Social Network Extraction: Superficial Method and Information Retrieval

Mahyuddin K. M. Nasution, * Shahrul Azman Mohd. Noah, and Saidah Saad

Information Technology Department,
Fakultas Ilmu Komputer dan Teknologi Informasi (Fasilkom-TI)
Universitas Sumatera Utara, Padang Bulan, Medan 20155, Sumatera Utara, Indonesia
mahyuddin@usu.ac.id

Knowledge Technology Research Group, Faculty of Information Science & Technology
Universiti Kebangsaan Malaysia, Bangi 43600 UKM Selangor Malaysia
samn@ftsm.ukm.my, saidah@ftsm.ukm.my

Abstract. Social network has become one of the themes of government issues, mainly dealing with the chaos. The use of web is steadily gaining ground in these issues. However, most of the web documents are unstructured and lack of semantic. In this paper we proposed an Information Retrieval driven method for dealing with heterogeneity of features in the web. The proposed solution is to compare some approaches have shown the capacity to extract social relation: strength relations and relations based on online academic database.

Key words: superficial, unsupervised, supervised, keyword, similarity, association rule, semantic.

1 Introduction

Social networks is an approach for representing relations between individuals, groups or organization [1]. The concept of social networks extraction is designed to map the relationship of entities among them that can be observed, to mark the patterns of ties between social actors, to measure the social capital: the values obtained by the entities individually or in groups, to present a variety of social structures according to the interests and its implementation, based on different domains or information sources [2]. The extracted social networks can further be processed in information systems, i.e., in data mining systems that detect valuable trends in the social networks, in cases when selecting content based on certain semantic classes, or in expert systems that reason with the extracted knowledge in order to support government decision making. For example, in counterterrorism and the terrorism require networks for learning exchange of information and actions [3]. In scientometric, so extraction of academic social

* Proceeding of International Conference on Informatics for Development (ICID’11), c2-110 - c2-115 (2011).
networks aims to see the performance of human resources for example in education [4].

Web has been chosen as main source to gain various digital information about entities and their relations from all over the world. The web has been becoming the largest text database contained information about social networks, but most of data on the web are so unstructured that they can only be understood by humans, the other way the amount of data is so huge that they can only be processed efficiently by machines. The semantic representation of documents currently forms the vision of semantic web [5], in the same manner as the social network is a nucleus of semantic relation, i.e. possibilities of attributes partnership between actors where domains generate the attributes and the web documents provide the descriptions of entities and their relations.

Extracting social network is more than the text mining process and has been one of Artificial Intelligent (AI) research agendas, involving the task of identifying entities and their relations for trusted social network. As the manual semantic relation annotation of web documents is impractical and unscaleable, and so in the semantic web area there are two research streams for extracting social networks from web documents: unsupervised method and supervised method. The unsupervised methods utilize the Cartesian product for clustering the nodes $A$ in the network. The concept of clustering is $\gamma : A \times A \rightarrow R$ such that $\gamma(a, b) \in R$, $a, b \in A$ [6–8]. However, the clustering approach is difficult to generate the labels of relations in a network. The supervised methods employ a function $\lambda$ for classifying $Z$, i.e. $\lambda : Z \rightarrow C$ such that $\lambda(z) = c$, $z \in Z$, and $c \in C$ is a class label, where $C = \{c_1, c_2, \ldots, c_{|C|}\}$ is data set as special target attributes, $|C| \geq 2$ is the number of classes, and $Z \cap C = \emptyset$ [9–11]. The classification approaches however only concern with extraction of network based on predefined labels only, and thus cannot be adapted to the other descriptions of relations. Therefore, our goal is to enhance the superficial approach, an approach belong to unsupervised method, for extracting social network from web by not only relying on the co-occurrence but to consider other information.

2 Motivation and Related Works

An approach for extracting social network is by involving the superficial strategy that depends heavily on the co-occurrence. An occurrence, a singleton event of $a$ is a for $a \in A$, $A$ is a set of social actors, $a \subset \Omega$, whereas a co-occurrence, the doubleton event of $a$ and $b$, i.e., $a \cap b$, the subset of $\Omega$, and $a, b \in A$. $\Omega$ is a set of web pages indexed by a search engine, and $P : \Omega \rightarrow [0, 1]$ to be an uniform mass probability functions. To make the semantic relation between $a$ and $b$ from Web, we can explore singleton and doubleton, following a pioneer work [12] by implementing similarity measure with conditions that $|a| \geq |a \cap b|$ and $|b| \geq |a \cap b|$ [13]. At the time of doing this experiment, a Yahoo! search for "Shahrul Azman Mohd Noah", returned 1,200 hits, whereas for "Abdullah Mohd Zin" Yahoo! Search Engine returned 3,870 hits. Searching the pages where both "Shahrul Azman Mohd Noah" and "Abdullah Mohd Zin" = 13 hits, and
showed that singletons is greater or equal to doubleton, but we have found that the conditions is not met. Moreover, singletons and doubleton always took along bias in relation, due to the limitations of search engine which due to ambiguity of results.

2.1 Strength Relation and Similarity (SRS)

Relations differ in strength. Such strength can be operationalized in a number of ways. With respect to event frequencies of the pairs of actors where they may exchange large or small amounts of social capital: money, goods, papers, or services. They may supply important or trivial information. Such aspects of relationships measure different types of relational strength. Flink system has been developed to extract, aggregate and visualize a social network [6]. In POLYPHONET system has been created the procedures to expand superficial strategy by providing keywords in query submitted to search engine [7]. However, these system took along special cases: Flink for building the social network of a Semantic Web community by utilizing the friend-of-a-friend (FOAF) semantic profile [6], and POLYPHONET developed to identify the relations in the Japan AI conference [8].

The content of a relation refers to the resource that is exchanged that may generate some labels of relation. In this case, each entity will be assigned with multiple labels extracted from information sources related to that entity. Suitable labels will be assigned to relations by using the Information Retrieval technique, mainly the generative probabilistic model (PGM) [14]. The parameters of GPM are used as modalities to get the knowledge from the corpus, and generating strength relation based on labels of entities. For example, the strength relation between entities based on participants in the same conference or workshop [15].

Similarity measures the relation between two entities are a strategy to generate strength relation. One of the used similarity measure widely is the Jaccard coefficient, in singleton and doubleton we have

$$sim_{jac}(a, b) = \frac{|a \cap b|}{(|a| + |b| - |a \cap b|)} \quad (1)$$

2.2 Underlying Strength Relation (USR)

Underlying strength relation is an approach to exploit URL (Uniform Resource Locator) addresses and its organization since URL address is always available in Web snippets, returned by any search engine [1]. Syntactically, an URL represents a resource in Internet. A composition of URL contains a set of tokens, $U = \{s, d_1, \ldots, d_m, p_1, \ldots, p_n\}$, satisfying a structure: $s : //d_m \ldots d_2.d_1/p_1/p_2/\ldots/p_n$; i.e. a string consists of scheme, authority, and path. The scheme is a token $s$, a component contains a protocol that is used for communicating in Internet. For example, http (Hypertext Transfer Protocol), https (http Security) and other protocols. The authority is string of tokens as $d_m \ldots d_2.d_1$, i.e. a component has three subcomponents: user information, host, and port:
1. The user information may consist of a user name and, optionally, scheme-specific information about how to gain authorization to access the resource. Usually, it is followed by a commercial at-sign('"@"') that delimits it from the host, if present like in an email address mahyuddin@usu.ac.id.

2. The host contains a location of a web server, where the location can be describe as either a domain name system (DNS) or internet protocol (IP) address.

3. The port is a specific number. For instance, a default port number (80 for the http protocol), i.e. \( s://d_m\ldots.d_1:80/\). The colon symbol ('":"') should be prefixed prior to the port number.

The last string of tokens \( d_1/p_1/p_2/\ldots/p_{n-1} \) is a path, i.e. a component contains the directories including a web page and a file name of the page, where a directory and a file are separated by the slash symbol ('"/"'). The last token of path sometimes comes with two other components: query and fragment. For example, http://search.yahoo.com/search;_ylt=AjoEJr09wuxK34pfA74_RvC bvZx4?vc=&fp_ip=mykp=mahyuddin+K.+Nasution&toggle=1&cop=mss&ei=UTF-8&fr=yfp-t-701. The query is a component containing parameter names and values that may to be supplied to web applications. The token of path and the query are separated by the question symbol ('"?"'). The form of query is name=[value], where there is equal symbol ('"="') between a parameter name and a parameter value. A pair of name=[value] is separated each other by the ampersand symbol ('"&"'). The fragment is a component for indicating a parameter part of a document. This last component and the part previously mediated by a sharp symbol ('"#"'). Consequently, be found same URL addresses, but presented in a different threads. Therefore, necessary to the canonicalization of the URL.

URL address indicates the layered structure of a web site which can be logically shown as a hierarchy. As such the URL of web pages which provides and indicator of its logical position in the hierarchy structure that can be considered as the underlying strength of the relationship, where site editors usually tend to put similar or related web pages as close as possible underlying relations among entities in the case that co-occurrence measures unable to provide such relations.

For any web snippet produces as a results of the entity name query, there exists a set of \( k \) URL addresses. Therefore, there will be \( n_1 + \ldots + n_k \) URL addresses whereby \( n_i, i=1,\ldots,k, \) is the number of layers for each \( i \). For these generated URLs, there is a possibility of redundant URLs. Let \( u \) is the number of same URL address. For each entity \( a \in A \) we can derive a vector space \( a = [a_1,\ldots,a_K] \) where \( a_j = u_{n_i} \). We, therefore, can measure the distance between the two entities based on the list of URL addresses from Web snippets [1].

### 2.3 Association Rule and Similarity (ARS)

Another implementation of co-occurrence is a formulation of basic data mining, e.g. the association rule [16]. Assume \( B = \{b_1,b_2,\ldots,b_H\} \) is a set of attribute literals, and a set of transaction \( M_i \) are subsets of attributes, or \( M_i \) be subsets of \( B \). Then, we define the implication, \( X \Rightarrow Y \) with two possible values \( T = \text{TRUE} \).
or \( F = \text{FALSE} \) as association rule if \( X \) be a subset of \( B \), \( Y \) be a subset \( B \) and \( X \cap Y \neq \emptyset \). Let \( q = "a \text{ AND } x" \) is a query, where \( x \) is a keyword and \( a \) is a name of actor \( a \in A \) as seeds, and \( Db_i \) is a collection of document containing names of actors \( b_i \in A \), then the transactions be \( M_i = \{(q, b_i)\} \) or \( \{(q \Rightarrow b_i)\} \), \( q \in X, b_i \in Y, M_i \in M \). So, by making \( q \) always \( T \), (see Table 1), we obtain a conditional probability as follows.

**Table 1. Transaction and implication**

| Transaction | \( q = (t_a, t_x) \) | \( b_i \) | Implication |
|-------------|----------------|--------|-------------|
| \( M_1 \)   | T             | T      | \( a \Rightarrow t_x \Rightarrow b_i \) |
| \( M_2 \)   | T             | T      | \( ... \Rightarrow T \Rightarrow ... \) |
| ...         | ...           | ...    | ...         |
| \( M_{|M|} \)| T             | T      | \( ... \Rightarrow ... \) |
$V = \{v_i | i = 1, \ldots, n\}, V \neq \emptyset$. The nodes in a social network refer to actor names such as authors, recipients, researchers, artists. Therefore, the first task of extracting social network, $\xi$, is to identify the actors. This can be achieved by providing a list of names as seeds in order to extract other names, recognize and disambiguate them [19, 20]. The actors play some role in a social based on their background, and they have some characteristics as attributes. Formally, such attributes of actors we define as $Z = \{z_j | j = 1, \ldots, m\}$ which are attributes / characteristics of entities whereby a pair of $\langle A, Z \rangle$ is the instance of actors, where $Z_i$ are subsets of $Z$, $Z_i$ are subsets of attributes of each entity $a_i$, i.e. $\langle a_i, Z_i \rangle$, $i = 1, \ldots, n$, or simply denotes a set of attributes of entity a as $Z_a$.

We developed an approach to extracting keywords from web snippets for disambiguating names in social networks. The methods for using keyword in query are [22]:

- noK : A name pair used as query without relation keywords.
- K1 : A name pair and top-weighted relation keyword used as a query.
- K2 : A name pair and second-weighted relation keyword used as a query.
- K1+K2 : A name pair and top-weighted strongest relation keyword.

By considering a set of snippets $L$, each contains a reference to a person. Let $P = \{P_1, \ldots, P_{|P|}\}$ be a partition of $\xi$ and $\zeta$ into references to the same person, so for example $P_1 = \{S_1, S_4, S_5, S_9\}$ might be a set of references to "Abdul Razak Hamdan" the information technology professor. We produce the current context for each actor where keywords have two vector, $\nu$ and $\upsilon$: first vector we define based on TF.IDF and second vector we generate from hit count, and then we define delta $\delta$ as distance of two vector. We use delta to select keyword from each classes of candidates that grouped by tree of network semantic of words [1].

Let $C = \{C_1, \ldots, C_{|C|}\}$ be a collection of disjoint subset of $L$ created by algorithm and manually validated such that each $S$ has an identifier, i.e. URL address. Then, we denote $L_C$ as references that be the clusters based on collection. Based on measure were introduced, we define an notation of recall $Rec()$ as follows

$$Rec(S_i) = \frac{|\{S \in P(S_i) : C(S) = C(S_i)\}|}{|\{S \in P(S_i)\}|} \quad (6)$$

and a notation of precision $Prec()$ as follows

$$Prec(S_i) = \frac{|\{S \in C(S_i) : P(S) = P(S_i)\}|}{|\{S \in C(S_i)\}|} \quad (7)$$

where $P(S_i)$ as a set $P_i$ containing reference $S_i$ and $C(S_i)$ to be the set $C_i$ containing $S_i$. Thus, the precision of a reference to "Abdul Razak Hamdan" is the fraction of references in the same cluster that are also to "Abdul Razak Hamdan". We obtain average of: recall (REC), precision (PREC), and F-measure for the clustering $C$ as follows:

$$REC = \frac{\sum_{S \in L_C} Rec(S)}{|L_C|} \quad (8)$$
Social Network Extraction: Superficial Method and ...

\[
P_{REC} = \frac{\sum_{S \in L_C} P_{rec}(S)}{|L_C|} \\
F = \frac{2 \cdot REC \cdot P_{REC}}{REC + P_{REC}}
\]

(9)  

(10)

Second step, for building any network is to determine an edge based on a concept of graph theory [18]. In social network, the edges refer to relationships between actors. Therefore, another main objective of extracting social network is to identify relation among entities, i.e., second task of social network extraction, \( \zeta \). If there exists \( R = \{r_1, \ldots, r_k\} \) as a set of possible relations between social actors, then we have \( R = A \rightarrow A \times A = r(a, b), \forall a, b \in A \), that can be depicted as the overlap principle, i.e. the intersection their attributes \( r_k(a, b) = Z_a \cap Z_b \).

This means that the relations among actors are formed by sharing attributes, ideas, and concepts between them.

Social network extraction provides the technology to identify and describe content. The technology is as an exertion to further acquire rich and trusted social network. Such social network formally is \( SN = (V, E, A, R, Z, \xi, \zeta) \), \( V \neq \emptyset \), \( A \neq \emptyset \) that satisfies the following conditions:

1. \( \xi : A \xrightarrow{1:1} V \), \( v = \xi(a) \), for every \( a \in A \) there is only one \( v \in V \);
2. \( \zeta : R \rightarrow E \) so that \( e_j = \zeta(r_k(a, b)) = \zeta(Z_a \cap Z_b) \), \( e_j \in E \), \( r_k \in R \), \( \forall a, b \in A \), where \( Z_a, Z_b, Z_a \cap Z_b \subset Z \).

The labeled social network is to present a variety of social structures according to the interests and its implementation. The strength relation in social network tends to ambiguity. One or more relations may connect a pair of actors, if it applies, we have social network with many ties. Pairs may maintain a tie based on one relation only, e.g., as members of the same organization, or they may maintain a multiplex tie, based on many relations, such as sharing information, giving financial support and attending conference together. Semantically, the composition of a relation or a tie is derived from the social attributes of both participants, for example is the tie between different or same sex dyads, between a supervisor and an underling or between two peers. Computer network tends to underplay the social cues of participants than the connections, but once a computer network is social network. Once again, we define the social network as \( SN_{des} = (V, E, A, R, Z, \xi, \zeta, \varphi_1, \varphi_2) \), \( V \neq \emptyset \), \( A \neq \emptyset \), a labelled social network satisfying the following conditions:

1. There exists \( \varphi_1 : Z \rightarrow V \) so that \( \forall z_i \in Z_a, \varphi_1(z_i) \in V \), and \( \forall a \in A \) and \( Z_a \subset Z \);
2. There exists \( \varphi_2 : Z \rightarrow E \) so that \( \forall z_j \in Z_a \cap Z_b, \varphi_2(z_j) \in E \), \( \forall a, b \in A \), and \( Z_a, Z_b, Z_a \cap Z_b \subset Z \).

Extracting social network from web documents need technologies. Those technologies will become an important component of any retrieval system or as Information Retrieval (IR) where the representations of documents and query are enriched with semantic information. From social network perspective, it will
give emersion that can adapted and advance the retrieval models, i.e. the representations of query and document relations and by the function that estimates the relevance of relations to a query. Therefore, we can consider relationship extraction as an information retrieval task. Let \( G_1 = (V_1, E_1) \) is a graph as the resulted social networks and \( G_2 = (V_2, E_2) \) is a benchmark graph, we have Jaccard-coefficient, or Eqs. (1) and (8)-(10) be [22]

\[
sim_G = \frac{|E_1 \cap E_2|}{|E_1| + |E_2| - |E_1 \cap E_2|}
\]

\[
Precision = \frac{|E_1 \cap E_2|}{|E_1|},
\]

\[
Recall = \frac{|E_1 \cap E_2|}{|E_2|}
\]

and

\[
F_{\text{measure}} = \frac{2(|E_1 \cap E_2|)^2}{|E_1| \cdot |E_1 \cap E_2| + |E_2| \cdot |E_1 \cap E_2|}.
\]

Information retrieval is concerned with answering information needs as accurately as possible, where relations retrieval typically involves the querying of structured data from unstructured information.

4 Result

4.1 Evaluation and Dataset

For evaluation of the approaches, we have gathered and labeled a dataset of 539 Web pages, Tab. 2, where the label we created based on URL address of each web page in snippet.

| Personal name                      | Position    | Number of pages |
|-----------------------------------|-------------|-----------------|
| Abdul Razak Hamdan                | Dean        | 85              |
| Abdullah Mohd Zin                  | Professor   | 90              |
| Shahrul Azman Mohd Noah           | Professor   | 134             |
| Tengku Mohd Tengku Sembok         | Professor   | 189             |
| Md Jan Nordin                     | Assoc. Prof.| 41              |
| **Total**                         |             | **539**         |

As benchmark of social network, we possess \( |V_2| = 67 \) nodes and \( |E_2| = 253 \) edges. This graph we derived from online database DBLP to 67 academic persons, we extract by using association rule (ARS) and then evaluated and corrected their relations based on author-coauthor relationship.
4.2 Experiment

Let us consider information context of actors that includes all relevant relationships as well as interaction history, where Yahoo! Search engine fall short of utilizing any specific information, especially context information, and just use full text index search in web snippets. In experiment, we use maximum of 600 web snippets for search an actor $a$, and we consider words where the TF.IDF value $> 0.3 \times$ highest value of TF.IDF, or maximum number is 30 words, see Fig. 1. We used Eqs. (6), (7) and (10) for computing recall, precision and F-measure.

| Method | Recall | Precision | F-measure |
|--------|--------|-----------|-----------|
| Delta (6) | 45.8 | 29.5 | 35.9 |

Under recall, precision and F-measure: Eqs. (8)-(10), and results in Table 3, this method shows something to consider, i.e. the number of words in the cluster should be limited so that an average value of measurement is not affected by the lower (see Fig. 1). Given the implementation of this method is done with the Yahoo! search, this result is still reasonable.

We test for 213 actor names and there are 22,683 potential relations, but at the time of doing this experiment by using SRS there are only 12,621 (53\%) relations which satisfy threshold $\alpha = 0.0001$. However, the method (USR) that involves URL-computation able to identify 19,513 (86\%) relations for $\alpha = 0.01$. 
Table 4. Recall and precision of social network

| Method | Recall | Precision | F-measure |
|--------|--------|-----------|-----------|
| SRS    | 120/253 (47%) | 120/12,621 (10%) | 27% |
| URS    | 176/253 (70%) | 176/19,513 (9%) | 18% |

Table 4 shows the result of correlation with the SRS and USR where based on Jaccard coefficient $\text{sim}(\text{SRS, ASR}) = 0.0094$ almost the same with $\text{sim}(\text{URS, ASR}) = 0.009$, but looking at those results USR shows good performance.

5 Conclusion and Future Work

The method used URL as object for extracting social network to be incorporated into existing tracting method of social network. It shows how to uncover underlying strength relations by exploiting Web snippets and URL structure. This showed that a well-known paradigm of querying a document web more simple for accessing by inputting keyword, but most simply by using URL address, because on the word or keyword we have typical problems of synonymy and ambiguity, but URL not exist. Our near future work is to further experiment he proposed method and look into the possibility of enhancing IR performance by using social network with developing benchmark manually.

References

1. Nasution, Mahyuddin K. M., and Noah, S. A. M.: Superficial method for extracting social network for academic using Web snippets. In: Yu, J. et al. (eds.): Rough Set and Knowledge Technology (RSKT), LNAI, vol. 6401, 483-390, Springer, Heidelberg (2010).
2. Nasution, Mahyuddin K. M., and Noah, S. A. M.: Extracting social networks from web documents. In Book of Abstracts (CAIT 2010), The 1st National Doctoral Seminar on Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 27-28 September, 278-281 (2010).
3. Meter, K van: Terrorists/liberators: Researching and dealing with adversary social networks, Connections, Vol. 24, No. 3, 2002, 66-78 (2002).
4. Mika, P., Elfring, T. and Groenewegen, P.: Application of semantic technology for social network analysis in the sciences, Scientometrics, Vol. 68, No. 1, 3-27 (2006).
5. Barner-Lee, T., Hendler, J., and Lassila, O.: The semantic web, Scientific American, Vol. 284, No. 5, 34-43 (2001).
6. Mika, P.: Flink: Semantic web technology for the extraction and analysis of social network, Journal of Web Semantics, Vol. 3, 211-223 (2005).
7. Matsuo, Y., Mori, J., Hamasaki, H., Nishimura, T., Takeda, H., Hasida, K. and Ishizuka, M.: POLYPHONET: An advanced social networks extraction system, Proceedings of the 15th International Conference on World Wide Web (WWW 2006), 397-406 (2006).
8. Matsuo, Y., Mori, J., Hamasaki, M., Nishimura, T., Takeda, T., Hasida, K., and Ishizuka, M.: POLYPHONET: An advanced social networks extraction system
from the Web. Journal of Web Semantics, Science, Services, and Agents on the World Wise Web 5, 262-278, (2007).
9. McCallum, A., Corrada-Emmanuel, A., and Wang, X.: Topic and role discovery in social networks, Proceedings of the 19th International Joint Conference on Artificial Intelligence, 786-791 (2005).
10. McCallum, A., Wang, X. and Corrada-Emmanuel, A.: Topic and role discovery in social networks with experiments on Enron and Academic email, Journal of Artificial Intelligence Research, Vol. 30, 249-272 (2007).
11. Tang, J., Zhang, D., Yao, L., Li, J., Zhang, L., and Su, Z.: ArnetMiner: Extraction and mining of academic social networks, KDD 2008, 900-998 (2008).
12. Kautz, H., Selman, B., and Shah, M.: The hidden web, AI Magazine, Vol.18, No. 2, 27-36, (1997).
13. Kautz, H., Selman, B., and Shah, M.: ReferralWeb: Combining social networks and collaborative filtering, Communications of the ACM, Vol. 40, No. 3, (1997).
14. McCallum, A., Corrada-Emmanuel, A. and Wang, X.: The author-recipient topic model for topic and role discovery in social network: Experiments with Enron and Academic email, Technical Report UM_CS-2004-096, Amherst: University of Massachusetts, (2004).
15. Tang, J., Zhang, D., Yao, L., and Li, J.: Extraction and mining of academic social network, WWW 2008, 1193-1194 (2008).
16. Nasution, Mahyuddin, K. M., and Noah, S. A. M.: Extraction of academic social network from online database, Proceeding of International Conference on Semantic Technology and Information Retrieval (STAIR’11), IEEE, 28-29 June, 64-69 (2011).
17. Lin, N.: Building a network theory of social capital, Connection, Vol. 22, No. 1, 28-51 (1999).
18. Newman, M. E. J., and Park, J.: Why social networks are different from other type of network, arXiv:cond-mat/03050612v1, 26 May, 1-9 (2003).
19. Balog, K., Azzopardi, L., and De Rijke, M.: Resolving person names in Web people search, in Weaving services and people on the World Wide Web, I. King and R. Baeza-Yates (Eds.), Heidelberg: Springer-Verlag, 301-323 (2009).
20. Cullota, A., Bekkerman, R., and McCallum, A.: Extracting social networks and contact information from email and the Web, Proceedings of the 1st Conference on Email and Anti-Spam (CEAS), (2004).
21. Jin, Y.-Z., Matsuo, Y., and Ishizuka, M.: Extracting social network among various entities on the web, ESWC 2007, LNCS Vol. 4519, Heidelberg: Springer, 251-266 (2007).
22. Mika, P.: Social Networks and the Semantic Web, Berlin: Springer-Verlag (2007).