

# Chapter 26

## Integrating Online Social Network Analysis in Personalized Web Search

**Omair Shafiq, Tamer N. Jarada, Panagiotis Karampelas, Reda Alhajj, and Jon G. Rokne**

**Abstract** With the emergence of high speed internet applications and advanced Web 2.0 based Rich Internet Applications (i.e., blogs, wikis, etc.), it has become much easier for the users to publish data over the Web. This brings a challenge for the Web search solutions to let individual users find the right information as per their preferences, because traditional Web search engines have been built on “one size fits for all” concept. Different users of the Web may have different preferences. Search results for the same query raised by different users may differ in priority for individual users. In this book chapter, we present the extended version and results of our proposal on community-aware personalized Web search. It is quite challenging to know the preferences of the users by the search engines. We have designed and developed our unique approach of finding the preferences of users from the relevant parts of the user’s social network and community. We believe that the information related to the queries posed by the users may have strong correlation with the relevant information in their social networks. In order to find out personal

---

O. Shafiq (✉) · J.G. Rokne  
Department of Computer Science, University of Calgary, Alberta, Canada  
e-mail: [moshafiq@ucalgary.ca](mailto:moshafiq@ucalgary.ca); [rokne@ucalgary.ca](mailto:rokne@ucalgary.ca)

T.N. Jarada  
University of Calgary, Calgary, Alberta, Canada  
e-mail: [tjarada@gmail.com](mailto:tjarada@gmail.com)

P. Karampelas  
Department of Information Technology, Hellenic American University, Manchester, NH, USA  
e-mail: [pkarampelas@gmail.com](mailto:pkarampelas@gmail.com)

R. Alhajj  
Department of Computer Science, University of Calgary, Alberta, Canada  
Department of Information Technology, Hellenic American University, Manchester, NH, USA  
Department of Computer Science, Global University, Beirut, Lebanon  
e-mail: [alhajj@ucalgary.ca](mailto:alhajj@ucalgary.ca)

interest and social-context, we find (1) activities of users in their social-network, and (2) relevant information from user's social networks, based on our proposed trust and relevance matrices. We have further developed a mechanism that extracts from user's social network information to be used to re-rank search results from a search engine. We also have discussed the implementation and evaluation details of our proposed solution.

## 26.1 Introduction

The amount of information available over the Web is increasing rapidly. Through the use of high-speed internet, high capacity network, and upcoming advanced interactive websites, like facebook, YouTube and blogging, it has become even easier for the internet users to publish data over the Web. With this information flooding, it becomes hard for a user to find out right information over the Web. Search engines like Google, Altavista, Yahoo, Bing, etc. bring thousands of search results for a particular search query. It is almost impossible and mostly frustrating for a human user to go through the content of all the search results manually. With the increase of information over the Web, scalability of Web search engines became important, e.g., [13]. Search engines, like Google, even provide manual mechanisms to rank up and down search results. But for each and every search result, it is not possible for a user to adjust the ranking manually in advance; as a result categorization and personalization of search results become an important issue [24]. Different efforts are being made in order to enable community and social network aware [1], user-terminology-aware [2] and group-aware [4,16] search engines that may help in bringing results closer to the user's interest. After providing the personalized version [9] of Web search by Google, researchers still seem to feel the need to move towards social network-aware search [17]. Not only Web search, but the personalized search has been explored in other real-life scenarios, e.g., medical and digital libraries [15].

Our goal is to find out a way to prioritize search results based on the interests, activities and community of users. There are different ways that exist in order to find out contextual information, i.e., location etc. But this information is most of the time not enough or limited to find out the right search results for the right user at the right time. Different users searching over the Web for the same thing may have different preferences. An executive going for a business meeting would probably be interested in a hotel closer to the meeting place; however a back-packer searches for a hotel in the same city might want to search for, rather cheaper hotel regardless of the location.

We propose to exploit user's social network in order to find out the interests, activities of the users and their community, i.e., friends in the social network. This information can help in bringing word-of-mouth recommendation system from trusted friends in the social network. Every group of users may have certain set of activities over the social network, i.e. postings, public profile information, tagging,

blogging, etc. Similarly different friends of the user may also have similar activities. Most of the times, friends of a user in a social network have to some extent similar interests and therefore could help in finding out the right information the user needs. The information about the related search query from friends in a social network, or activities of a user in social network could help in finding out the relevant information over the Web and prioritize it for the user. It could help the user to find out the right information at the right time. Such kind of community-aware personalized Web search utility could be used in many different ways and in many applications involving Web Intelligence and Business Intelligence, i.e., for summarizing important search results, caching of important search results, as well as collaboration techniques.

We have developed methodology to find out activities of users over a social network as well as their friends/community. Secondly, we find out how to use this information to prioritize relevant information over the Web. This way, we are able to use social networks for finding relevant information over the Web in a personalized way. There are different kinds of friends and communities in a social network based on which we can find out the relevant information. Same as in real-life, a user can have different trust matrices for different friends in a social network. This means, information collected from a more trusted friend is more important than that of the one from lesser trusted one. As a part of this research, we have found out metrics based on which the trust can be built and used in a social network. Similarly, different metrics to rank the results over the Web have also been sought.

The rest of the chapter is organized as follows. Section 26.2 provides an overview of the related work together with the pros and cons. Section 26.3 describes our proposed solution of community-aware personalized Web search and discusses how to model the information based on user's individual preferences as well as trusted and relevant friends of the users in their social network. Section 26.4 describes the details implementation and evaluation, followed by conclusion.

## 26.2 Related Work

Various approaches have been developed for personalized Web search based on contextual information, i.e., interests and preferences of users [20, 23]. A number of approaches are based on modeling and collecting contextual information. For example, context-aware search methods [18] use special ontology that is constructed from the table of contents of the book, to help formulating the query for efficiency and to refine query search results that are relevant to the user's interests. Location-awareness is another form of context-awareness that is used to filter out search results based on the current location of the users. There are different ways in which clients access the services over the Internet. With the advancements of requirements and complexity in consumer applications, consumers expect the services to be aware of their current environment and situation, i.e., the type of device through

which consumers are communicating, their preferences, locations, etc. The context information can be used by services to provide their clients with a customized and personalized behavior. Chen et al. [6] describe an ontology for supporting pervasive context-aware systems. An ontology is developed as a part of the Context Broker Architecture (CoBrA), a broker-centric agent architecture that provides knowledge sharing, context reasoning, and privacy protection supports for pervasive context-aware systems.

Other approaches take context into account one way or the other. For instance, Almeida and Almeida [1] propose an algorithm for search results ranking based on community-based evidence; they call it query contextualization which is obtained from user's interactions. First, the algorithm identifies interest-based community using session interest graphs followed by community identification in the graphs. Based on community identification, search results are ranked. This way, community information is used as a way to provide context for the query specified by the user. Authors claim to improve the search results up to 48 %.

Computing word-of-mouth recommendations is another form of context-awareness that has been introduced recently. The idea is to identify who knows what in the social-network of a user, and based on that some information is brought up. Not only this, the process also includes finding out who is trustworthy source of information. For example, Granovetter [10] proposes how social networks can serve as a source of new information to which an individual may not otherwise have access and Sugiyama et al. [21] proposes adaptive Web search based on automatically created users profiles. This is especially useful in use-cases like job-hunting where weak social ties might be better than strong social ties to bring some new and interesting information. Heath [11] proposes algorithms to compute trust metrics in his developed system based on social-network. The information is collected from tags used in the system to comment on and recommend some information. Another algorithm helps in calculating the prevalence of an individual in the reviews of items that have been tagged with a particular tag, thereby providing a relative measure of their experience with the topic. Calculation of affinity score between one individual and another is based on the analysis of reviews where the information is provided to the algorithm in the form of FOAF (Friend of a friend) description.

With the evolution of Semantic Web [3] over the last decade, several approaches for personalized Web search from social networks based on Semantic Web have also been initiated. For example, Carminati et al. [5] propose the usage of tags as a basis to enforce advanced personalization policies for Web based search. A multi-layer framework has been proposed where data collected by social tagging communities is complemented with additional services that provide users the ability of expressing their agreement or disagreement with existing tags, denoting the members that they trust based on their characteristics and relationships, or specifying policies enforcing the criteria adopted by a given end user to access the resources based on the associated tags and ratings. The framework helps in representing the tagging information as well as the user policies based on Semantic Web standards so that its expressivity could be exploited easily. The layers include data layer which consists

of services in-charge of managing social network data, Web metadata as well as ratings by the users. Second is Rule layer that consists of modeling user-preferences and trust policies specified by the users. This is the key of the information that helps in identifying contextual information based on social network. Last is Application Layer where the output of data and rule layers is used by a set of agents corresponding to the application layer, for a variety of purposes, i.e., social network data represented in Semantic Web is used by semantic search engines in order to find users and resources matching given queries.

Zeng et al. [25] first introduce user interest models which are then used to build a social network to track the group interest related to the given user. The authors then refine the search task based on the semantic dataset using the acquired group interest. The search refinement is carried out based on the interest of the individual user, as well as interest of the group which the user is part of.

After covering the related approaches, in the next section, we present our methodologies for contextualizing users' interest based on the information from their social network and use it to refine Web search results. The novelty of our approach is that our methodologies are generic enough and can be customized for any social network and any Web search engine. Secondly, the proposed methodologies prescribe the usage of Semantic Web technologies to make the modeling of user preferences more expressive that could be used during the search refinement process.

The key problems observed in the related works as described above could be articulated as follows. Most of the related works consider location information as the only important part of context [6]. However, in case of user-based customized Web search, location is not the only important factor that should be taken into account. Therefore, most of the related work approaches have not considered user context to its full potential. Secondly, several user based context calculation techniques are computation exhaustive. For example, the work described in [1] proposes to generate interest graphs and then tries to identify community, which makes the whole approach highly computational extensive, and hence causes high resource consumption and hence not cost effective with respect to time. Further, most of the approaches are limited to the interest represented in the form of keywords only. Considering keyword based approaches, this is a quite straight forward, simple and cost effective method. However, it imposes a limitation because complex information about user's preferences cannot be modeled using simple keywords; therefore, we have taken into account the modeling of complex preferences of users using semantic expressions.

Most of the approaches blindly take into account the information from social-network of a user, regardless of its relevance. There could be some information in the social network which might be relevant from syntactic point-of-view, however, might be quite different from semantic point-of-view. Therefore, most of the approaches are unable to identify relevant and irrelevant information from the social network of a user. We have introduced the use of relevance metrics to take into account only information which is relevant to users' interests.

## 26.3 Community Aware Personalized Web Search

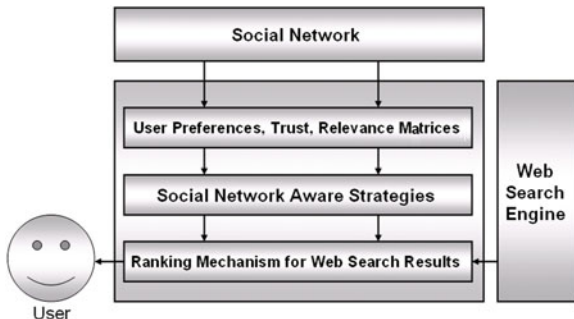
Currently available Web search solutions treat users mostly the same, in a “one size fits all” manner. Web search engines give same search results for the same queries by different users. However, in reality it is very likely that different users may have different preferences. Therefore, we believe that the search mechanisms need to be aware of the contextual information, preferences and interests of the users. In this section, we describe our proposed solution to tackle this problem.

Figure 26.1 depicts our solution in a layered manner that acts as a middleware between the user and the Web search engine. The first layer takes the information from the social network of users, and finds out the activities of the users in their social network, e.g., what kind of postings a user is doing, tagging, and keywords from user’s own profile. It helps us in knowing the current state of personal interests of the user him/herself. Secondly, we have built a similar mechanism to find out the activities and information posted by the community or friends of the user in the social network; it can be tagging, posting a video, blogging and information from friends’ public profile, etc. This may help us in finding out what kind of information and interest does members of the community of a user have. We believe that the information collected from the community of the user may be of direct interest to the user, and hence may serve as preference of the user. However, there may be several cases where some part of the community is not relevant for some information for the user (e.g., a user searching for computer science material might not find as relevant information from his/her friends belonging to the medical field). In order to tackle this, we have developed trust and relevance schemes to find out the relevant information from the user’s social network. This mechanism allows users to specify trust relationship with their friends or community in the social network. For example, there can be some friends who are most close or trustworthy, from whom the user might want to use more information whereas some of the friends may be of lesser trust, i.e., not close enough, from whom a user might not want to consider the information, similar to what happens in real-life.

We allow users to assign weights, which describe the level of trust, to the friends and other related communities in a social network. This helps in finding out what information from the community is more trustworthy than others. The trust mechanism allows to pullout trustworthy information from user’s social network, which is then used to re-rank Web search results. Similarly, users are also provided with mechanisms to specify the level of relevance of their friends in their social network, with respect to different search topics. This allows the personalized mechanism to consider only relevant subset of the user’s social network, and hence avoid irrelevant information.

Our approach is to build our proposed solution in a generic manner, such that it is not dependent on any particular social-network or Web search engine. Our solution can be customized for any search scenario, and hence can be applied on any kind of known social networks (e.g., facebook, youtube, blogs, etc.) as well as any kind of Web search engine (i.e., Google, Yahoo, Alta Vista, etc.). Our Social-Network

**Fig. 26.1** Layered model for community-aware personalized Web search



Analysis based contextual mechanism does not require any change or interference in the Web search engine because it acts as a search-engine front-end for users searching information over the Web; it helps them in ranking and reordering of Web search results according to their preferences obtained from their social-network.

Figure 26.1 provides an overview of our proposed solution in the form of layered model architecture. Each user and his/her friends have their activities on the social network from which the preferences and the trust are identified. Based on the trust and relevance schemes, preferences of a user and his/her community from the social-network are obtained. Certain standard social-network analysis based techniques are used to select relevant subset preferences from the user’s community. The Social Network Aware Strategies are applied to Web search data with the information coming from the user’s social network, and Web search results are ranked accordingly. While ranking the search results, our solution does not hide or change any results from the user. It just re-ranks and re-orders Web search results for the user based on his/her social-network information, which might be interesting for the user. We define the social network, in a generic way, as follows:

**Definition 1 (Social Network SN).** Let SN be a social network that shows connectivity between different members in the social network. It can be represented as a two-dimensional matrix of  $n$  rows and  $n$  columns.

$$Social\ Network := SN(n, n)$$

where, SN is the two-dimensional matrix. Each of the entries in SN is referred to as,  $SN_{i,j}$ , where  $1 \leq i \leq n$  and  $1 \leq j \leq n$ . The value of each entry  $SN_{i,j}$  in the matrix describes the connectivity between two particular members of the social network, namely  $i$  and  $j$ . Null values refer to no connectivity, whereas, positive values indicate the existence of some connectivity between the members.

In the subsections below, the sub-modules of the community-aware personalized Web search model shown in Fig. 26.1 are described in detail.

### 26.3.1 User Preferences and Trust Model

In order to know user's preferences in a social network, activities of users and their friends in the social-networks are to be modeled in a systematic way. This information could be used for the purpose of ranking and re-ordering Web search results. The design methodology behind our solution is generic. Therefore, our solution is not restricted or dependent on any particular type of social network. In any social network, a user has certain activities which can be postings of certain information, doing some activities, attending some events, tagging and commenting on photos of the users, etc. The same holds for the friends of the considered user in the social-network; they are expected to have similar kind of activities. We have adopted a simpler way to model user preferences by having this information as key-pair values. This is a practical approach where all the activities of a user are summarized in the form of a set of keywords. For example, if a user is blogging on the topic of soccer in his/her social network about the upcoming FIFA World Cup in 2010 in South-Africa, in that case a simple set of keywords mentioning interests of the user may be listed as {FIFA, World Cup, soccer, 2010, South Africa}. Similar sets of keywords are collected for each of the users performing their activities on the social network.

#### 26.3.1.1 User Preferences Metric

A sample list of keywords of interest for users is given in Table 26.1 that shows the id of the users against the important keywords collected based on their activities in the social network.

Based on the activities of the users, the keywords provided in Table 26.1 represent their interests. These keywords, once collected, can be taken into account for ranking Web search results for the users. Similarly, keywords of other users (which may be friends of each other) could be taken into account while ranking based on the interest of a group of people.

**Definition 2 (Preference List).** Let PL be the list of preferences of a member in the social network, which may also be a user searching for information over the Web. PL can be represented as tuple:

$$PL (index, keywords (m))$$

where, index is a numeric item that corresponds to the user ID in the social network SN, and keywords is a one dimensional array of keywords, having each item  $keywords_i$  as preference of the user, with length  $1 \leq i \leq m$ .

#### 26.3.1.2 User Trust Matrix

Each user in a social network has his/her friends. However, as in real-life, there are different ties with different friends. These ties may also vary with time. Close



**Table 26.1** List of keyword sets for users in a social network

| User ID | Set of keywords                      |
|---------|--------------------------------------|
| 1       | FIFA, World Cup, soccer, 2010        |
| 2       | Cricket, World Cup, Australia, 1992  |
| 3       | Politics, North America              |
| 4       | Programming, Java, Web, Enterprise   |
| 5       | Traveling, Airline tickets, holidays |
| 6       | Electronics, Boxing day, sales       |
| 7       | Movies, Hollywood, Bollywood, Avatar |

friends or important friends normally have closer ties and hence higher trust factor than that of others. Our approach takes into consideration the social ties and trust of users with their friends in the social network. This information is necessary and is required in ranking the information. For some of the searches, friends of a user with closer ties and high trust metric might be of more importance than others, e.g., while buying a computer. However, there might be cases where weak social ties are more important for a user than stronger social ties, e.g., in case of job search. Our approach is to represent the social ties based on trust in a numerical scale from 0 to 10, ranging from 0 as the no/weakest ties to 10 as the strongest ties. Such information is calculated for each user with each of his/her friends in the form of a matrix; it is taken into account while performing social network aware personalized Web search ranking.

**Definition 3 (Trust Matrix TM).** Let TM be the social network (as described in Definition 1) that shows connectivity between different members in the social network. The value of each entry  $TM_{i,j}$  in the matrix describes the level of trust between the two connected members.

$$TM (n, n)$$

As described in Definition 1, null values mean no connectivity; whereas, positive values describe connectivity between members, with its value ranging from 1 to 10 as the level of trust (i.e., 1 as for lower level of trust and 10 as for higher level of trust). We describe standard representation of a social network as follows:

1. Vertex  $v$  represents a user in social network  $n$
2. Edge  $e$  represent connection between two particular users in social network  $n$
3. Weight  $w$  of each edge represents the level of trust between the two connected users in social network  $n$
4. Preferences  $p$  denote the list of preferences/interest values or logical expressions for each of the users in the social network

Table 26.2 shows a one-mode adjacency matrix of a sample social network. Rows and columns show users A to G and each of the cells shows the weight of the connecting edge, i.e., the strength of trust relationship between the users.

**Table 26.2** TM – Trust matrix for a simple social network

|   | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| A | X | 9 | 9 | 9 | 4 | 2 | 3 |
| B | 2 | X | 7 | 6 | 6 | 5 | 1 |
| C | 6 | 2 | X | 4 | 9 | 6 | 8 |
| D | 4 | 5 | 3 | X | 8 | 1 | 4 |
| E | 0 | 8 | 4 | 4 | X | 7 | 6 |
| F | 8 | 5 | 7 | 2 | 2 | X | 2 |
| G | 2 | 4 | 6 | 5 | 5 | 2 | X |

The diagonal of the matrix shows the highest weight as we assume that a user may have his/her own interests at highest priority.

### 26.3.1.3 User Relevance Matrix with Respect to Topics

As described in the previous section about user trust matrix, where the level of trust between different users in a social network is quantified with numbers, in a similar way, we represent the relevance of user or user-groups relationship with each other, with respect of a particular topic, in a quantifiable manner. For example, the user is interested in sports (particularly football), however, there are couple of friends in the user’s social network relevant to football as well as cricket. Now, in order to perform group-level preference calculation, it will be important to include the information obtained from friends relevant to football, rather than cricket. In order to represent this level of relevance, we propose the use of relevance matrix.

**Definition 4 (Relevance Matrix RM).** Let RM be the relevance matrix that shows the level of relevance between different members in a social network, with respect to a particular topic.

$$RM (n, n, topic)$$

where, RM is the two-dimensional matrix,. Each of the entries in RM is referred to as,  $RM_{i,j}$ , where  $1 \leq i \leq n$  and  $1 \leq j \leq n$ . The value of each entry  $RM_{i,j}$  in the matrix describes the level of relevance between members i and j, with respect to a particular topic, with the value ranging from 1 to 10 (i.e. 1 as for lower level of relevance and 10 as for higher level of relevance).

Tables 26.3a and b shows two relevance matrices that show relationships between various users in a social network, based on different topics.

After representing the level of relevance as well as the trust between the different users, we calculate balance between the trust and relevance matrices, and hence identify the most relevant and trust friends of a user in his/her social network to find the list of preferences. The balance between trust and relevance is calculated using the expression given as follows:

**Table 26.3** RM – Relevance matrix of users based on topics ‘Football’ and ‘Cricket’

|                     | A | B | C | D | E | F | G |
|---------------------|---|---|---|---|---|---|---|
| (a) Topic: Football |   |   |   |   |   |   |   |
| A                   | X | 5 | 7 | 2 | 5 | 3 | 2 |
| B                   | 5 | X | 6 | 8 | 2 | 6 | 6 |
| C                   | 3 | 2 | X | 5 | 6 | 5 | 3 |
| D                   | 8 | 5 | 3 | X | 2 | 4 | 1 |
| E                   | 0 | 8 | 4 | 4 | X | 2 | 2 |
| F                   | 1 | 5 | 7 | 2 | 2 | X | 4 |
| G                   | 2 | 4 | 6 | 5 | 5 | 2 | X |
| (b) Topic: Cricket  |   |   |   |   |   |   |   |
| A                   | X | 7 | 3 | 5 | 8 | 9 | 2 |
| B                   | 5 | X | 7 | 6 | 4 | 6 | 6 |
| C                   | 4 | 2 | X | 4 | 7 | 5 | 5 |
| D                   | 7 | 5 | 3 | X | 2 | 3 | 3 |
| E                   | 3 | 8 | 4 | 4 | X | 2 | 7 |
| F                   | 4 | 5 | 7 | 2 | 2 | X | 6 |
| G                   | 2 | 4 | 6 | 5 | 5 | 2 | X |

**Table 26.4** Calculated measures based on relevance and trust

|   | A         | B         | C         | D         | E         | F         | G         |
|---|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| A | X         | 0         | 0         | 0         | 0         | 6         | 4         |
| B | 10        | X         | <b>42</b> | <b>48</b> | 12        | <b>30</b> | <b>36</b> |
| C | 18        | 4         | X         | 20        | <b>54</b> | <b>30</b> | 9         |
| D | <b>32</b> | 25        | 9         | X         | 16        | 4         | 1         |
| E | 0         | <b>64</b> | 16        | 16        | X         | 14        | 4         |
| F | 0         | 25        | <b>49</b> | 4         | 4         | X         | 16        |
| G | 4         | 16        | <b>36</b> | 25        | 25        | 4         | X         |

$$\alpha TM_{i,j} + \beta RM_{i,j} = P_{i,j} \tag{26.1}$$

where TM is the trust matrix and RM is the relevance matrix. Moreover,  $\alpha$  and  $\beta$  have been introduced as two threshold measures, which can be fine tuned by the user to reflect whether he/she gives more preference to trust, relevance or both.

Table 26.4 shows the calculated measures between different users in a social network, based on the relevance as well as trust. Once the balance of trust and relevance between a user and his/her friends is calculated, most trustworthy as well as highly relevant part of the social network is selected to deduce preference information from the relevant part of user’s social network. This way, we manage to focus on the key and required part of the social network of the user, rather than performing calculation for whole network, and hence avoid non-trusted and irrelevant information.

### 26.3.2 Social Network Analysis Strategies

Once the important part of the social network (with higher relevance and trust) is selected, social network analysis (SNA) strategies [8] constitute the next important step of our proposed solution. They further help in narrowing-down the required members of the user’s social network, for whom the preferences are to be collected in order to re-rank Web search results. SNA measures help in analyzing the social network in terms of network and graph theory consisting of nodes and ties. Nodes are individual actors in social networks. For example, betweenness centrality of a node calculates the extent to which the node lies between other nodes in the network. This measure takes into account the connectivity of the node’s neighbors, giving a higher value for nodes which bridge clusters. An edge that is said to be a bridge, if deleted, would cause its endpoints to lie in different components of the social network. Closeness measure shows the degree to which an individual is near all other individuals in a network (directly or indirectly). Degree measure counts the number of direct ties to other actors in the network. This is also known as the “geodesic distance”. We have used such SNA measures to identify for a given user his/her friends whose preferences and interests can help in better ranking Web search results. Below we list a couple of general purpose strategies which could be effective by employing SNA measures:

**Strategy 1:** A user may want to get preferences from his/her friends to rank Web search results.

**Technique:** Preferences of friends with higher value of *betweenness* centrality should be taken into account while ranking Web search results. List of preferences would be calculated by considering Eq. (26.1) and Definition 2 as:

From Eq. (26.1) ...  $\alpha TM_{i,j} + \beta RM_{i,j} = P_{i,j}$  and based on Definition 2 ...  $PL(\text{index, keywords}(m))$

*Betweenness Centrality is calculated as*

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where  $\sigma_{st}$  is the number of shortest paths from  $s$  to  $t$ , and  $\sigma_{st}(v)$  is the number of shortest paths from  $s$  to  $t$  that pass through vertex  $v$ .

$$PL(\text{Max}(C_b(P)))$$

where  $P$  is the sub-part of the social network that was identified after calculating members with higher relevance and trust. Preference list of members with maximum betweenness centrality in the identified  $P$  sub-social network would be considered for re-ranking Web search results.

**Strategy 2:** A user may want to get preferences from his/her friends having higher degree of dependence within the social network; this is to be used in ranking Web search results.

**Technique:** Preferences of the friends connected with *Bridge edge* should be taken into account while ranking Web search results.

*The bridge measure is computed as  $B(SN)$ ; its deletion increases the number of connected components within the social network  $SN$ . The algorithm follows from Tarjan [22].*

$$PL((B(P)))$$

*where  $P$  is the sub-part of the social network that was identified after calculating members with higher relevance and trust. Preference list of members in the Bridge vertex in the identified  $P$  sub-social network would be considered for re-ranking Web search results.*

**Strategy 3:** A user may want to get preferences from his/her friends, who are found closely related in the social network, to rank Web search results.

**Technique:** Preferences of the friends with higher *Closeness* measure should be taken into account while ranking Web search results.

*Closeness is calculated as*

$$C_C(v) = \sum_{s \neq v \neq t \in V} \frac{1}{\sum_{t \in V/v} d_G(v, t)}$$

*which is the reciprocal of the sum of geodesic distances to all other vertices in the social-network.*

$$PL(\text{Max}(C_c(P)))$$

*where  $P$  is the sub-part of the social network that was identified after calculating members with higher relevance and trust. Preference list of members with the highest Closeness, in the identified  $P$  sub-social network, would be considered for re-ranking Web search results.*

**Strategy 4:** A user may want to get preferences from his/her friends, who are found receiving higher degree of connections in the social network, to rank Web search results.

**Technique:** Preferences of the friends, with higher *In-degree* measure, should be taken into account while ranking Web search results.

*In-degree is calculated as  $\text{indeg}^-(v) =$*

$$\sum_{v \in V} \text{deg}^-(v)$$

$$PL(\text{Max}(\text{indeg}^-(P)))$$

where  $P$  is the sub-part of the social network that was identified after calculating members with higher relevance and trust. Preference list of members with the highest in-degree value, in the identified  $P$  sub-social network, would be considered for re-ranking Web search results.

**Strategy 5:** A user may want to get preferences from his/her friends who are found actively communicating with other friends in the social network, to rank Web search results.

**Technique:** Preferences of the friends, with higher *Out-degree* measure should be taken into account while ranking Web search results.

*Out-degree* is calculated as  $\text{outdeg} + (v) =$

$$\sum_{v \in V} \text{deg}^+(v)$$

$$PL(\text{Max}(\text{outdeg} + (P)))$$

where  $P$  is the sub-part of the social network that was identified after calculating members with higher relevance and trust. Preference list of members with the highest out-degree, in the identified  $P$  sub-social network, would be considered for re-ranking Web search results.

**Strategy 6:** A user may want to get preferences from his/her close friends to rank Web search results.

**Technique:** Preferences of the friends in the social network that are directly connected to the user should be taken into account while ranking Web search results.

*Directly connected friends  $D$  are the ones at distance  $d=0$*

$$PL(\text{Max}(D(d = 0)(P)))$$

where  $P$  is the sub-part of the social network that was identified after calculating members with higher relevance and trust. Preference list of members with the friends connected at distance  $d = 0$ , in the identified  $P$  sub-social network, would be considered for re-ranking Web search results.

**Strategy 7:** A user may want to have recommendation for ranking search results based only on his/her own interests.

**Technique:** Only preferences of the user him/herself should be taken into account while ranking Web search results.

*PL (user) shows the Preference list of the user him/herself only, and does not take into account the social-network of the user.*

**Strategy 8:** A user may want to have recommendations for ranking search results based on the interest of closely connected friend in his/her community.

**Technique:** Preferences of the friends forming a *Clique*, should be taken into account while ranking Web search results.

*Clique is calculated as  $Cl = \text{subset of at least three members in the social-network, such that every two members in the subset are directly connected.}$*

$$PL(Cl(P))$$

*where  $P$  is the sub-part of the social network that was identified after calculating members with higher relevance and trust. Preference lists of members who are part of the clique, in the identified  $P$  sub-social network, would be considered for re-ranking Web search results.*

The above mentioned strategies use standard SNA measures to help the user getting preference list from the members of his/her social network in different possible ways, while performing Web search. These measures are generic and can be applied to any kind of social network and any Web search engine. These strategies help the user to know the preferences as a set of recommendations that are taken into account while performing re-ranking of Web search results.

### 26.3.3 Ranking Mechanism for Web Search

In this section, we provide the details of our next key component of the social-network aware personalized Web search solution. It takes into account the preferences list which is retrieved as described in the previous sub-sections, uses the preferences to re-rank Web search results. This way, Web search results are re-ordered, based on the information obtained from the social network of a user, to help in displaying more relevant and interesting Web search results at the top of the retrieved list.

Ranking of information is carried out on individual as well as group basis. In case of individual basis, the user performing the search is the point of reference and all the calculations from the social network measures are carried out accordingly. However, in case of group-based ranking, the calculations from the social network measures are carried out with reference to all users taking part in the group.

#### 26.3.3.1 Individual Ranking

Ranking of Web search results based on the preferences of an individual user is carried out while taking the user as the point of reference and all the calculations from the social network measures are carried out accordingly. Preferences of the users, after performing the calculations from the social network measures, are taken

---

**Algorithm 1** Individual Ranking of Web search results
 

---

**Input:** User Query Q, Social-Network SN represented as the adjacency matrix

**Output:** List of re-ranked Web Search results

**begin**

1.  $TM$  ( $SN$ ) is the adjacency matrix of the user's social network representing the trust level as described in the previous sub-sections
2.  $RM$  ( $SN, topic$ ) is the adjacency matrix of the user's social network representing the level of relevance with respect to a topic, as described in the previous sub-sections
3.  $P = \alpha TM + \beta RM$  is the subpart of the social-network, as specified in Eq. (26.1)
4.  $PL = (index, keywords(m)) P$ , where PL is the list of preferences from the members of the sub social-network P
5. List  $R' = Re-ordered(R, PL)$ , where  $R'$  is the re-ordered list of Web search results based on the list of preferences PL

**end**

---

into account while ranking Web search results. Listing 1 is proposed for ranking Web search results based on individual ranking.

Listing 1. Algorithm for individual ranking of Web search results Source Codewww

Algorithm 1 is described below step-wise.

- Step 1:** Calculate the sub-social network based on higher level of trust and relevance matrices as P
- Step 2:** Calculate the list of members from the sub-social network P from the social network analysis strategy selected by the user
- Step 3:** Retrieve the preferences/interest keywords of all the users identified by the calculation in Step 1 and Step 2
- Step 4:** Take top 100 search results from the search engine based on the search query provided by the user
- Step 5:** Match the list of preferences/interest keywords obtained in Step 3 with the Web search results
- Step 6:** Re-order the Web search results in descending order to show the most relevant Web search results on the top, with respect to the selected social-network

There might be the possibility of having thousands of Web search results for most Web search queries; therefore, matching the list of preferences with each of the returned Web search results is not feasible. Hence, we propose to perform the matching of the preference/interest keywords for the top 100 Web search results.

### 26.3.3.2 Group Based Ranking

Ranking Web search results based on a certain group in the social network is carried out while taking into account all users who are part of the group. Calculations for



**Algorithm 2** Group Ranking of Web search results**Input:** User Query Q, Social-Network SN, represented by the adjacency matrix**Output:** List of re-ranked Web Search results**begin**

1.  $SN(n, e)$  is the network with  $n$  vertices and  $e$  edges
2.  $SN'$  = Social Network SN with vertices merged into one for the members identified in a group
3.  $TM(SN')$  is the adjacency matrix of the user's social network representing the trust level among members, as described in the previous sub-sections
4.  $RM(SN', topic)$  is the adjacency matrix of the user's social network representing the level of relevance among members, with respect to a topic, as described in the previous sub-sections
5.  $P = \alpha TM + \beta RM$ , is the sub part of the social-network, as described in Eq. (26.1)
6.  $PL = (index, keywords(m)) P$ , where PL is the list of preferences from the members of the sub social-network P
7. *List*  $R' = Re\text{-ordered}(R, PL)$ , where  $R'$  is the re-ordered list of Web search results based on the list of preferences PL

**end**

the social network strategies are done for each of the user in the group. Then for a particular social network strategy, preferences of the users identified by the group in the social network are taken into account while ranking Web search results. Listing 2 is proposed for ranking Web search results based on group ranking.

The key difference between the two algorithms for individual ranking and group-based ranking is the additional step in the group-based ranking algorithm, where the calculations are carried out for all users taking part in the group in the social network.

Listing 2. Algorithm for group ranking of Web search results

The algorithm is described below step-by-step.

- Step 1:** Retrieve the social network of the members belonging to a particular group.
- Step 2:** Merge vertices representing the group as one node in the social network, and calculate the sub social-network based on higher level of trust and relevance matrices as P
- Step 3:** Calculate the list of members from the sub social-network P derived by applying selected the social network analysis strategy, with respect to the group
- Step 4:** Retrieve the preferences/interest keywords of all the users identified by the calculation in Step 3
- Step 5:** Take the top 100 search results from the search engine based on the search query provided by the user
- Step 6:** Match the list of preferences/interest keywords obtained in Step 4 with the Web search results
- Step 7:** Re-order the Web search results in descending order to show, with respect to the selected social-network, the most relevant Web search results on top

### 26.3.3.3 Fractional Cascading for Ranking Search Results

After calculating the preferences based on personal and group level information from the user's social network, the final step is to rank the relevant Web search results based on the best matches with the preferences obtained. However, looking at the huge size of Web search results, this task can become very cumbersome and time-consuming to do (i.e., if we start matching each Web search result with each of the preferences and then calculate and rank them in the order of best match). Especially if the list of preferences is higher, the ranking process can take a significantly large amount of time, in comparison to the time in which the results are obtained from the Web search engines. Therefore, we employ the techniques of orthogonal range search to perform the ranking of Web search results for ordering the best match.

Orthogonal range searching formalizes the search based on multiple dimensions (from one to  $n$  dimensions). We take each preference item as one dimension. So, if we have three items in the calculated preference list, then we have a three-dimensional match/ranking mechanism. Similarly, if there are 4, 5 or  $n$  items in the preference list, the dimensions will be 4, 5 or  $n$ , respectively. In order to keep our solution as generic as possible, we have chosen to use a dynamic data structure for matching the search results based on  $n$  dimensions. In this case, we use fractional cascading mechanism, which takes into account each dimension, and calculates the match with the first dimension; if a good match is found with the first dimension, only then proceeds to the other dimensions, otherwise it drops the calculation for the rest of the dimensions (i.e., preference items) for the particular search result, and then proceeds to the next search result accordingly. In this case, we save the time, by calculating the necessary dimension match only, and drop the search results where no good match is found with any of the dimensions.

The Listing 3 sketches the fractional cascading based algorithm for ranking Web search results based on the preferences mentioned as dimensions. It calculates the match of each Web search result  $R$  with each dimension  $D$ . As soon as a Web search result  $R$  does not match any of the dimensions, the algorithm stops calculating the further matches for that Web search result and continues to the next search result; hence it saves time by calculating only for the required dimensions. Once dimension match check is performed for each Web search result  $D$ , the counter array items having count equal to the number of dimensions is the answer, i.e., it ranks those Web search results as the highest ones; they matches with every dimension, and vice versa. Web search results are then re-ordered based on the ranking obtained from this mechanism.

#### Listing 3: Fractional-cascading based Ranking algorithm

As already described by fractional cascading technique [19], cost estimation of the algorithm is  $O(k) + n$ , where  $n$  is the time spent in matching dimension  $D$  with a Web search result.

---

**Algorithm 3** Fractional cascading for ranking search results

---

```
Counter [1 . . . D]  $\downarrow$  0
For each Web search result R
For each Preference as dimension D
if (D matches R)
counter[D] ++
else
break and continue to next Web search result
```

---

## Evaluation and Discussion

Ranking algorithms are the ones that involve maximum processing in re-ordering the Web search results based on the preferences of the user's community. As described in Chap. 3, the complexity and cost estimation of our ranking algorithm, as being based on fractional cascading technique [19], is  $O(k) + n$ , where  $n$  is the time spent in matching dimension  $D$  with a Web search result. The increase in the cost of the algorithm, with increasing dimension is linear, therefore, the approach is sustainable, as the list of preferences increases or extends. Moreover, we have limited the re-ordering of only top 100 results, rather than attempting to re-order all of the search results. This is because, after analyzing click-through and click-stream datasets, we found that, although many of the Web search engines provide millions of the search results. However, users mostly visit the top range of the Web search results, and do not normally visit the results going beyond certain initial range, normally in terms of hundreds. Therefore, re-ordering of the top 100 search results is beneficial for the users.

From implementation point-of-view, we have evaluated our solution using a real-life data set of social network, based on wikipedia where different users posted information, and compared the personalized Web search results with the original search results in many different ways. One way to evaluate was the comparison of ranked search results with original search results from a search engine, through user feedback, i.e. by first allowing the users to build a sample social network of web blogs through a wiki, and then using the community information collected from wiki pages, to rank the search results. We assumed that different users posting on same wiki page, were taken as connected to each other as "friends" in the social network.

For this purpose, we used the history of wiki page updates, to find out which users have been posting and sharing information of different wiki pages. In order to evaluate the usability of personalized search results, each participant was required to issue a certain number of test queries and determine whether each result is relevant to his/her interest or not. Users compared the original search results from a search engine versus ranked search results, and rated the usefulness of the personalized Web search results in different aspects. The important observation was that, not only users found a significant number of re-ranked Web search results more relevant, and also did not find our results irrelevant as compared to the original search results.

|  |
|--|
| <b>Search Query:</b> Football  |
| <b>Original Search Result titles:</b>  |
| <ol style="list-style-type: none"> <li>1. NFL.com – Official Site of the National Football League</li> <li>2. BBC SPORT   Football</li> <li>3. FIFA.com – Fédération Internationale de Football Association (FIFA)</li> <li>4. CBC.ca Meyer’s all-in approach a theme among college football coaches</li> <li>5. Official Site of the Premier League – Barclays Premier League News</li> <li>6. The Official Website of Arsenal Football Club   Arsenal.com</li> <li>7. AFL – The official site of the Australian Football League – AFL.com.au</li> <li>8. Manchester United Football Club</li> <li>9. American football – Wikipedia, the free encyclopedia</li> <li>10. Sky Sports   Football News</li> <li>11. Chelsea Football Club – Official Site for News, Tickets &amp; Fixtures</li> </ol> |

|  |
|--|
| <b>Search Query:</b> Football  |
| <b>Personal Preferences:</b> FIFA, American Football, Chelsea  |
| <b>Re-ordered Search Results:</b>  |
| <ol style="list-style-type: none"> <li>1. FIFA.com – Fédération Internationale de Football Association (FIFA)</li> <li>2. American football – Wikipedia, the free encyclopedia</li> <li>3. Chelsea Football Club – Official Site for News, Tickets &amp; Fixtures</li> <li>4. NFL.com - Official Site of the National Football League</li> <li>5. BBC SPORT   Football</li> <li>6. CBC.ca Meyer’s all-in approach a theme among college football coaches</li> <li>7. Official Site of the Premier League – Barclays Premier League News</li> <li>8. The Official Website of Arsenal Football Club   Arsenal.com</li> <li>9. AFL – The official site of the Australian Football League – AFL.com.au</li> <li>10. Manchester United Football Club</li> <li>11. Sky Sports   Football News</li> </ol> |

This is because, our approach does not hide any search results, but only shows the same search results, but in a re-ordered fashion. Therefore, the risk of bringing any irrelevant search results goes almost to none. The advantage of this evaluation approach, based on user feedback, was that the relevance of the search results was explicitly judged by the participants, and hence, there was no compromise on calculating the rating for the relevance of a Web search result, as per user needs.

Given below are the listings that show the search results obtained from Web search engine against the reordered search results based on the preferences of the user in its social network. Listing 4 shows the re-ordered search results based on the preferences of the user itself. Listing 5 shows the re-ordered search results based on the preferences of the user. While Listing 6 shows the re-ordered search results based on the preferences computed from the friends in the user’s social network.

Listing 4. Original search results for a sample query

|  |
|--|
| <b>Search Query:</b> Football  |
| <b>Community Preferences:</b> FIFA, American Football, Australian Football, Chelsea, Arsenal   |
| <b>Re-ordered Search Results:</b>  |
| <ol style="list-style-type: none"> <li>1. FIFA.com – Fédération Internationale de Football Association (FIFA)</li> <li>2. American football – Wikipedia, the free encyclopedia</li> <li>3. Chelsea Football Club – Official Site for News, Tickets &amp; Fixtures</li> <li>4. AFL – The official site of the Australian Football League – AFL.com.au</li> <li>5. The Official Website of Arsenal Football Club   Arsenal.com</li> <li>6. NFL.com – Official Site of the National Football League</li> <li>7. BBC SPORT   Football</li> <li>8. CBC.ca Meyer’s all-in approach a theme among college football coaches</li> <li>9. Official Site of the Premier League – Barclays Premier League News</li> <li>10. Manchester United Football Club</li> <li>11. Sky Sports   Football News</li> </ol> |

Listing 5. Reordered search results for a sample query based on individual user’s interests

Listing 6. Reordered search results for a sample query based on interests of users and its social network

However, we believe that the constraints on the number of participants and test queries did effect and may have caused some level of biasness in the evaluation results on accuracy and reliability of the ranking algorithms. In order to avoid this limitation, we have used click-through data that is recorded in search engine logs to simulate user experiences in Web search, and analyze it, as described in [19]. In general, when a user issues a query, the user usually checks search results in top to bottom manner. The user may click one or more results that look relevant and skips those documents that the user is not interested in. If our personalization mechanism ranks relevant results for the users higher in the search results list, the user might be more satisfied and browse through the re-ordered search results more than that of original search results. Therefore, we may utilize the user click history as a way to judge the level of satisfaction of the user to evaluate the ranking search accuracy. In this way, it has become easier for us to perform a large-scale evaluation. Furthermore, click-through data reflects the real world distribution of queries, users, and search scenarios. Thus, in our evaluation, usage of click-through data has brought results closer to the practical use cases in the evaluation of personalized search than that of using user surveys.

Click-through data in search engines can be seen as a set of triples  $(q, r, c)$  consisting of the query  $q$ , the ranking  $r$  presented to the user, and the set  $c$  of links the user clicked on. Listing 5 shows an example, the user asked a query, and received the ranking shown as order of results, and then clicked on the links. Since every query corresponds to one triple, the amount of data that is potentially available is virtually unlimited. Users, normally, do not click on links at random, but make a (somewhat) informed choice. While click-through data is typically noisy and clicks

**Table 26.5** Average DGC for search results

| Result set                | DGC   | Standard deviation |
|---------------------------|-------|--------------------|
| Original search results   | 0.518 | 0.119              |
| Re-ordered search results | 0.892 | 0.147              |

are not perfect relevance judgments, the clicks are likely to convey at least some information. Click-through data can be recorded with little overhead and without compromising the functionality and usefulness of the search engine. In particular, compared to explicit user feedback, it does not add any overhead for the user. The query  $q$  and the returned ranking  $r$  can easily be recorded whenever the resulting ranking is displayed to the user. For recording the clicks, a log file can be kept by a simple proxy system.

In order to collect the data and use it for experiments and testing, a simple Web search engine was built, that uses well-known Web search engines (like Google, Yahoo and Altavista) to collect the Web search results, and performs the re-ordering of the Web search results (as per user preferences) on client side. Our search engine works as follows: the users type query into the basic interface. This query is forwarded to the search engine. The results are returned by one of these basic search engines and the top 100 suggested links are extracted. Re-ordering of the search results is then performed and top 10 results are shown to the user. For each of the presented link, the system displays the title of the page along with its URL. The clicks of the user are recorded using the proxy system.

To be able to compare the quality of different retrieval functions, methods prescribed in [14] have been used. The idea is to present two rankings as A and B at the same time and then compare them after combining them as C. This particular form of presentation leads to a blind statistical test so that the clicks of the user demonstrate unbiased preferences. If the user scans the links of C from top to bottom, at any point he has seen almost equally many links from the top of A as from the top of B. To measure the ranking quality, we use the Discounted Cumulative gain (DCG) [12], which is a measure that takes into consideration the rank of relevant documents and allows the incorporation of different relevance levels. DCG is defined as follows:

$$DCG(i) = \begin{cases} G(1) & \text{if } i = 1 \\ DCG(i - 1) + G(i) / \log(i) & \text{otherwise} \end{cases}$$

where  $i$  is the rank of the result within the result set, and  $G(i)$  is the relevance level of the result. We used  $G(i) = 2$  for highly relevant documents,  $G(i) = 1$  for relevant ones, and  $G(i) = 0$  for non-relevant ones.

As shown in Table 26.5, automatic re-ordering of search results clearly improves search result quality. For example, the search results for query “football” with different users having different preferences have been reformulated by our system resulting in a better DCG.

The pure personalized results slightly overdue personalization, however, when combined with the original web ranks using the personalization, the original Google results are outperformed. The ranking quality obtained by re-ordering the search results based on the user preferences indicates the need for our approach of community-aware personalization strategies based on the information need. Click entropy, as introduced in [7], is defined as a measure of the variation in the information needs of users for a query as follows:

$$ClickEntropy(q) = \sum_{p \in P(q)} -P(p|q) \log P(p|q)$$

where,  $ClickEntropy(q)$  is the click entropy of query  $q$ .  $P(q)$  is the collection of web pages clicked on query  $q$ .  $P(p|q)$  is the percentage of clicks on web page  $p$  among all clicks on  $q$ , i.e.,  $P(p|q) = \frac{|Clicks(q,p,\bullet)|}{|Clicks(q,\bullet,\bullet)|}$

Click entropy is a direct indication of query click variation. If all users click only one identical page on query  $q$ , we have  $ClickEntropy(q) = 0$ . Smaller click entropy means that the majority of users agree with each other on a small number of web pages. In such cases, there is no need to do any personalization. Large click entropy indicates that many web pages were clicked for the query. This may mean that either a user has to select several pages to satisfy his/her information need, which means that the query is most likely an informational query. Therefore, personalization-based reordering of search results can help to in bringing more relevant results to the users or different users have different selections for a search query, which means that the query is an ambiguous query. In this case, personalization-based reordering of search results can be used to provide different web pages to different users, as per their needs.

We show the average search performance improvement of different personalization strategies on test queries with different click entropies in Fig. 26.2 that shows the improvement of personalized search performance increases with click entropy being larger, especially when click entropy is higher than 1.5. For click-based methods, individual-level personalization and group-based personalization, the improvement is limited for the queries with click entropy being lower, i.e. between 0 and 0.5. In the case of low click entropy, our approach brings only 0.37% of improvement which is not statistically significant. It shows that the users have general consistent clicks for the queries with low click entropy, and thus, no personalization is necessary for the current search results. However, for the queries where click entropy is significantly higher, say greater than 3.0, the result is obviously different. Both individual-level personalization and group-based personalization gets a dramatic improvement by having more relevant search results. In terms of ranking and re-ordering of search results, both, the group-based personalization and the individual-level personalization improve by 25%. This shows that the search queries with higher click entropy benefit more from personalization. Therefore, we perform personalization and re-ordering of search results, only when it is beneficial to do so (i.e. click entropy is significantly high).

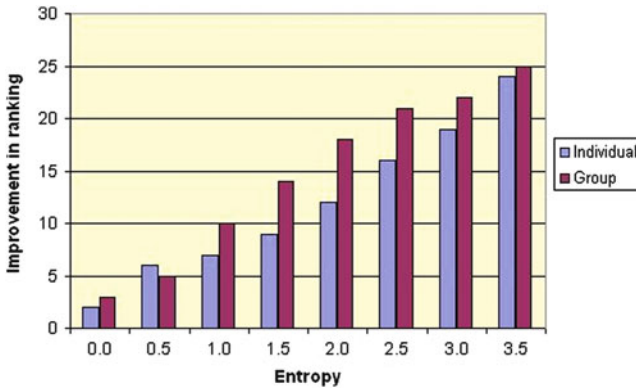


Fig. 26.2 Click entropy vs. improvement in ranking

## 26.4 Conclusions

In this book chapter, we presented our proposed solution to re-order Web search results based on the contextual information collected from a user's social network in a systematic manner. We have shown how the proposed system can help in bringing more relevant and interesting information for different users by re-ordering the search results from Web search engines, while avoiding irrelevant information. It enables the users to find out the information according to their interests in an easier and automated manner. We have developed various strategies, based on standard social network analysis techniques, for pulling out important publicly available information from the community of a user that is taken as important contextual information. It helps in extracting the required information of the community of a user from his/her social network and uses it to rank and re-order the search results of a Web search engine, based on the personal interests of the user and his/her community. Our design methodology in developing the community-aware personalized Web search kept our solution generic, to be able to customize it and apply it to any search scenario. We presented the overall architecture, described the ranking algorithms, and discussed the implementation and evaluation using well-known techniques and real-life datasets to show the viability of our proposed solution in real-life search scenarios.

## References

1. Almeida, R.B., Almeida, V.A.F.: A community-aware search engine. In: Proceedings of International World Wide Web Conference (WWW 2004). New York
2. Anick, P.: Using terminological feedback for web search refinement: a log-based study. In Proceedings of WWW 2004, New York, pp. 89–95 (2004)



3. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. *Scientific American Magazine* (2001). <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>
4. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. In: *Proceedings of WWW, Santa Clara* (1997)
5. Carminati, B., Ferrari, E., Perego, A.: Combining social networks and semantic web technologies for personalizing web access. In: *Proceedings of the 4th International Conference on Collaborative Computing (CollaborateCom 2008), Orlando* (2008)
6. Chen, H., Finin, T., Joshi, A.: An ontology for context-aware pervasive computing environments. *Special Issue on Ontologies for Distributed Systems, Knowledge Engineering* (2004)
7. Dou, Z., Song, R., Wen, J.R., Yuan, X.: Evaluating the effectiveness of personalized web search. *IEEE Trans. Knowl. Data Eng.* **21**(8), 1178–1190 (2009)
8. Freeman, L.: *The development of social network analysis*. Empirical Press, Vancouver (2006)
9. Google Personal (2009). <http://labs.google.com/personalized>
10. Granovetter, M.S.: The strength of weak ties. *Am. J. Soc.* **78**, 1360–1380 (1973)
11. Heath, T., Motta, E., Petre, M.: Computing word-of-mouth trust relationships in social networks from semantic Web and Web2.0 data sources. In: *Proceedings of the Workshop on Bridging the Gap Between Semantic Web and Web 2.0, 4th European Semantic Web Conference (ESWC2007), Innsbruck* (2007)
12. Jarvelin, K., Kekkonen, J.: Ir evaluation methods for retrieving highly relevant documents. In: *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, July 2000, Athens* (2000)
13. Jeh, G., Widom, J.: Scaling personalized web search. In: *Proceedings of WWW 2003*, pp. 271–279 (2003)
14. Joachims, T., Granka, L., Pan, B., Hembrooke, H., Gay, G.: Accurately interpreting click-through data as implicit feedback. In: *Proceedings of the Conference on Research and Development in Information Retrieval (SIGIR), Salvador* (2005)
15. McKeown, K.R., Elhadad, N., Hatzivassiloglou, V.: Leveraging a common representation for personalized search and summarization in a medical digital library. In: *Proceedings of ICDL 2003*, pp. 159–170. New Delhi, India (2003)
16. Morris, M.R., Teevan, J., Bush, S.: Enhancing collaborative Web search with personalization – Groupization, smart splitting, and group hit-highlighting. In: *Proceedings of CSCW 2008. SIGCHI/Association for Computing Machinery, New York* (2008)
17. Parr, B.: BREAKING News: Google announces social search. October 21st, 2009. <http://mashable.com/2009/10/21/breaking-google-launches-social-search>
18. Pattanasri, N., Jatowt, A., Tanaka, K.: Context-aware search inside e-learning materials using textbook ontologies. In: *Advances in Data and Web Management. LNCS, vol. 4505/2009*. Springer, Berlin (2009). doi:10.1007/978-3-540-72524-4, ISBN:978-3-540-72483-4
19. Shafiq, O., Alhadj, R., Rokne, J.G.: Community aware personalized web search. In: *Proceedings of the 2010 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2010), Odense* (2010)
20. Speretta, M., Gauch, S.: Personalizing search based on user search history. In: *Proceedings of CIKM, Washington* (2004)
21. Sugiyama, K., Hatano, K., Yoshikawa, M.: Adaptive web search based on user profile constructed without any effort from user. In: *Proceedings of WWW, New York*, pp. 675–684 (2004)
22. Tarjan, R.E.: A note on finding the bridges of a graph. *Inf. Process. Lett.* **2**, 160–161 (1974)
23. Teevan, J., Dumais, S.T., Horvitz, E.: Beyond the commons: investigating the value of personalizing Web search. In: *Proceedings of the Workshop on New Technologies for Personalized Information Access (PIA), Edinburgh* (2005)
24. Teevan, J., Dumais, S.T., Horvitz, E.: Characterizing the value of personalizing search. In: *Proceedings of SIGIR 2007, Amsterdam*, pp. 757–756 (2007)
25. Zeng, Y., Ren, X., Qin, Y., Zhong, N., Huang, Z., Wang, Y.: Social relation based scalable semantic search refinement. In: *Workshop on Scalable Semantic Data Processing, Asian Semantic Web Conference (ASWC 2009), 7th December 2009, Shanghai* (2009)