-01174843v4

HAL Id: lirmm-01174843

https://hal-lirmm.ccsd.cnrs.fr/lirmm-01174843v4

Submitted on 3 Oct 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Information Systems 56 (2016) 1–18

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/infosys

Social networks and information retrieval, how are they converging? A survey, a taxonomy and an analysis of social information retrieval approaches and platforms

Mohamed Reda Bouadjenek [a,b,*], Hakim Hacid [c], Mokrane Bouzeghoub [d]

[a] INRIA & LIRMM, University of Montpellier, France
[b] Department of Computing and Information Systems, University of Melbourne, Parkville 3010, Victoria, Australia
[c] Zayed University, United Arab Emirates
[d] PRiSM Laboratory, University of Versailles-Saint-Quentin-en-Yvelines (UVSQ), France

ARTICLE INFO

Article history:
Received 18 October 2014
Accepted 9 July 2015
Available online 28 August 2015

Keywords:
Information Retrieval
Social networks
Social Information Retrieval
Social search
Social recommendation

ABSTRACT

There is currently a number of research work performed in the area of bridging the gap between Information Retrieval (IR) and Online Social Networks (OSN). This is mainly done by enhancing the IR process with information coming from social networks, a process called Social Information Retrieval (SIR). The main question one might ask is *What would be the benefits of using social information (no matter whether it is content or structure) into the information retrieval process and how is this currently done?*

With the growing number of efforts towards the combination of IR and social networks, it is necessary to build a clearer picture of the domain and synthesize the efforts in a structured and meaningful way. This paper reviews different efforts in this domain. It intends to provide a clear understanding of the issues as well as a clear structure of the contributions. More precisely, we propose (i) to review some of the most important contributions in this domain to understand the principles of SIR, (ii) a taxonomy to categorize these contributions, and finally, (iii) an analysis of some of these contributions and tools with respect to several criteria, which we believe are crucial to design an effective SIR approach. This paper is expected to serve researchers and practitioners as a reference to help them structuring the domain, position themselves and, ultimately, help them to propose new contributions or improve existing ones.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

With the emergence of the social Web, the Web has evolved from a static Web, where users were only able to consume information, to a Web where users are also able to produce information. This evolution is commonly known as Social Web or Web 2.0. Thus, the Web 2.0 has introduced a new freedom for the user in his relation with the Web by facilitating his interactions with other users who have similar tastes or share similar resources. Social platforms and networks (such as *MySpace*,[1] *Facebook*,[2] and *LinkedIn*[3]), collaborative tagging sites (like *delicious*,[4] *CiteULike*,[5] and *Flickr*[6]), and microblogging sites (like *Twitter*[7] and *Yammer*[8]) are

* Corresponding author. Tel.: +33 6 43 83 76 91.
E-mail addresses: rbouadjenek@gmail.com (M.R. Bouadjenek), hakim.hacid@zu.ac.ae (H. Hacid), mokrane.bouzeghoub@prism.uvsq.fr (M. Bouzeghoub).

http://dx.doi.org/10.1016/j.is.2015.07.008
0306-4379/© 2015 Elsevier Ltd. All rights reserved.

[1] https://myspace.com/
[2] https://www.facebook.com
[3] https://www.linkedin.com/
[4] https://delicious.com/
[5] http://www.citeulike.org/
[6] https://www.flickr.com/
[7] https://twitter.com

certainly the most adopted technologies in this new era. These platforms are commonly used as ameans to communicate with other users, exchange messages, share resources (photos and videos), comment news, create and update profiles, interact and play online games, etc. In addition to dedicated social platforms, traditional content providers sites like newspapers, tend to be more social since they provide to users means for sharing, commenting, constructing, and linking documents together [1,2], e.g., through the buttons *"share on social networks"*. This has been also facilitated by initiatives like OpenID[9] and OpenSocial.[10]

These collaborative tasks that make users more active in generating content are among the most important factors for the increasingly growing quantity of available data. In such a context, a crucial problem is to enable users to find relevant information with respect to their interests and needs. This task is commonly referred to as Information Retrieval (IR). IR is performed every day in an obvious way over the Web [3], typically using a search engine. However, classic models of IR do not consider the social dimension of the Web. They model web pages[11] as a mixture of static homogeneous terms generated by the same creators, i.e., the authors of the web pages. Then, the ranking algorithms are often based on (i) a query and document text similarity (e.g., the cosine measure and the Okapi BM25 [4]), and (ii) the existing hypertext links that connect these web pages (e.g., PageRank [5], and HITS [6]).

Therefore, classic models of IR and even the IR paradigm have to be adapted to the socialization of the Web in order to fully leverage the social context that surrounds web pages and users. Indeed, exploiting social information has a number of advantages (for IR in particular): first, feedback information in social networks is provided directly by the user, so users' interests accurate information can be harvested as people actively express their opinions on social platforms. Second, a huge amount of social information is published and available with the agreement of the publishers. Exploiting this information should not violate user privacy, in particular social tagging information, which does not contain sensitive information about users. Finally, social resources are often accessible, as most of social networks provide APIs to access their data (even if often, a monetized contract must be established before any large scale use).

There is currently a number of research works undertaken to improve the IR process with information coming from social networks. This is commonly known as "Social Information Retrieval" (SIR). This paper intends to provide a clear understanding of the various efforts performed in the domain of SIR, and in this perspective, we propose

1. An objective review of some of the most representative research contributions and existing tools in this domain to understand the principles of SIR as they are currently formulated.
2. A taxonomy to categorize these contributions in order to structure this wide domain.
3. An analysis of some of these contributions and tools with respect to several criteria considered as crucial to design an effective SIR approach or to appreciate one.

This study is useful for researchers, engineers, and practitioners in this domain to help in understanding trends, challenges and expectations of a SIR approach. The rest of this paper is organized as follows: in Section 2, we discuss the main concepts used throughout this paper. In Section 3, we introduce our taxonomy of SIR approaches and tools, while describing each of these categories in Sections 4–6. In Section 7, we elaborate on an analysis of some of these contributions and tools based on several dimensions. Finally, we give some future directions in Section 8 and we conclude in Section 9.

2. Background

Information Retrieval is the process of recovering stored information from large datasets to satisfy users' information needs. Salton [7] and Baeza et al. [3] defined IR as follows:

Definition 1 (*Information Retrieval*). Information Retrieval (IR) is the science that deals with the representation, storage, organization of, and access to information items in order to satisfy the user requirements concerning this information.

An IR system is evaluated in its accuracy and ability to retrieve high quality information/documents, which maximize users' satisfaction, i.e. the more the answers correspond to the users' expectations, the better the system is. Information Retrieval is a very well established domain. Several articles and books are devoted to IR, e.g., [3,7]. Since our focus is on the social part of IR, we do not go into more detail in the description of the IR domain.

This section introduces and defines the basic concepts used throughout this paper. In Section 2.1 we define what is a social network, and discuss the main models of social relationships.[12] Then, in Section 2.2, we introduce the notion of Social Information Retrieval that bridges social networks and information retrieval.

2.1. Social networks and social network analysis

Nowadays, social networks are at the heart of the Web 2.0. A social network is defined as follows:

Definition 2 (*Online social network*). An online social network is the social structure, which emerges from human interactions through a networked application.

[8] https://www.yammer.com/

[9] http://www.openid.net/

[10] http://www.opensocial.org/

[11] In this paper, we also refer to web pages as documents.

[12] Note that we do not intend to provide a complete review of the SNA domain, but rather to provide the main underlying principles, which could be leveraged in IR systems and which are helpful to understand the analysis provided in this paper.

Example 3 (*Online social network*). *Facebook* is certainly the most popular social network that handles relationships between individuals. There are many social network sites that manage in addition to users, objects. These include documents on *CiteULike*, images on *Flickr*, videos on *YouTube*, etc.

The structure of a social network can be constructed in two ways: (i) either explicitly declared by the user, e.g., friendship links in *Facebook*, or (ii) implicitly inferred from the behavior and the common interests of users, e.g., Social Network of Web services [8]. To understand the underlying social structures and phenomena, a set of techniques and methods exist, which are known as Social Networks Analysis (SNA) techniques [9]. SNA introduces methods and metrics (e.g., centrality and influence) for analyzing a Social Network, e.g., measuring the role of individuals and groups of individuals in a social network.

Each social network might be characterized by the relationships that link its users, e.g., friend relationships, follower–following relationships, and publisher–subscriber relationships. Hence, we distinguish mainly three models of social relationships that we describe hereafter.[13]

2.1.1. Symmetric relationships

Many social networks manage symmetric relationships that translate the same consideration of relations between entities, i.e. users, participating in the relation. Social networks that include these relationships allow for example users to maintain a list of friends and thus create friendship relations. The *Friendship* relation is of the form Alice considers Bob as a friend and she explicitly adds Bob to her list of friends. This relation is instantiated once Bob accepts the request. Thus, the friendship strength of a link between two users can reflect for example, the degree of trustiness, and the degree of mutual interest. This model is more dedicated to create and maintain personal relationships with trusted persons and with whom one is expected to share personal elements and contents.

Facebook is a good example of a social network that includes these relationships. Let us consider the users *Alice* and *Bob* on *Facebook*, if *Alice* has a relationship with *Bob,* then *Bob* has the same relationship with *Alice*.

2.1.2. Asymmetric relationships

Similarly, many social networks manage social relationships that can link two users with two different perspectives depending on the users. These relationships illustrate the concept of followers–following or publisher–subscriber and are generally at the heart of microblogging platforms, e.g., *Twitter* and *Yammer*. These social networks allow a user to create and maintain a list of following people by permitting him to subscribe to their information stream. This model of social networks is more dedicated to the dissemination of information than to mutual share of information. For example, if we consider a relationship between a user *Alice* and a user *Bob* on *Twitter*. *Alice* may subscribe to the content published by *Bob*, while *Bob* does not necessarily subscribe to the content of *Alice* as he considers her from a different perspective, e.g., his perspective.

2.1.3. k-partite relationships

The two previous models of social relationships involve generally only one type of nodes, i.e. users. In this last model, social relationships imply k types of nodes, e.g., users, resources, and tags. Social bookmarking websites are representatives of such models.

Social bookmarking websites are based on the techniques of *social tagging or collaborative tagging*. The principle behind social bookmarking platforms is to provide the user with means to annotate resources on the Web, e.g., URIs in *delicious*, videos in *youtube*, images in *flickr*, or academic papers in *CiteULike*. Thus, there are three different kinds of entities: users, tags, and resources, involved in 3-partite relationships. The annotations (also called tags) can be shared with other users. This unstructured (or free structured) approach to classification with users assigning their own labels is often referred to as a *folksonomy* [10,11].

It is important to notice that most of the existing Social Networking platforms are not restricted to manage only one kind of social graph but they may manage several social relationships. For example, *delicious* which is a tagging based platform, manages (1) a social graph of ternary relationship between users, tags, and documents and (2) a social graph of publisher–subscriber to provide to users, with means to view all the bookmarks saved by interesting people, such as friends, co-workers, and favorite bloggers. Finally, besides these social relationships, a social content is generated and exchanged between users. This content is the second important component of online social networks, which is leveraged in several domains, such as SIR that we are discussing in this paper.

In the next section, we discuss the concept of social information retrieval, which links IR and social networks.

2.2. Social information retrieval

With the intrinsic social property of the Web, a large range of applications and services make the user becoming more interactive with Web resources, and a lot of information that concerns both users and resources is constantly generated. This information can be very useful in information retrieval tasks for both user and resources modeling. However, classical models of IR are blind to this social context that surrounds both users and resources.

Therefore, the domains of IR and SNA have been bridged resulting in *Social Information Retrieval (SIR)* models [12]. Very often, SIR models extend conventional IR models in order to incorporate social information. However, as we shall discuss in the next sections, new IR paradigms and concepts have emerged in order to provide a new way of operating information retrieval.

The meaning of the concept *Social Information Retrieval* can be very broad, but we propose the following definition:

Definition 4 (*Social Information Retrieval*). *Social Information Retrieval* is the process of leveraging social information (both social relationships and the social content), to

[13] We only describe the main social relationships models that we believe are the most adopted in social networks.

perform an IR task with the objective of better satisfying the users' information needs.

SIR aims to provide relevant content and information to users in the domains of information retrieval, research, and recommendation; covering topics such as social tagging, collaborative querying, social network analysis, subjective relevance judgments, Q&A systems, and collaborative filtering [12].

Several existing platforms investigate this track in order to improve the search paradigm as illustrated in Fig. 1. These include *Social Bing, Google+, Aardvark, Yahoo! Answers*, etc. The research in this domain has emerged and became very present in the daily (virtual) life of users. Investigating the IR domain from this perspective seems to be very promising to improve the representation, the storage, the organization, and the access to information. The large number of work in this domain is certainly a good indicator of the interest its taking.

In order to provide a better overview and understanding of all the work done in the SIR domain, we propose a taxonomy to categorize and classify the proposed methods. Basically, this taxonomy considers these different contributions from the way they exploit social information. Each category in the taxonomy is then discussed and examples will be given. This taxonomy is discussed in the next section.

3. A taxonomy for social information retrieval

There are many contributions in the domain of SIR. Each of these contributions considers a particular social network type and uses social information to perform an IR task differently. For example, on one hand, tags in folksonomies have been found useful for Web search, personalized search and contextualized enterprise search. Indeed, Gupta et al. [13] provide a survey summarizing different properties of tags along with their usefulness for IR like tag semantics, recommendations using tags, tag profiling, etc. Also, Heymann et al. [14] analyze folksonomies, and conclude that social bookmarking systems can provide search results not currently provided by a conventional search engine (approximately 25% of URLs posted by users are new, unindexed pages). On the other hand, micro-blogging systems like *Twitter* have been also found useful to users to share or submit specific questions to be answered by friends, families, colleagues, or even an unknown person (using a hashtag for a specific topic).

From these two examples, we understand that different models of social networks have been used differently for IR tasks. Thus, through the bibliographical study that we have performed, we distinguished several SIR approaches that can be categorized according to the way social information is used. Therefore, we propose a taxonomy for grouping

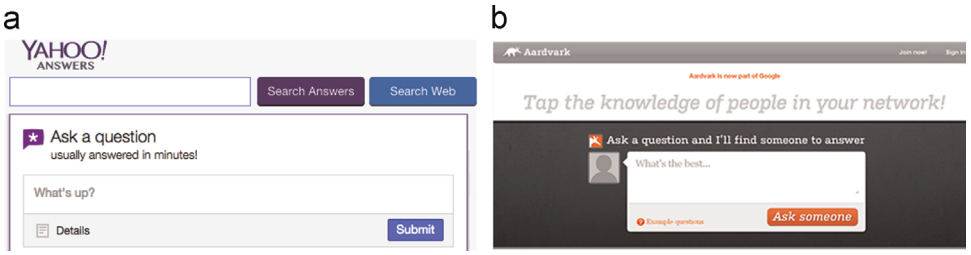

Fig. 1. Examples of two social search engine that allow users to both submit questions to be answered and answer questions asked by other users. (a) Yahoo! Answers. (b) Aardvark.

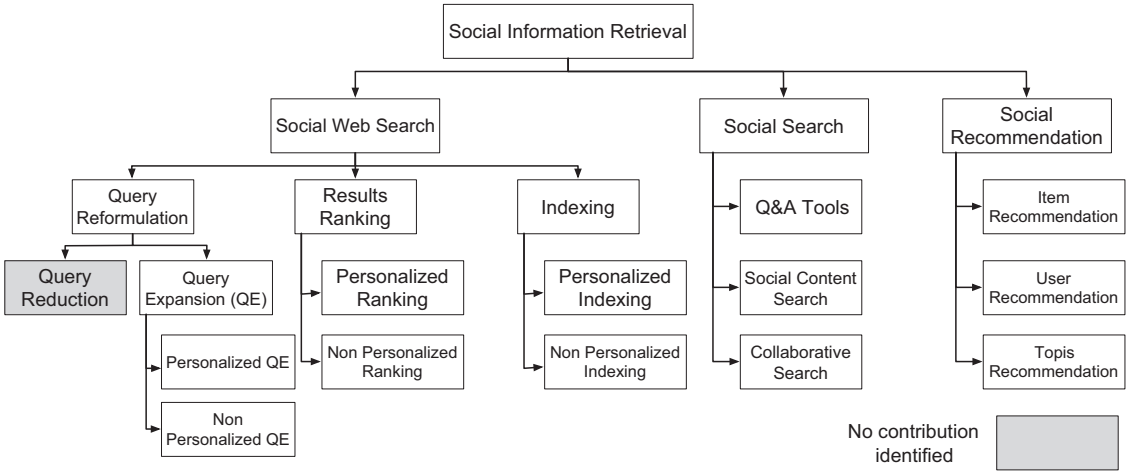

Fig. 2. A taxonomy for Social Information Retrieval Models.

these different initiatives and building a common understanding of this domain. Fig. 2 summarizes this taxonomy of SIR models, which is mainly composed of three categories:

1. *Social Web Search*, in which social information is used in order to improve the classic IR process, e.g., documents re-ranking, query reformulation, and user profiling. We discuss this category of SIR approaches in Section 4.
2. *Social Search*, in which it is a matter of finding information only with the assistance of social resources, such as by asking friends, or unknown persons on-line for assistance [15]. This third category is discussed in Section 5.
3. *Social Recommendation*, in which the user's Social Network is used to provide better recommendation, e.g., using a social trust network [16]. This category is discussed in Section 6.

Several contributions closely related to the domain of Social Information Retrieval are discussed in this paper. The objective is not to discuss the whole set of contributions but to point the closest ones as illustrative contributions. In the following, we discuss and detail these different categories, while giving some illustrative examples.

4. Social web search

We consider this category to include techniques that improve the conventional IR process using social information. In existing IR systems, queries are usually interpreted and processed using document indexes and/or ontologies, which are hidden for users. The resulting documents are not necessarily relevant from an end-user perspective, in spite of the ranking performed by the search engine.

To improve the classic IR process and reduce the amount of irrelevant documents, there are mainly three possible improvement tracks: (i) query reformulation, i.e. which includes expansion or reduction of the query, (ii) post-filtering or re-ranking of the retrieved documents (based on the user profile or context), and (iii) improvement of the IR model, i.e. the way documents and queries are represented and matched to quantify their similarities. Here, we consider the use of social information in these three tracks.

4.1. Query reformulation

In IR systems, users express generally their needs through a set of keywords that summarize their information needs. Thus, different users are expected to use different keywords to express the same need (e.g., synonyms), and vice versa (i.e., the same keywords can be used by different users to express different information needs). Query reformulation can then bring a solution to this problem. It is defined as follows:

Definition 5 (*Query reformulation*). *Query reformulation* is the process which consists of transforming an initial query Q to another query Q'. This transformation may be either a *reduction* or an *expansion.Query Reduction* (QR) [17] reduces the query such that superfluous information is removed, while *Query Expansion* (QE) [18] enhance the query with additional information likely to occur in relevant documents.

To the best of our knowledge, there are no contributions in *Query Reduction* using social information, but all the existing work focuses on *Query Expansion*. In this latter, we distinguish two types of approaches: (i) non-personalized QE and (ii) personalized QE.

4.1.1. Non-personalized social query expansion

In traditional query expansion methods, the database used for the expansion is often constructed according to the comparison between terms' distributions in the retrieved documents and in the whole document collection/database. Here, this database of terms is enhanced using social information without any personalization. Therefore, the underlying idea is to leverage the interactions of users with the system to implicitly and collaboratively build a database of terms. This database is expected to feed the expansion process. This yields to a user-based vocabulary source for query expansion.

Many methods have been proposed in this area. Lioma et al. [19] provide Social QE by considering query expansion as a logical inference and by considering the addition of tags as an extra deduction to this process. In the same spirit, Jin et al. [20] propose a method in which the used expansion terms are selected from a large amount of social tags in folksonomy. A tag co-occurrence method for similar terms selection is used to choose good expansion terms from the candidate tags directly according to their potential impact on the retrieval effectiveness. The work in [21] proposes a unified framework to address complex queries on multi-modal "social" collections. The approach they proposed includes a query expansion strategy that incorporates both textual and social elements. Finally, Lin et al. [22] propose this to enrich the source of terms expansion initially composed of relevant feedback data with social annotations. In particular, they propose a learning term ranking approach based on this source in order to enhance and boost the IR performances.

4.1.2. Personalized social query expansion

In query expansion, providing merely a uniform expansion to all users is often not really suitable nor efficient as relevance of documents is relative for each user [23]. Thus, a simple and uniform query expansion is not enough to provide satisfactory search results for each user. Personalized social query expansion refers to the process of expanding the same query differently for each user using social information.

Example 6. Let us consider the query Q= *"Computer science"*, the user Bob may have the expanded query Q' = "*Computer science technology programming java*", whereas the expanded query q' = "*Computer science technology Internet information*" may be more suitable for Alice, depending on their topics of interest.

Several efforts have been made to tackle this problem of personalized query expansion, in particular in the context of folksonomies. Hence, authors in [24,25] consider SIR from both the query expansion and results ranking. Briefly, this strategy consists of adding to the query q, k possible expansion tags with the largest similarity to the

original tags in order to enrich its results. For each query, the query initiator u, ranks results using BM25 and tag similarity scores. Bertier et al. [26] propose *TagRank* algorithm, an adaptation of the *PageRank* algorithm, which automatically determines which tags best expand a list of tags in a given query. This is achieved by creating and maintaining a *TagMap* matrix, a central abstraction that captures the personalized relationships between tags, which is constructed by dynamically computing the estimation of a distance between taggers, based on cosine similarity between tags and items.

Biancalana et al. [27] proposed *Nereau*, a query expansion strategy where the co-occurrence matrix of terms in documents is enhanced with meta-data retrieved from social bookmarking services. The system can record and interpret users' behavior, in order to provide personalized search results, according to their interests in such a way that allows the selection of terms that are candidates of the expansion based on original terms inserted by the user. The input queries are analyzed according to collected data, then the system makes different query expansions, each one to a different semantic field, before carrying out the search. The final result is a page in which results are grouped in different blocks, each of them categorized through keywords to facilitate for the user the choice of the result that is most coherent with his interests. This method has been improved in [28] using three-dimensional matrices, where the added dimension is represented by semantic classes (i.e., categories comprising all the terms that share a semantic property) related to the folksonomy extracted from social bookmarking services.

Finally, Bouadjenek et al. [29] propose an approach that considers ranking terms for expansion purposes. The ranking process takes into account: (i) the semantic similarity between tags composing a query, and (ii) a social proximity between the query and the user for a personalized expansion.

4.2. Results ranking

In IR, ranking results consist in the definition of a function that allows quantifying the similarities among documents and queries. We distinguish two categories for social results ranking that differ in the way they use social information. The first category uses social information by adding a social relevance to the ranking process, while the second uses it to personalize search results.

4.2.1. Non-personalized ranking (ranking using social relevance)

Social relevance refers to information socially created that characterizes a document from a point of view of interest, i.e. its general interest and its popularity. Two formal models for folksonomies and ranking algorithm called *folkRank* [30] and *SocialPageRank* [31] have been proposed. Both are an extension of the well-known *PageRank* algorithm adapted for the generation of rankings of entities within folksonomies. *SocialPageRank* intends to compute the importance of documents according to the mutual enhancement relation among popular resources, up-to-date users and hot social annotations. In the same spirit, relying on social bookmarking systems, Takahashi et al. [32,33] propose *S-BIT* and *FS-BIT*, which are extensions of the well-known HITS approach [6]. Yanbe et al. [34] proposed *SBRank*, which indicates how many users bookmarked a page, and use the estimation of *SBRank* as an indicator of Web search.

All these algorithms are in the context of folksonomies, and a number of them are reviewed and evaluated in [35]. The work in [36] proposes a method to use microblogging data stream to compute novel and effective features for ranking fresh URLs, i.e., "uncrawled" documents likely to be relevant to queries where the user expects documents which are both topically relevant as well as fresh. The proposed method consists of a machine-learning based approach, that predicts effective rankings for queries–url pair. Recently He et al. [37] propose a new method to predict popularity of items (i.e., webpages) based on users' comments, and to incorporate this popularity into a ranking function.

4.2.2. Personalized ranking

Several approaches have been proposed to personalize ranking of search results using social information [24,38–43]. Almost all these approaches are in the context of folksonomies and follow the common idea that the ranking score of a document d retrieved when a user u submits a query Q is driven by (i) a term matching, which calculates the similarity between Q and the textual content of d to generate a user unrelated ranking score; and (ii) an interest matching, which calculates the similarity between u and d to generate a user related ranking score. Then a merge operation is performed to generate a final ranking score based on the two previous ranking scores. A number of these algorithms are reviewed and evaluated in [44], while considering different social contexts.

4.3. Indexing and modeling using social information

In the social Web, a social context is often associated to web pages, which can tell a lot about their content. As an example, this social context includes annotations, comments, like mentions, etc. Consequently, as pointed in [45], *social contextual summarization* is required to strengthen the textual content of web pages. Several research work [46–49] reported that adding a tag to the content of a document enhances the search quality as they are good summaries for documents [49] (e.g., document expansion [50,14]). In particular, social information can be useful for documents that contain few terms where a simple indexing strategy is not expected to provide good retrieval performances (e.g., the *Google homepage*[14]).

Throughout our analysis of the state-of-the-art, we noticed that social information has been mainly used in two ways for modeling and enhancing document representation: (i) either by adding social meta-data to the content of documents, e.g., document expansion, or (ii) by personalizing the representation of documents, following the

[14] http://www.google.com/

There are only a very few terms on the page itself but thousands of annotations available on *delicious* are associated to it. Eventually, social annotations of the *Google* homepage are more useful for indexing.

intuition that *each user has his own vision of a given document*.

4.3.1. Document expansion (non-personalized indexing).

Some works investigate the use of social metadata for enriching the content of documents. In [47,51,48,52], authors index a document with both its textual content and its associated tags modeled as in the Vector Space Model (VSM) [53]. However, each method uses a different algorithm for weighting social metadata, e.g., tf-idf [47,52], term quality [47,51], etc. Also, Zhang et al. [46] propose a framework to enhance document representation using social annotations. The framework consists in representing a Web document in a dual-vector representation: (i) enhanced textual content vector and (ii) enhanced social content vector. Each component being calculated from the other.

4.3.2. Personalized indexing and modeling of documents

Given a document, each user has his own understanding of its content. Therefore, each user employs a different vocabulary and words to describe, comment, and annotate this document. For example, if we look at the homepage of Youtube, a given user can tag it using "video", "Web" and "music" while another can tag it using "news", "movie", and "media".

Following this observation, Bouadjenek et al. [45] proposed a framework for modeling personalized representations of documents based on social annotations. This framework is based on matrix factorization, where the idea is to provide a personal representation of a given document for a particular user, which is then used for query processing. Also, Amer-Yahia et al. [54] investigate efficient top-k processing in collaborative tagging sites. The idea is that the score of an answer is computed as its popularity among members of a seeker's network. Basically, the solution is to create personalized indexes based on clustering strategies, which achieve different compromises between storage space and processing time.

Finally, in [55] the authors proposed a dual personalized ranking function, which adopts two profiles: an extended user profile and a personalized document profile. Briefly, for each document, the method computes for each individual user a personalized document profile to better summarize his/her perception about this document. The solution proposed is to estimate this profile based on the perception similarities between users.

5. Social search

Social platforms like *Twitter* and *Facebook* allow users to share and publish information with their friends and often with general public. In addition to this, users use them to answer very precise and highly contextualized queries, or queries for which the relevant content has not been authored yet, e.g., asking about a conference event using its hashtag on *Twitter*. We refer to such a process of finding information as *Social search*, and we define it as follows.

Definition 7 (*Social search*). Social search is the process of finding information only with the assistance of social entities, by considering the interactions or contributions of users.

Thus, social search is associated to platforms that are defined as search engines specifically dedicated to social data management such as Facebook. The main ingredient to perform a social search is the user interactions, including (i) social content (e.g., comments and tweets) and (ii) social relations (e.g., finding a person with a certain expertise). Hence, social search systems index either social content and offer a means for users to search that content [56], or social relations and allow the user finding persons, who are likely to respond to specific needs [57]. We divide social search into three main categories: (1) question/answering tools, (2) content search, and (3) collaborative search. These categories are detailed in the following.

5.1. Social question/answering (Q&A)

Despite the development of techniques and methods for Web search assistance such as navigational queries [58] and query auto-completion [59,60], for helping users to express their needs, many queries still remain unanswered. Dror et al. [61] argue that this is mainly due to two reasons: (i) the intent behind the query not being well expressed/captured and (ii) the absence of relevant content.

To tackle these issues, *Question/Answering Systems (Q&A)* have emerged to connect people for helping each other answering questions. Examples of such systems include *Yahoo Answer!*,[15] *WikiAnswers*,[16] *Aardvark*,[17] *Chacha*,[18] and *Ask.com*.[19] Basically, Q&A systems provide a means for answering several types of questions such as recommendation, e.g., building a new play list, any ideas for good running songs?, opinion seeking, e.g., *I am wandering if I should buy the Kitchen-Aid ice cream maker?*, factual knowledge, e.g., *Does anyone know a way to put Excel charts into LaTeX?* problem solving, e.g., *How do I solve this Poisson distribution problem?* Morris et al. [15] conducted a survey in which they study the type of asked questions, the frequency of these questions, and the motivations for users asking their social network rather than using a traditional search engine.

The main problem facing such systems is in the response time as well as the quality of the answers. For example, Zhang et al. [62] report that on the Java Developer Forum, the average waiting time of a high expertise user to get a reply for a question is about 9 h, compared with 40 min for a low expertise user. As for Microsoft's Live Q&A site, Hsieh and Count [63] state that 80% of queries receive an answer, with an average response time of 2 h and 52 min. As for Aadvark [57], 87.7% of questions received at least 1 answer, and 57.2% received their first answer in less than 10 min. Dror et al. [61], study a dataset of 4 month period of Yahoo

[15] http://answers.yahoo.com/
[16] http://wiki.answers.com/
[17] Aardvark has been acquired by Google on February 11, 2010. In September 2011, Google announced it would discontinue a number of its products, including Aardvark.
[18] http://www.chacha.com/
[19] http://www.ask.com/

Answer! interactions, and state that the average time for answering queries is 10 min, while almost all answers were given within 60 min of the question's creation time. However, in order to properly rank answers in Q&A systems, an interesting method has been proposed by Dalip et al. [64]. The authors proposed a learning approach for ranking using specific features of Q&A domain, which was able to significantly outperform state of the art baselines.

5.2. Social content search

Social platforms allow users to provide, publish and spread information, e.g., commenting or tweeting about an event. In such a context, a huge quantity of information is created in social media, which represents a valuable source of relevant information. Hence, many users use social media to gather recent information about a particular event by searching collection of posts, messages and statuses. Therefore, social content search systems come as a mean to index content explicitly created by users on social media and provide a real-time search support [65].

There are several social content search engines, which index real-time content spreading systems. This includes *TwitterSearch*,[20] *Social Bing*,[21] *collecta* [66], *OneRiot* [67],[22] etc. Social content search systems deal with a different kind of content than classic search engines. Indeed, posts and statutes published on social media are often short, frequent, and do not change after being published, while web pages are rich, generated more slowly, and evolve after creation [56]. Dealing with such content is challenging, because it requires real-time and recency sensitive queries processing. Sensitive query refers to a query where the user expects documents, which are both topically relevant as well as fresh [36,68]. A study has been performed by Teevan et al. [56] that gives an overview of "What is the motivation behind a user to use a social content search system rather than a classic search engine?". This study reveals that *social content search systems* are interrogated with queries, which are shorter, more popular, and less likely to evolve as part of a session than Web queries. The main goal is to find temporally relevant information (e.g., breaking news, real-time content, and popular trends) and information related to people (e.g., content directed at the searcher, information about people of interest, and general sentiment and opinion).

Dong et al. [36] propose a method to use microblogging data stream to compute novel and effective features for ranking fresh URLs, i.e., "uncrawled" documents likely to be relevant to recency sensitive queries. The proposed method consists of a machine-learning based approach, that predicts effective rankings for queries-url pair. Also, Efron et al. [68] propose to estimate the temporal density of relevant documents, starting with an initial set of results from a baseline retrieval model. A reranking of results is then used. Their contributions lie in a method to characterize this temporal density function using kernel density estimation, with and without human relevance judgments, and an approach to integrating this information into a standard retrieval model. Finally, Vosecky et al. [69] propose a framework for collaborative personalized Twitter Search, which exploits the user's social connections in order to obtain a comprehensive account of her preferences. This framework includes a novel user model structure to manage the topical diversity in Twitter and to enable query disambiguation.

5.3. Social collaborative search

One of the weaknesses of search engines available today (e.g., Google, Yahoo!, Bing) is the fact that they are designed for a single user who searches alone. Thus, users cannot benefit from the experience of each other for a given search task. Morris [70,71] conducts a survey on 204 knowledge workers in a large technology company in which she revealed that 97% of respondents reported engaging in one of collaborative search task described in the survey. For example, 87.7% of respondents reported having watched over someone's shoulder for query suggestion, and 86.3% of respondents reported having e-mailed someone to share the results of a Web search.

In such a context, Morris [72] developed *SearchTogether*, a collaborative search interface, where several users who share an information need collaborate and work together with others to fulfill that need. The authors discuss the way *SearchTogether* facilitates collaboration by satisfying criteria like awareness, division of labor, and persistence. Similarly, Filho et al. [73] proposed *Kolline*, a search interface that aims at facilitating information seeking for inexperienced users by allowing more experienced users to collaborate together.

Paul and Morris [74] investigate *sensemaking* for collaborative Web search, which is defined as the act of understanding information. The study revealed several themes regarding the sensemaking challenges of collaborative Web search, e.g., Awareness, Timeliness and sensemaking handoff. Based on their finding, they proposed *CoSense*, a system that supports *sensemaking* for collaborative Web search tasks that provides enhanced group awareness by including a time-line view of all queries executed during the search process. Even though these features help to enhance participants' communication and *sensemaking* during their search activities, users still have to sort among different documents and analyze them one by one to find relevant information.

Finally, *SearchTeam*[23] is a concrete example of a collaborative search system currently available. Collaborative search in *SearchTeam* is conducted within a SearchSpace. Collaborators search the Web together, save and edit their results into the SearchSpace, and pick up next time where they left off. Results are organized into folders within the SearchSpace. Within these folders, collaborators can comment on search results, "like" them, post their thoughts, add search results, post-links, or upload documents. Users can log in using their Facebook, Twitter, Google, LinkedIn, and

[20] https://twitter.com/search-home
[21] http://www.bing.com/social
[22] OneRiot has been acquired by Walmart in September 2011.
[23] http://searchteam.com/

Yahoo accounts, and can pick collaborators from their contacts and social networks.

In summary, in this section we discussed the Social Search category of SIR. We showed that the search paradigm has evolved to provide users with tools and methods for asking more sophisticated and contextualized queries. Social content search systems allow to mine and find relevant information within posts and users comments, while Q&A systems allow users to answer very contextualized queries, whereas social collaborative search systems come as a mean to help users share their experiences and findings. In the next section, we present Social Recommendation in which recommendation is done based on the user's social network.

6. Social recommendation

The second category of SIR models considers filtering and recommendation domains (e.g., content-based filtering, collaborative filtering, recommender systems). Basically, recommendation aims at predicting the interest that users would give to an item/entity they had not yet considered explicitly. There are two main methods of recommendation: (i) an approach based on recommending items that are similar to those in which the user has shown interest in the past, which is known as *"content-based"* approach (CB), and (ii) an approach that intends to recommend items to the user based on other individuals who are found to have similar preferences or tastes, which is known as *"collaborative filtering"* approach (CF). We define social recommendation as follows:

Definition 8 (*Social Recommendation*). Social recommendation is a set of techniques that attempt to suggest: (i) items (e.g., movies, music, books, news, web pages), (ii) social entities (e.g., people, events, groups), or (iii) topics of interest (e.g., sport, culture, and cooking) that are likely to be of interest to the user through the use of social information.

On one hand, in recent years, many personalized recommendation features based on the user's social network have been developed and integrated to popular web sites. This is mainly done by prompting the user to connect his social networks' accounts to their services; the ultimate goal being collecting as much data as possible that characterize the user. For example, videos recommendation in *YouTube based on the Google+profile*, or movies recommendation on *IMDb* through the registration using a *Facebook* account. This functionality has been reported to lead to an improvement of the recommendation services.

On the other hand, many social platforms understood the power of learning over their data for building recommender services. For example, targeted advertising in some social platforms like *Facebook*, group recommendation again on *Facebook*, follower recommendation on *Twitter*, or topics and web pages recommendation on *delicious*. Moreover, there are other social web services, whose recommendation is at the heart of their business. For example, social news aggregation services, like *Digg*, which presents stories expected to be most interesting to a user, based on preferences of similar users. It is clear that all these services have been improved by exploiting their social information.

In the following, we categorize social recommender systems according to the type of output they intend to recommend: (i) items recommendation, (ii) users recommendation, and (iii) topics recommendation.

6.1. Items recommendation

Items recommendation has probably attracted the most attention in recommender systems. Classical items recommendation methods are based on the assumption that users are independent entities and identically distributed. Thus, they do not suppose any additional structure, including the social network that surrounds users. This does not reflect the real behaviors of users, since they normally ask friends for recommendations before acting, e.g., buying a product. Many researchers have then started exploring social relations to improve recommender systems (including implicit social information, which can be employed to improve traditional recommendation methods [75]), essentially to tackle the cold-start problem [76,77,16,78]. However, as pointed in [79], only a small subset of user interactions and activities are actually useful for social recommendation.

In collaborative filtering based approaches, Liu and Lee [80] proposed very simple heuristics to increase recommendation effectiveness by combining social networks information. Guy et al. [81] proposed a ranking function of items based on social relationships. This ranking function has been further improved in [82] to include social content such as related terms to the user.

Another common approach to CF attempts to factorize an (incomplete) matrix R of dimension $N \times M$ containing observed ratings $r_{i,j}$ (which represent the rate of the user i to the item j) into a product $R \approx U^T V$ of latent feature matrices U and V. In this context, following the intuition that a person's social network will affect his behaviors on the Web, Ma et al. [16] propose to factorize both the users' social network and rating records matrices. The main idea is to fuse the user-item matrix with the users' social trust networks by sharing a common latent low-dimensional user feature matrix. This approach has been improved in [83] by taking into account only trusted friends for recommendation while sharing the user latent dimensional matrix. Almost a similar approach has been proposed in [84,85] who include in the factorization process, trust propagation and trust propagation with inferred circles of friends in social networks respectively. In this same context, other approaches have proposed to consider *social regularization terms* while factorizing the rating matrix. The idea is to handle friends with dissimilar tastes differently in order to represent the taste diversity of each user's friends [86–88]. A number of these methods are reviewed, analyzed and compared in [89].

6.2. Users recommendation

As mentioned above, social network platforms have adopted the strategy of suggesting friends (or group of friends) to increase the connectivity among their users. Hence, content-based approaches were proposed in [90] to match the content of user profiles and determine user

Table 1
Summary of the considered dimension. (×) means the dimension (i.e. functionality) is provided, (—) means the dimension is not provided. These marks do not imply any "positive" or "negative" information about the tools except for the presence or the absence of the considered dimension. The mark (?) means we can say nothing about the dimension because of lack of information.

				Social Information Retrieval																			
				Social Web Search							Social Search						Social Recommendation						
				QE		Ranking			Indexing		Q&A		Content		Collaborative		Item		User		Topic		
				[22]	[29]	[39]	[31]	[43]	[46]	[47]	[57]	ChaCha	[66]	TwitterSearch	[72]	[74]	[16]	[87]	[90]	[103]	[99]	[92]	
1	Social	Symmetric		–	–	–	–	–	–	–	×	?	×	–	–	–	–	×	×	×	–	–	
	Networks	Asymmetric[a]	CS	–	–	–	–	–	×	–	×	?	×	×	–	–	–	–	–	–	–	–	
	Relationships		TR	–	–	–	–	–	–	–	×	?	–	–	–	–	×	×	–	–	–	–	
		Ternary		×	×	×	×	×	×	×	–	?	–	–	–	–	–	–	–	–	×	×	
		Their own SN		–	–	–	–	–	–	–	×	?	–	×	×	×	–	–	–	–	–	–	
2	Social	Content		×	×	×	×	×	×	×	×	?	×	×	–	–	–	–	×	×	×	×	
	Network Data	Structure		–	×	×	×	–	–	×	–	?	–	?	×	×	×	×	×	×	×	×	
		Metadata		–	–	–	–	–	–	–	×	×	×	×	–	–	–	–	–	–	–	–	
3	Datasource[c]	Social Net		×	×	×	×	×	×	×	×	?	×	×	–	–	×	×	×	×	×	×	
		Web		–	–	–	–	×	×	×	–	?	–	×	–	–	–	–	–	–	–	–	
		Provided by users		–	–	–	–	–	–	–	×	×	–	–	×	×	–	–	×	–	–	–	
4	Personalization	Profile		–	×	×	–	×	–	–	×	?	–	?	–	–	–	–	–	–	–	–	
		Content-based		–	–	–	–	–	–	–	–	?	–	?	×	×	–	–	×	–	–	×	
		Colla-filtering		–	–	–	–	–	–	–	–	?	–	?	–	–	×	×	×	×	–	×	
5	Complexity & adaptability	Scalability		×	×	×	–	×	×	–	×	×	×	×	×	×	×	×	×	×	×	–	
		Dynamicity		–	×	×	–	×	–	–	×	×	×	×	×	×	–	–	×	–	–	–	
		Data sparsity		×	×	–	×	×	×	×	×	×	×	×	–	–	×	×	–	×	–	×	
		Cold start		×	×	×	×	×	×	×	×	×	×	×	–	–	×	×	–	–	–	–	
6	Evaluation	Formal		×	×	×	×	×	×	×	–	?	×	?	–	–	×	–	–	×	×	×	
		End-user		–	–	×	–	–	–	–	×	?	–	?	×	×	–	×	×	–	–	–	
7	Socialization			–	×	–	–	–	–	–	×	×	×	×	×	×	×	×	×	×	–	×	
8	Privacy			×	×	×	×	×	×	×	–	–	–	–	–	–	–	–	–	–	–	–	
9	Industrialization			–	–	×	–	–	–	–	×	×	×	×	×	×	–	–	–	–	–	–	

[a] CS=Content Subscription; TR=Trust relations.

similarities for recommendation. Groh et al. [91] generated user neighborhood information from known social network structures and demonstrated that collaborative filtering based on such neighborhood outperforms classic collaborative filtering methods. Symeonidid et al. [92] proposed a ternary semantic analysis unified framework to perform users recommendation. Guy et al. [90] describe an user interface for providing users with recommendations of people to invite into their social network. The approach is based on aggregated information collected from various sources. Hannon et al. [93] utilize content and collaborative-based approaches to evaluate a range of different user profiling and recommendation strategies. Many of these strategies have been implemented in [94]. Finally, in the context of tagging systems, Wang et al. [95] propose to connect users with similar tastes by measuring their similarities based on the tags they share in an inferred network of tags.

6.3. Topics and tags recommendation

Recently, topic and tag recommendation has attracted significant attention to produce high quality hot-lists to users. Recommendation of tags also allows users to choose the right tags as tagging is not constrained by a controlled vocabulary and annotation guidelines. Hence, the work in [96] provides a comprehensive evaluation and comparison of several state-of-the-art tag recommendation algorithms in three different real world datasets. A content-based collaborative filtering technique has been proposed in [97] to automate tags assignments to blogs. Hotho et al. [30] propose to project the three-dimensional correlations to three 2D correlations. Then, the two-dimensional correlations are used to build conceptual structures similar to hyperlink structures that are used by Web search engines.

The work in [98,92] has shown to generate high quality tags recommendations that outperform baseline methods such as the most-popular models and collaborative filtering [96]. Also, Krestel et al.[99] proposed an approach to use Latent Dirichlet Allocation to expand tag sets of objects annotated by only a few users. The work in [100] proposed to learn tag relevance by taking into account three kinds of correlations: tag co-occurrence, tag visual correlation, and image conditioned tag correlation. Specifically, they adopted the Rankboost [101] to learn an optimal combination of these multi-modality correlations, and generated a ranking function for tag recommendation. Recently, Zhu et al. [102] proposed a method of tag recommendation based on the neighbor voting graph of tags. They casted the social tag relevance learning problem as an adaptive teleportation random walk process on the voting graph.

Finally, several approaches have been proposed making these three types of recommendations, i.e. social items, users, and topics recommendations, unified under one framework. Carmel et al. [47] propose a framework for social bookmarks weighting, which allows estimating the effectiveness of each of the bookmarks individually for several IR tasks. To do this, they propose several recommendation strategies such as tag recommendation, user recommendation, and document recommendation. The obtained values from the three strategies are merged in order to effectively estimate the bookmarks quality. In the same spirit, the framework proposed by Symeonidid et al. [92] acts in the same way. This framework models the three types of entities by a 3-order tensor, on which multiway latent semantic analysis and dimensionality reduction is performed using both the Singular Value Decomposition method and the Kernel-SVD smoothing technique. Also, Wei et al. [103] propose to leverage a quaternary relationship among users, items, tags and ratings to provide recommendation. They propose a unified framework for user recommendation, item recommendation, tag recommendation and item rating prediction by modeling the quaternary relationship among users, resources, tags, and ratings as a 4-order tensor and cast the recommendation problem as a multi-way latent semantic analysis problem.

It is clear that social networks represent a valuable source of information to improve and develop effective and efficient recommendation algorithms. In the next section, we give an overview of many platforms, methods and approaches of SIR as well as their potential drawbacks. The main objective is to understand the impact of the social dimension on the IR process, the weaknesses and the possible future contributions in this domain.

7. Analysis of SIR methods and platforms

The objective in this section is to analyze the richness and weaknesses of some SIR tools and approaches. Among the discussed SIR tools and methods in the previous sections, only few of them have spawned to a concrete commercial prototype. The selected tools for analysis are provided in Table 1, and their description was provided in the previous sections. We have selected these approaches and tools following these main criteria: (i) they are the most popular ones when this analysis is performed (judging from our research), (ii) they are the most referenced ones and are published in top venue conferences,[24] and finally, (iii) this limit is motivated by the fact that our objective is not to analyze all the existing approaches and tools, but the most representative ones as illustration of the underlying principles and ideas.

For this analysis, we choose the dimensions listed in Table 1 covering various aspects of the social information used and the problem to be solved. We consider these dimension as crucial for designing an effective and efficient SIR approach. For illustration purposes, we only discuss some of the above tools for each dimension. Table 1 summarizes the analysis dimensions and the classification of the different tools according to these dimensions. In the following sections, we discuss these dimensions, their meaning, their importance, and the extent to which they are considered by the tools.

[24] A highly selective conference like SIGIR, WWW, CIKM, VLDB, WSDM, etc.

7.1. Social networks

This dimension is related to the kind of social networks leveraged by a SIR approach. Each SIR approach relies on almost one kind of social network either by using its social content or by exploiting its social relations. Depending on the considered approach, its application and its purpose, it can use

- Social networks with symmetric relationships in which users explicitly declare their social relations of friends. Users can also express their opinions, comment news, and share resources. This represents a valuable source of information for user modeling and profiling [57,66], which can be reused to build many interesting services, e.g., services of recommendation and personalization.
- Social networks with asymmetric relationships, which can be divided into two categories: (i) content subscription social networks, e.g., microblogging systems, in which most of the considered SIR approaches relies on using the temporal aspect of these networks, e.g., answering recency sensitive queries [36] and a (ii) trust-based Social Network, e.g., *epinion*,[25] in which SIR approaches use the trust degree between users, e.g., for collaborative filtering [16].
- Social networks with ternary relations, e.g., bookmarking systems. As discussed in Section 2.1, the generated structures in these systems have been proven to be a valuable knowledge for building many SIR approaches. Exploiting annotations and social metadata explicitly provided by users has been used in different way, for example to extract correlated terms [22], build users' profiles [39], compute social relevance [31], and enhance documents [46].
- Their own social networks built upon information gathered from (i) the user, by explicitly providing information about him, (ii) the aggregation and crawling of both social networks and the Web, e.g., building a FOAF (Friend Of A Friend) ontology, which is a machine-readable structure that describes people, or (iii) by an inference process based on the user behavior, e.g., Aardvark [57].

7.2. Social data

This dimension concerns the kind of social information leveraged by SIR approaches. We distinguish the following three kinds of social information, which are embedded within a social network:

- Social content, which is the content generated by users through their interactions and activities, i.e. activities of publishing, annotating, commenting or rating content and entities. This social content is most of the time useful to: enrich entities and thus enhance their logical representation [46,47,36], extract correlated terms, e.g., through co-occurrence of terms [22], and pull out user profiles [39,57].
- Social relations explicitly or implicitly declared by users. Indeed, social networks exhibit various relationships, between entities of their social graphs. These relationships can be within entities of the same type, e.g., friendship relations between users and similarity between resources, or between entities of different types, e.g., an authorship relation between a user and a document, a relation of description between a term and a user, etc. Social relations are useful for building a SIR approach, and have been used in different ways, e.g., for recommendation [16] by handling trust relations, for extracting correlated terms by leveraging their relations over document [22], etc.
- Social metadata refer to information embedded inside social platforms and data such as geo-location information and time-stamps. Metadata can be easily exploited for a social search task (e.g., forwarding a geographically contextualized query to the right person [57]) or a social recommendation purpose (e.g., recommending the right event at the right place [104]).

7.3. Data sources

The data sources dimension refers to the source of information used by SIR approaches. These latter are most of the time not only based on social data sources, but also on other sources of information like:

- The content of web pages, which contains valuable information that can be extracted by performing classic textual treatments [22,47,46].
- Ontologies, e.g., FOAF ontology (Friend Of A Friend), which can be used to infer a social network [105].
- Information explicitly provided by users. Some SIR approaches ask users to explicitly provide information to enrich their internal data model [57].

Note that basically, social information is only used to improve and enhance information retrieval tasks. Hence, we believe that the use of trusted sources of information (e.g., information provided by authors of web pages) is necessary to the proper functioning of SIR approaches (actually, almost all SIR approaches combine such data sources [47,46]). A trade-off should be found to properly weight each data source, while avoiding overfitting.

7.4. Personalization

As discussed in the previous sections, some SIR approaches are based on personalization, e.g. [29,45,38,39,57,26,43]. Personalization allows differentiating between individuals by emphasizing on their specific domains of interest and their preferences. It is a key point in IR and its demand is constantly increasing by numerous users for adapting their results [106]. Several techniques exist to provide personalized services among which: (i) the user profiling, (ii) content based approaches, and (iii) collaborative based approaches.

7.4.1. Profile based methods

A user profile is a collection of data describing a specific user, which are explicitly or implicitly provided by him. Therefore, a profile refers to the digital representation of a person's identity, which includes mainly the description of

[25] http://www.epinions.com/

his characteristics and his domains of interest and expertise. We distinguish two types of profiles that differ in the way they are constructed:

- *Profile constructed offline*: Some SIR approaches construct profiles offline and maintain them incrementally, which make them more efficient regarding the execution time. However, profiles computed offline decreases the dynamics of the approach since new data are not instantly taken into account.
- *Profile constructed on the fly (online)*: In contrast, profiles computed online increase the dynamics of a SIR approach and its efficiency, while degrading the execution time.

Examples of work focusing their strategies on user profiling include [39,43,29,38,45]. The main issue with the usage of profiles is related to the ability of updating the user's preferences quickly, especially because users interact a lot on social media. Thus, this problem has to be considered in order to make use of dynamic profiles.

7.4.2. Content-based methods

Personalization in content based approaches aims at providing to users, information similar to that they previously consumed. This category of methods best suits for recommender systems where items and users' preferences are usually described with keywords. The algorithms try then to recommend items that are similar to those that a user liked in the past. In particular, various candidate items are compared with items previously rated by the user and the best-matching items are recommended [92,81].

7.4.3. Collaborative-filtering based methods

Personalization based on collaborative filtering process aims to provide to users, information consumed by many similar users. This process of collaborative filtering can be summarized to two steps: (1) look for users who share the same interests and behavior with a given user, i.e. the user who is currently using the SIR approach, and (2) provide to the considered user, contents and information based on the users found in step (1). Examples of representative efforts in this category include [16,87,52].

7.5. Complexity and applicability

The complexity of a method tends to characterize it from several perspectives of its applicability. We consider four dimensions to describe and characterize the complexity and the applicability of a method.

7.5.1. Scalability

Scalability is the ability of a method to continue to work well when its context is changed. This refers to its ability to scale to a very large dataset (number of users, resources, objects, documents, etc.) while continuing to meet the users' needs (both precision and recall) and to take full advantage of it in terms of performance, e.g., execution time and relevance of information. Some SIR approaches are able to scale to very large datasets [16,86,87], while other not [47] because of their algorithms' complexity.

7.5.2. Dynamicity

Dynamicity refers to the ability of a method to consider new data and to quickly update its model. Considering new data is a key problem for SIR methods since they are based on social information, which is growing quickly with the intense activity of users who are constantly in the process of commenting, editing, publishing and sharing information. Some SIR methods have a model, which can be easily updated [29,39,57], whereas other methods not, e.g., methods based on machine learning techniques, which are among the most difficult to update since we have to rebuild the model each time in order to consider new data [16,83,86,22].

7.5.3. Data sparsity

Data sparsity is a term used to designate how much data we have for the dimensions of a dataset. The term of sparse data is most of the time associated with matrices, where a sparse matrix is a matrix populated primarily with zeros [107]. Therefore, this dimension refers to the ability of an approach to process over sparse data. Indeed, sparse data leads to the problem of poor results quality since there is no enough information to process. However, some approaches handle effectively this problem by considering other sources of data, e.g., content of web pages or other social networks [16,29,57,22].

7.5.4. Cold-start problem

The cold start is a potential problem of systems to handle effectively new entities [78], e.g., users and items. In other words, it concerns the issue that the system cannot draw any inference for users or items about which it has not yet gathered sufficient information. Recommender systems are among the most affected systems by the cold-start problem, since they need a lot of information to make predictions. However, many approaches deal efficiently with this problem by relying on other data sources [39,16,83,86].

7.6. Evaluation

Evaluation is a critical part of any research work. This dimension intends to track the evaluation methodology followed by the proposed methods. We distinguish two main categories of approaches for evaluation: (i) formal evaluation, where the evaluation is performed off-line using a public or a specifically built dataset, and (ii) end-user evaluation, where we generally ask users to judge and provide feedback on a method. In general, almost all the methods we studied involve a formal evaluation, where different aspects are studied with a comparison to the state-of-the-art. Few of them, such as [39,14,72] consider end-user evaluation.

7.7. Socialization

Some SIR approaches put users in contact, and encourage them to socialize and collaborate in order to satisfy a particular goal of information needs. This dimension refers to this particularity of socialization between users. Thus, many approaches put users in contact in order to benefit from the experience of each other. Especially, many social

recommender systems offer this feature, since they provide explanations of why a recommendation has been done, and what yielded to the recommendation [81,90,82]. Then, if the recommendation is based on the experience of other users, e.g., rating of other users, the current user can contact them to obtain other information and precision. Also, as stated before, social search approaches are exclusively based on the assistance of social entities. This will automatically put users in contact as it is the case in *Q&A* systems (Aadvark [57]), and social collaborative search systems (*SearchTogether* [72], *Kolline* [73], and *sensemaking* [74]).

7.8. Privacy management

Social information is sensitive to the privacy of users. Some approaches do not consider the privacy of users, as they spread sensitive information about users. Even if users make their social accounts public, we believe that reusing these data in other value-added services is still a user privacy problem. Hence, probably social recommender systems and social search systems are the most related and exposed to users' privacy problems. Examples where we need to divulge information about users include

1. In recommender systems, we need to justify why a recommendation has been done.
2. In *Q&A* system, we need to justify why we put two users in contact.
3. In social content search systems, we show who said what on what.
4. In social collaborative search systems, we show who searched what.

Finally, many approaches are using social information in back-office, and thus, do not spread sensitive information about users. This includes mainly social Web search approaches [22,39,47,46,43].

7.9. Industrialization

Developing effective and efficient algorithms is good, but putting them in action is better. Hence, this final dimension is related to an aspect of production and commissioning. Many SIR approaches studied in this paper have spanned to a concrete and commercial prototype, e.g., Aadvark [57], *SearchTogether* [72], or collecta [66], while other approaches are still as research contributions. This is certainly a big limitation of most of the contributions, i.e., their inability to span to an industrial level even with the interest the approaches may have in this domain.

8. Discussion and future directions

In this section, we discuss some thoughts and future directions for research related to SIR. We consider two perspectives: (i) category perspective, where the discussed aspects are related the categories of the SIR approaches, and (ii) dimension perspective, where the discussion is directed by considering the analysis dimensions.

8.1. Approaches' category perspective

In the social Web Search, especially in the social query reformulation part, we reported that there is no approach that considers query reduction. Query reduction is a technique to reduce long queries, to queries, which are shorter and more effective [17]. Hence, investigating a social query reformulation approach by considering both query expansion and reduction, seems to be promising and provide good research perspectives. This can be by providing users with a way to suggest them other queries that can better match their requirements.

Example 9. The query “interesting tickets for four people to warm weather countries for vacation” can be rewritten as “cheap family vacation ticket” or “luxury family vacation plan”. The new proposed query can be more precise/concise and is eventually expected to provide personalized, similar or even better results than the long one.

Regarding the IR modeling part (documents representation and ranking functions), we believe that the temporal dimension is a key aspect, which has not been deeply investigated. This includes considering the evolution of users' behavior, and profiles in time. Indeed, users are expected to evolve in time, which make their interests/tastes different. IR systems should constantly learn information about users in order to adapt their results.

As for the social recommendation track, this topic has been deeply investigated, whether for items, tags, or users recommendation. Items recommendation is probably the part that attracted the most attention. However, we also believe that the temporal dimension of recommender systems can help to tackle the problems of (i) hot topics, i.e. news, fresh information, (ii) the evolution of user profiles along time, i.e. the user interests evolve and change along time, and (iii) the diversity of information, i.e. in order not to annoy users with similar information. Indeed, for the first problem, information is time-dependent, meaning that it attracts much attention at a given moment and will be quickly forgotten after a while. Many users of social media state that the freshness of information is a key point in a recommender system. The second problem deals with the evolution and the update of user profiles. For example, the opinion of a user concerning something may change in time when he grows, or reads news about this topic. The third problem is also a feeling that users have when they use social media (Facebook in particular). Most of the time, when a recent information appears, all users begin to publish articles that deal with the same information, and users are quickly overwhelmed by similar information published mutually by each other. We believe that, at a given time, the recommender system should know that a given user is already aware about this information and consequently it should be hidden [87].

Finally, with the advent of social networks, many tools and approaches have emerged to deal with the weakness of classic search engines. This includes social content search engines (for opinion seeking), query recency search (Q&A tools for contextualized search), and collaborative search engine (for collectively answering a common information

need). In these areas, many improvement still possible like finding the right persons for answering a given question in Q&A tools or how to find users who share common information needs and relate them to work together to fulfill this need.

8.2. Dimension perspective

From the data perspective, most of the existing approaches use either content or structure of social networks. Although few techniques started leveraging both parts, there is still a room of improvement to encourage this usage, since it has been shown that it may help [82]. Furthermore, it is interesting to note that social meta-data have been almost not used. Considering this type of data may bring additional capabilities to a method, e.g., contextual information retrieval. The same observation can also apply on the data sources dimension, where approaches tend to leverage either one source or another. Since the social Web can be seen as a complementary level of the traditional Web, combining the different sources can provide more capabilities for any approach.

On the other hand, exploiting personal information resulting from social interactions should naturally benefit to the user. This is achieved through personalization, which is added to the SIR process. More efforts have to be put towards reinforcing personalization in SIR. This is also justified by the huge amount of information available to the user, which need to be handled on behalf of him, e.g., to display the most interesting piece of information at the right moment and at the right place. Moreover, it is pretty obvious to consider the capability of any approach to provide a socialization functionality as bootstrapping mechanism. We observed that many methods use social information but do not generate or favor the generation of social information in their turn. This can certainly be benefit to them, e.g., by having a better characterization and contextualization of the user as done in [57].

Social networks contain sensitive data directly related to the user. Data privacy is a big challenge that has to be considered when designing a new SIR method. To handle this issue, the current techniques and methods leverage the “public” part of the social data, i.e., the part of the data that is open to public and for which the user gave an explicit consent to share it publicly. However, we believe that there is still huge part of the data, which can be leveraged to further improve the IR process. Investigations have to be made in this area following innovative techniques such as information granularity (similar to what is practiced in medical informatics [108]).

Dealing with social data implies being confronted to their large size, their diversity, and their dynamics. This dimension has to be considered as a priority when designing a new SIR approach, as the social data is inherently large and complex. Some existing techniques are considering a part of it, e.g., tackling the size but not considering the dynamics. Thus, there is a need for more investigation to solve scalability issues related to SIR, especially if any method targets a production objective. These issues can be handled with existing technologies such as *Hadoop* and its associated technologies, but has to be also investigated from the algorithmic perspective as the considered data and structures are often complex.

There is also a room for contribution from the evaluation perspective of SIR systems and approaches. In fact, although heavy evaluation is operated using different data sets and comparisons to other approaches, more evaluations and protocols have to be operated and set-up to involve real users in this process. As mentioned in the previous sections, some contributions started involving users but not in heavy manner. Beyond involving users, working on elaborated protocols to involve them may help a lot in this process. Last but not least, and in order to ensure a certain reproduction of the evaluation results, there should be an effort in sharing the code of the approaches in the community to ensure transparency, efficiency, and objectivity in this domain.

9. Conclusion

In this paper, we proposed a deep review of the topic of Social Information Retrieval (SIR). Especially, we proposed a taxonomy to classify and categorize SIR approaches into three main categories, namely: (i) social Web search, (ii) social search, and (iii) social recommendation. We showed that these three sub-categories are fundamentally different in the way they leverage and use social information. Many methods have been proposed to improve the classic IR process, while others proposed new search paradigms based on the socialization between users.

Social platforms represent valuable sources of information and knowledge that can be reused to improve many services (especially search services). On one hand, initiatives like OpenID, and the Mashup concept are promoting the possibility to social platforms to share their data with other applications, e.g., the full user name, the profile's picture, the gender, the username and user id (account number), and even the list of friends and contacts. On the other hand, users are encouraged to share and publish content on social platforms, and make their social information publicly accessible. Therefore, this social dimension of the Web attracted the attention of many researchers in the IR community.

We discussed many SIR methods with respect to some dimensions that we judge to be essential in order to assess the robustness and the effectiveness of a SIR approach. We consider these dimensions, as factors that indicate the extent to which a SIR approach can be applicable, and in what context, i.e. size of the data and sparsity of the data. Finally, we discussed some interesting improvement for some SIR categories for a research perspective.

References

[1] S. Amer-Yahia, J. Huang, C. Yu, Building community-centric information exploration applications on social content sites, in: Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data, SIGMOD '09, ACM, New York, NY, USA, 2009, pp. 947–952.
[2] S. Amer-Yahia, L.V.S. Lakshmanan, C. Yu, Socialscope: enabling information discovery on social content sites, in: CIDR, 2009.

[3] R.A. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval, 2nd edition, . Addison-Wesley Longman Publishing Co., Inc, Boston, MA, USA, 2010.
[4] S.E. Robertson, K. Sparck Jones, Document Retrieval Systems, Taylor Graham Publishing, London, UK, 1988 Ch. Relevance weighting of search terms, pp. 143–160.
[5] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, Comput. Netw. ISDN Syst. 30 (1–7) (1998) 107–117.
[6] J.M. Kleinberg, Authoritative sources in a hyperlinked environment, J. ACM 46 (5) (1999) 604–632.
[7] G. Salton, Automatic Information Organization and Retrieval, McGraw Hill Text, 1968.
[8] A. Maaradji, H. Hacid, R. Skraba, A. Vakali, Social web mashups full completion via frequent sequence mining, in: SERVICES, 2011, pp. 9–16.
[9] W. Stanley, F. Katherine, Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences), 1st edition, Cambridge University Press, New York, 1994.
[10] T. Hammond, T. Hannay, B. Lund, J. Scott, Social bookmarking tools: a general review, D-Lib Mag. 11 (4),. (2005).
[11] B. Lund, T. Hammond, T. Hannay, M. Flack, Social bookmarking tools (ii): a case study—connotea, D-Lib Mag. 11 (4),. (2005).
[12] D. Goh, S. Foo, Social Information Retrieval Systems: Emerging Technologies and Applications for Searching the Web Effectively, Information Science Reference—Imprint of: IGI Publishing, Hershey, PA, 2007.
[13] M. Gupta, R. Li, Z. Yin, J. Han, Survey on social tagging techniques, SIGKDD Explor. Newsl. 12 (1) (2010) 58–72.
[14] P. Heymann, G. Koutrika, H. Garcia-Molina, Can social bookmarking improve web search? in: Proceedings of the 2008 International Conference on Web Search and Data Mining, WSDM '08, ACM, New York, NY, USA, 2008, pp. 195–206.
[15] M.R. Morris, J. Teevan, K. Panovich, What do people ask their social networks, and why? A survey study of status message Q&A behavior, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '10, ACM, New York, NY, USA, 2010, pp. 1739–1748.
[16] H. Ma, H. Yang, M.R. Lyu, I. King, Sorec: social recommendation using probabilistic matrix factorization, in: Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM '08, ACM, New York, NY, USA, 2008, pp. 931–940.
[17] G. Kumaran, V.R. Carvalho, Reducing long queries using query quality predictors, in: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '09, ACM, New York, NY, USA, 2009, pp. 564–571.
[18] Efthimis N. Efthimiadis, Interactive query expansion: a user-based evaluation in a relevance feedback environment, J. Am. Soc. Inf. Sci. 51, 11 (2000) 989–1003, http://dx.doi.org/10.1002/1097-4571.
[19] C. Lioma, R. Blanco, M.-F. Moens, A logical inference approach to query expansion with social tags, in: ICTIR, 2009.
[20] S. Jin, H. Lin, S. Su, Query expansion based on folksonomy tag co-occurrence analysis, in: 2009 IEEE International Conference on Granular Computing, 2009.
[21] A. Mantrach, J.-M. Renders, A general framework for people retrieval in social media with multiple roles, in: R. Baeza-Yates, A. de Vries, H. Zaragoza, B. Cambazoglu, V. Murdock, R. Lempel, F. Silvestri (Eds.), Advances in Information Retrieval, Lecture Notes in Computer Science, vol. 7224, Springer, Berlin, Heidelberg, 2012, pp. 512–516.
[22] Y. Lin, H. Lin, S. Jin, Z. Ye, Social annotation in query expansion: a machine learning approach, in: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '11, ACM, New York, NY, USA, 2011, pp. 405–414.
[23] J. Pitkow, H. Schütze, T. Cass, R. Cooley, D. Turnbull, A. Edmonds, E. Adar, T. Breuel, Personalized search, Commun. ACM 45 (9) (2002) 50–55.
[24] M. Bender, T. Crecelius, M. Kacimi, S. Michel, T. Neumann, J.X. Parreira, R. Schenkel, G. Weikum, Exploiting social relations for query expansion and result ranking, in: ICDE Workshops, 2008.
[25] R. Schenkel, T. Crecelius, M. Kacimi, S. Michel, T. Neumann, J.X. Parreira, G. Weikum, Efficient top-k querying over social-tagging networks, in: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '08, ACM, New York, NY, USA, 2008, pp. 523–530.
[26] M. Bertier, R. Guerraoui, V. Leroy, A.-M. Kermarrec, Toward personalized query expansion, in: SNS, 2009.
[27] C. Biancalana, A. Micarelli, C. Squarcella, Nereau: a social approach to query expansion, in: WIDM, 2008.
[28] C. Biancalana, F. Gasparetti, A. Micarelli, G. Sansonetti, Social semantic query expansion, ACM Trans. Intell. Syst. Technol. 4 (4) (2013) 60:1–60:43.
[29] M.R. Bouadjenek, H. Hacid, M. Bouzeghoub, J. Daigremont, Personalized social query expansion using social bookmarking systems, in: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '11, ACM, New York, NY, USA, 2011, pp. 1113–1114.
[30] A. Hotho, R. Jäschke, C. Schmitz, G. Stumme, Information retrieval in folksonomies: search and ranking, in: Y. Sure, J. Domingue (Eds.), The Semantic Web: Research and Applications, 2006.
[31] S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, Z. Su, Optimizing web search using social annotations, in: Proceedings of the 16th International Conference on World Wide Web, WWW '07, ACM, New York, NY, USA, 2007, pp. 501–510.
[32] T. Takahashi, H. Kitagawa, S-bits: social-bookmarking induced topic search, in: Proceedings of the 2008 the Ninth International Conference on Web-Age Information Management, WAIM '08, IEEE Computer Society, Washington, DC, USA, 2008, pp. 25–30.
[33] T. Takahashi, H. Kitagawa, A ranking method for web search using social bookmarks, in: Proceedings of the 14th International Conference on Database Systems for Advanced Applications, DASFAA '09, Springer-Verlag, Berlin, Heidelberg, 2009, pp. 585–589.
[34] Y. Yanbe, A. Jatowt, S. Nakamura, K. Tanaka, Towards improving web search by utilizing social bookmarks, in: Proceedings of the 7th International Conference on Web Engineering, ICWE'07, Springer-Verlag, Berlin, Heidelberg, 2007, pp. 343–357.
[35] F. Abel, N. Henze, D. Krause, Ranking in folksonomy systems: can context help? in: Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM '08, ACM, New York, NY, USA, 2008, pp. 1429–1430.
[36] A. Dong, R. Zhang, P. Kolari, J. Bai, F. Diaz, Y. Chang, Z. Zheng, H. Zha, Time is of the essence: improving recency ranking using twitter data, in: Proceedings of the 19th International Conference on World Wide Web, WWW '10, ACM, New York, NY, USA, 2010, pp. 331–340.
[37] X. He, M. Gao, M.-Y. Kan, Y. Liu, K. Sugiyama, Predicting the popularity of web 2.0 items based on user comments, in: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '14, ACM, New York, NY, USA, 2014, pp. 233–242.
[38] M.R. Bouadjenek, H. Hacid, M. Bouzeghoub, SoPRa: a new social personalized ranking function for improving web search, in: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '13, ACM, New York, NY, USA, 2013, pp. 861–864.
[39] D. Carmel, N. Zwerdling, I. Guy, S. Ofek-Koifman, N. Har'el, I. Ronen, E. Uziel, S. Yogev, S. Chernov, Personalized social search based on the user's social network, in: Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM '09, ACM, New York, NY, USA, 2009, pp. 1227–1236.
[40] M.G. Noll, C. Meinel, Web search personalization via social bookmarking and tagging, in: ISWC'07/ASWC'07, 2007.
[41] D. Vallet, I. Cantador, J.M. Jose, Personalizing web search with folksonomy-based user and document profiles, in: Proceedings of the 32nd European Conference on Advances in Information Retrieval, ECIR'2010, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 420–431.
[42] Q. Wang, H. Jin, Exploring online social activities for adaptive search personalization, in: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, CIKM '10, ACM, New York, NY, USA, 2010, pp. 999–1008.
[43] S. Xu, S. Bao, B. Fei, Z. Su, Y. Yu, Exploring folksonomy for personalized search, in: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '08, ACM, New York, NY, USA, 2008, pp. 155–162.
[44] M.R. Bouadjenek, A. Bennamane, H. Hacid, M. Bouzeghoub, Evaluation of personalized social ranking functions of information retrieval, in: F. Daniel, P. Dolog, Q. Li (Eds.), Web Engineering, Lecture Notes in Computer Science, vol. 7977, Springer, Berlin, Heidelberg, 2013, pp. 283–290.
[45] M.R. Bouadjenek, H. Hacid, M. Bouzeghoub, A. Vakali, Using social annotations to enhance document representation for personalized search, in: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '13, ACM, New York, NY, USA, 2013, pp. 1049–1052.
[46] X. Zhang, L. Yang, X. Wu, H. Guo, Z. Guo, S. Bao, Y. Yu, Z. Su, SDOC: exploring social wisdom for document enhancement in web

mining, in: Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM '09, ACM, New York, NY, USA, 2009, pp. 395–404.
[47] D. Carmel, H. Roitman, E. Yom-Tov, Social bookmark weighting for search and recommendation, VLDB J. 19 (6) (2010) 761–775.
[48] P.A. Dmitriev, N. Eiron, M. Fontoura, E. Shekita, Using annotations in enterprise search, in: Proceedings of the 15th International Conference on World Wide Web, WWW '06, ACM, New York, NY, USA, 2006, pp. 811–817.
[49] K. Bischoff, C.S. Firan, W. Nejdl, R. Paiu, Can all tags be used for search? in: Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM '08, ACM, New York, NY, USA, 2008, pp. 193–202.
[50] A. Singhal, F. Pereira, Document expansion for speech retrieval, in: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '99, ACM, New York, NY, USA, 1999, pp. 34–41.
[51] D. Carmel, H. Roitman, E. Yom-Tov, Who tags the tags? A framework for bookmark weighting, in: Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM '09, ACM, New York, NY, USA, 2009, pp. 1577–1580.
[52] S.-Y. Chen, Y. Zhang, Improve Web Search Ranking with Social Tagging, MSM.
[53] G. Salton, A. Wong, C.S. Yang, A vector space model for automatic indexing, Commun. ACM 18 (11) (1975) 613–620.
[54] S. Amer-Yahia, M. Benedikt, L.V.S. Lakshmanan, J. Stoyanovich, Efficient network aware search in collaborative tagging sites, Proc. VLDB Endow. 1 (1) (2008) 710–721.
[55] Sihem Amer Yahia, Michael Benedikt, Laks V.S. Lakshmanan, Julia Stoyanovich, Efficient network aware search in collaborative tagging sites, Proc. VLDB Endow 1 (1) (2008) 710–721. http://dx.doi.org/10.14778/1453856.1453934.
[56] J. Teevan, D. Ramage, M.R. Morris, #twittersearch: a comparison of microblog search and web search, in: Proceedings of the Fourth ACM International conference on Web Search and Data Mining, WSDM '11, ACM, New York, NY, USA, 2011, pp. 35–44.
[57] D. Horowitz, 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, ACM, New York, NY, USA, 2010, pp. 431–440.
[58] A. Broder, A taxonomy of web search, SIGIR Forum 36 (2) (2002) 3–10.
[59] P. Anick, R.G. Kantamneni, A longitudinal study of real-time search assistance adoption, in: Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '08, ACM, New York, NY, USA, 2008, pp. 701–702.
[60] R.W. White, G. Marchionini, Examining the effectiveness of real-time query expansion, Inf. Process. Manag. 43 (3) (2007) 685–704.
[61] G. Dror, Y. Koren, Y. Maarek, I. Szpektor, I want to answer; who has a question? Yahoo! answers recommender system, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '11, ACM, New York, NY, USA, 2011, pp. 1109–1117.
[62] J. Zhang, M.S. Ackerman, L. Adamic, K.K. Nam, Qume: a mechanism to support expertise finding in online help-seeking communities, in: Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology, UIST '07, ACM, New York, NY, USA, 2007, pp. 111–114.
[63] G. Hsieh, S. Counts, Mimir: a market-based real-time question and answer service, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '09, ACM, New York, NY, USA, 2009, pp. 769–778.
[64] D.H. Dalip, M.A. Gonçalves, M. Cristo, P. Calado, Exploiting user feedback to learn to rank answers in q&a forums: a case study with stack overflow, in: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '13, ACM, New York, NY, USA, 2013, pp. 543–552.
[65] B.J. Jansen, A. Chowdury, G. Cook, The ubiquitous and increasingly significant status message, Interactions 17 (3) (2010) 15–17.
[66] B.J. Jansen, Z. Liu, C. Weaver, G. Campbell, M. Gregg, Real time search on the web: queries, topics, and economic value, Inf. Process. Manag. 47 (4) (2011) 491–506.
[67] T. Peggs, The inner workings of a real time search engine: Thoughts on realtime search, 2009 (white paper).
[68] M. Efron, J. Lin, J. He, A. de Vries, Temporal feedback for tweet search with non-parametric density estimation, in: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '14, ACM, New York, NY, USA, 2014, pp. 33–42.
[69] J. Vosecky, K.W.-T. Leung, W. Ng, Collaborative personalized twitter search with topic-language models, in: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '14, ACM, New York, NY, USA, 2014, pp. 53–62.
[70] M.R. Morris, Collaborating alone and together: investigating persistent and multi-user web search activities, Technical Report, Microsoft Research, 2007.
[71] M.R. Morris, A survey of collaborative web search practices, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '08, ACM, New York, NY, USA, 2008, pp. 1657–1660.
[72] M.R. Morris, E. Horvitz, Searchtogether: an interface for collaborative web search, in: Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology, UIST '07, ACM, New York, NY, USA, 2007, pp. 3–12.
[73] F.F. Filho, G.M. Olson, P.L. de Geus, Kolline: a task-oriented system for collaborative information seeking, in: Proceedings of the 28th ACM International Conference on Design of Communication, SIGDOC '10, ACM, New York, NY, USA, 2010, pp. 89–94.
[74] S.A. Paul, M.R. Morris, Cosense: enhancing sensemaking for collaborative web search, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '09, ACM, New York, NY, USA, 2009, pp. 1771–1780.
[75] H. Ma, An experimental study on implicit social recommendation, in: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '13, ACM, New York, NY, USA, 2013, pp. 73–82.
[76] L. Chen, P. Wei, W. Zhe, Hete-cf: Social-based collaborative filtering recommendation using heterogeneous relations, in: The IEEE International Conference on Data Mining, ICDM'14, 2014.
[77] J. Lin, K. Sugiyama, M.-Y. Kan, T.-S. Chua, Addressing cold-start in app recommendation: latent user models constructed from twitter followers, in: Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '13, ACM, New York, NY, USA, 2013, pp. 283–292.
[78] S. Sedhain, S. Sanner, D. Braziunas, L. Xie, J. Christensen, Social collaborative filtering for cold-start recommendations, in: Proceedings of the 8th ACM Conference on Recommender Systems, RecSys '14, ACM, New York, NY, USA, 2014, pp. 345–348.
[79] S. Sedhain, S. Sanner, L. Xie, R. Kidd, K.-N. Tran, P. Christen, Social affinity filtering: Recommendation through fine-grained analysis of user interactions and activities, in: Proceedings of the First ACM Conference on Online Social Networks, COSN '13, ACM, New York, NY, USA, 2013, pp. 51–62.
[80] F. Liu, H.J. Lee, Use of social network information to enhance collaborative filtering performance, Expert Syst. Appl. 37 (2010) 4772–4778.
[81] I. Guy, N. Zwerdling, D. Carmel, I. Ronen, E. Uziel, S. Yogev, S. Ofek-Koifman, Personalized recommendation of social software items based on social relations, in: Proceedings of the third ACM conference on Recommender systems, RecSys '09, ACM, New York, NY, USA, 2009, pp. 53–60.
[82] I. Guy, N. Zwerdling, I. Ronen, D. Carmel, E. Uziel, Social media recommendation based on people and tags, in: Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '10, ACM, New York, NY, USA, 2010, pp. 194–201.
[83] H. Ma, I. King, M.R. Lyu, Learning to recommend with social trust ensemble, in: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '09, ACM, New York, NY, USA, 2009, pp. 203–210.
[84] M. Jamali, M. Ester, A matrix factorization technique with trust propagation for recommendation in social networks, in: Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys '10, ACM, New York, NY, USA, 2010, pp. 135–142.
[85] X. Yang, H. Steck, Y. Liu, Circle-based recommendation in online social networks, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12, ACM, New York, NY, USA, 2012, pp. 1267–1275.
[86] H. Ma, D. Zhou, C. Liu, M.R. Lyu, I. King, Recommender systems with social regularization, in: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM '11, ACM, New York, NY, USA, 2011, pp. 287–296.
[87] J. Noel, S. Sanner, K.-N. Tran, P. Christen, L. Xie, E.V. Bonilla, E. Abbasnejad, N. Della Penna, New objective functions for social collaborative filtering, in: Proceedings of the 21st International

Conference on World Wide Web, WWW '12, ACM, New York, NY, USA, 2012, pp. 859–868.
[88] L. Yu, R. Pan, Z. Li, Adaptive social similarities for recommender systems, in: Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys '11, ACM, New York, NY, USA, 2011, pp. 257–260.
[89] X. Yang, Y. Guo, Y. Liu, H. Steck, A survey of collaborative filtering based social recommender systems, Comput. Commun. 41 (0) (2014) 1–10.
[90] J. Chen, W. Geyer, C. Dugan, M. Muller, I. Guy, Make new friends, but keep the old: recommending people on social networking sites, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '09, ACM, New York, NY, USA, 2009, pp. 201–210.
[91] G. Groh, C. Ehmig, Recommendations in taste related domains: collaborative filtering vs. social filtering, in: Proceedings of the 2007 International ACM Conference on Supporting Group Work, GROUP '07, ACM, New York, NY, USA, 2007, pp. 127–136.
[92] P. Symeonidis, A. Nanopoulos, Y. Manolopoulos, "A Unified Framework for Providing Recommendations in Social Tagging Systems Based on Ternary Semantic Analysis," in Knowledge and Data Engineering, IEEE Transactions on, 22 (2) (2010) 179–192, http://dx.doi.org/10.1109/TKDE.2009.85.
[93] J. Hannon, M. Bennett, B. Smyth, Recommending twitter users to follow using content and collaborative filtering approaches, in: Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys '10, ACM, New York, NY, USA, 2010, pp. 199–206.
[94] J. Hannon, K. McCarthy, B. Smyth, Finding useful users on twitter: twittomender the followee recommender, in: Proceedings of the 33rd European Conference on Advances in Information Retrieval, ECIR'11, Springer-Verlag, Berlin, Heidelberg, 2011, pp. 784–787.
[95] X. Wang, H. Liu, W. Fan, Connecting users with similar interests via tag network inference, in: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM '11, ACM, New York, NY, USA, 2011, pp. 1019–1024.
[96] Robert Jäschke, Leandro Marinho, Andreas Hotho, Lars Schmidt-Thieme, Gerd Stumme, Tag recommendations in social bookmarking systems, AI Commun 21 (2008) 231–247.
[97] G. Mishne, Autotag: a collaborative approach to automated tag assignment for weblog posts, in: Proceedings of the 15th International Conference on World Wide Web, WWW '06, ACM, New York, NY, USA, 2006, pp. 953–954.
[98] A. Hotho, R. Jäschke, C. Schmitz, G. Stumme, Folkrank: a ranking algorithm for folksonomies., in: LWA, 2006.
[99] R. Krestel, P. Fankhauser, W. Nejdl, Latent Dirichlet allocation for tag recommendation, in: Proceedings of the Third ACM Conference on Recommender Systems, RecSys '09, ACM, New York, NY, USA, 2009, pp. 61–68.
[100] L. Wu, L. Yang, N. Yu, X.-S. Hua, Learning to tag, in: Proceedings of the 18th International Conference on World Wide Web, WWW '09, ACM, New York, NY, USA, 2009, pp. 361–370.
[101] Y. Freund, R. Iyer, R.E. Schapire, Y. Singer, An efficient boosting algorithm for combining preferences, J. Mach. Learn. Res. 4 (2003) 933–969.
[102] X. Zhu, W. Nejdl, M. Georgescu, An adaptive teleportation random walk model for learning social tag relevance, in: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '14, ACM, New York, NY, USA, 2014, pp. 223–232.
[103] C. Wei, W. Hsu, M. L. Lee, A unified framework for recommendations based on quaternary semantic analysis, in: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '11, ACM, New York, NY, USA, 2011, pp. 1023–1032.
[104] M. Sarwat, J. Levandoski, A. Eldawy, M. Mokbel, Lars: an efficient and scalable location-aware recommender system, IEEE Trans. Knowl. Data Eng. 26 (6) (2014) 1384–1399.
[105] Peter Mika, Social Networks and the Semantic Web (Semantic Web and Beyond), Springer-Verlag New York, Inc, Secaucus, NJ, USA, 2007.
[106] N.J. Belkin, Some(what) grand challenges for information retrieval, SIGIR Forum 42 (1) (2008) 47–54.
[107] J. Stoer, R. Bulirsch, Introduction to Numerical Analysis, 3rd edition, Springer, New York, 2002.
[108] K. Caine, R. Hanania, Patients want granular privacy control over health information in electronic medical records, J. Am. Med. Inf. Assoc. 20 (1).