Deriving user's profile from sparse egocentric networks: using snowball sampling and link prediction

Sirinya On-At, Marie-Françoise Canut, Florence Sèdes, André Péninou

To link to this article : DOI: 10.1109/ICDIM.2014.6991421
URL : http://dx.doi.org/10.1109/ICDIM.2014.6991421

Deriving User's Profile from Sparse Egocentric Networks

Using Snowball Sampling and Link Prediction

Sirinya On-at, Marie-Françoise Canut, André Péninou, Florence Sèdes
Institut de Recherche en Informatique de Toulouse (IRIT)
118 route de narbonne 31602 Toulouse, France
(sirinya.on-at, canut, peninou, sedes)@irit.fr

***Abstract*— Several studies demonstrate effectiveness and benefits of using user's social network information to enrich user's profile. In this context, one of our contributions [1] proposes an algorithm enabling to compute user's interests using information from egocentric network extracted communities. Therefore, mining information from a small or a sparse network remains challenging because there is not enough information to enrich a relevant user's profile. So, one of the main lock is to cope with the lack of information that is considered as an important issue to extract a relevant community and could lead to misinterpretations in the user's profile modeling process. We aim to improve the performance of [1], regarding the lack of information problem, in the case of a small and/or a sparse network. We propose to add more information (i.e. relations) into user's network before extracting the data and enriching his profile. To achieve this enrichment, we suggest using snowball-sampling technique to identify and add user's distance-2 neighbors (friends of a friend) into the user's egocentric network. Our experimentation conducted in DBLP demonstrates the interest of node integration into small and sparse network. This leads to the study of link prediction that enables us to provide better performances and results compared to the existing work.**

Keywords*— *User modeling, social network analysis, link prediction, snowball sampling

I. Introduction

With social content sharing explosion, the amount information shared becomes a rich source of knowledge but sometimes superfluous for a user. To propose information adapted to the user's specific needs, information adaptive mechanisms (e.g. recommendation, personalization) rely on user's profile modeling enabling to collect and identify his personal information and interests.

As social networks are growing in terms of active users, resources and interactions, several studies are interested in using this kind of information to enrich user' profile (e.g. [1]–[6]). In this context, one of our contributions [1] proposes a community-based algorithm to derive user's profile from his egocentric network (a social network widely used in sociology that focuses on user's neighbor's immediate interconnections). This algorithm computes user's interests using information from communities extracted from his egocentric network. We are interested in improving the results of this work in case of small and/or sparse networks. In fact, communities that are detected from such characteristics of network are usually proved irrelevant. Thus, this could lead to misinterpretations in the user's profile's modeling process.

To overcome these drawbacks, we focus on the user's egocentric network construction phase, in order to provide a relevant user's egocentric network before applying [1]'s proposed algorithm. We propose to integrate more users' neighbors into user's egocentric network, in order to facilitate the community extraction phase. There are two techniques considered in our works. Firstly, we propose using snowball sampling to integrate the distance-2 neighbors (friends of a friend) of the user into the network. Secondly, in order to filter and add the relevant ties and individuals into the user's egocentric network, we apply link prediction technique on the extended egocentric network computed from the first approach.

Our evaluation has been conducted in the co-authors network DBLP. This evaluation allows us to demonstrate the interest of integrating the distance-2 neighbors of the user into very small and/or sparse networks before running the user's profiling process. However, the technique does not perform well in other networks having larger size. This leads us to combine the first technique with the link prediction that enable to provide better performances and results compared to the first technique and the existing work.

The rest of this paper is structured as follows: in the next section we present related works. In the third section we describe a preliminary study on snowball sampling and link prediction that leads to our proposed extended approach. In section four, we illustrate the experimentation. In section five, we present and comment the results of our experiments. The last section concludes and presents the perspectives of our work.

II. Related Works

A. User profiling from social network

User's profile is generally represented with weighted user's interests in one or several domain (e.g. sport, music). It can contain also the user contextual information (e.g. location, time). In an information system, a user's profile is built and enriched over time upon his interactions with the systems (e.g. sharing information, annotating resources, publishing scientific papers). Nevertheless, the profile of new or less-active user could be empty and do not contain any useful interests for a mechanism of personalization or recommendation. This problem is identified as the *cold start problem* [7].

To solve this problem, it's necessary to find additional user information in order to complete non-existent/missing profile. One of effective techniques is to use information from user's social network. Different approaches of user's profile enrichment using user's social network information have been proposed. [4]–[6] select individually the user's social network to describe his interests. Each approach strongly depends on underlying mechanism that uses built profiles and on each application domain (e.g. search engines, products recommendation), not much has been done to build a generic profile that can be applied independently of the application mechanisms.

A. *User profiling from egocentric network*

Instead of considering only some individually selected people in the user's social network to describe the user, [1] proposes a community (extracted from his egocentric network) based algorithm to enrich user's interests. This algorithm can be reused in any application context and in any mechanism. Obviously, this work considers that the user is better described by communities of people around him, especially the people that are directly connected to him as demonstrated in [8], [9]. We describe the ego-centric network of a user as follows: for each user (u) we consider the undirected graph G(u) = (V, E) where V is the set of nodes directly connected to u, and E is the set of relationships between each node pair of V. We emphasize that u is not included in V.

[1] presents a user's profile as a vector of weighted user's interests. Each user's profile is composed of two dimensions. Firstly, the user dimension which contains user's interests computed by using only the user behavior. Secondly, the social dimension which contains user's interests computed by using the behavior of people in the user's egocentric network (G). This work introduced a user' social dimension building process (named CoBSP – Community based Social Profile), consisting of four steps:

Step 1 consists in extracting communities from the user's egocentric network. This phase is realized by applying iLCD algorithm [10] which performs very well with overlap and outperforms other algorithms particularly for egocentric networks.

Step 2 consists in computing the profile of each community found in the first step. The profile of a community is computed by analyzing the behavior of all members of this community.

Step 3 consists in computing the weight of each interest in the social dimension of the user's profile. The weight of an interest i in the community c depends on structural score and semantic score vary with the parameter α as presented in the following formula.

$$w(i,c) = \alpha \text{ Struct\_Score}(c) + (1-\alpha)\text{Semantic\_score}(i,c) \quad (1)$$

Finally, we derive the social dimension according to the weights calculated in the third step.

The performance of CoBSP algorithm has been proved with empirical results. Therefore, their experiment considered only the users having at least 50 persons in their egocentric network because the information from the smaller egocentric-network is considered insufficient.

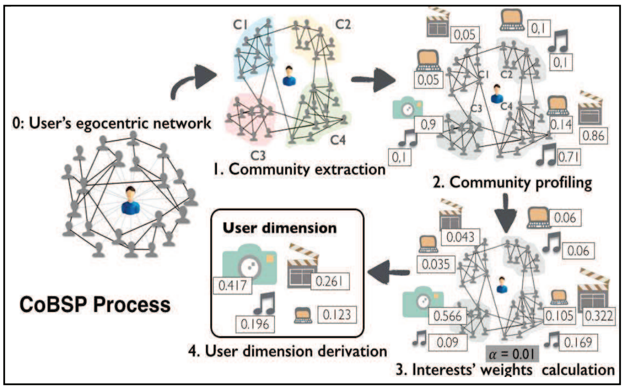

Fig. 1. Exemple of CoBSP's process using α = 0.01

We are interested in extending this existing work so that the algorithm performs better in such environment. We suggest adding more relevant relations into the user's egocentric graph.

III. EXTENSION APPROACHES

A. *Preliminary study*

Our contribution leads to the study of snowball sampling and link prediction that allow us to integrate more relations into the user' egocentric network and select only the relevant ones.

1) *Snowball sampling*

"Snowball sampling" is a sampling technique mainly used in research to locate further information in order to increase research sample's member. This method allows collecting information like a small snowball rolling on a snow-covered space and pick up more and more snow as it rolls. The process consists in searching for additional members of a research sample by using the existing members: we use existing members of the sample to identify the other external individuals who are, then, used to refer to other individuals until we reach the (significant) sample size [11]–[13]. In the other hand, we can consider this technique as transitive trust chain mining [14], [15]. Based on the knowledge transfers chain, we can identify people who share the same interest or information.

With the same concept, in social network, we can take advantage of interactions links between individuals in the network to identify ones who are in the same target (i.e. link-tracking). [14] uses the snowball sampling technique to build the egocentric networks of a user with the following principle: given k, the distance between two nodes in the network, we start from the user and search for others individuals whom he has connection with (k=1). Then, we start from the latter one to search for the others individuals (k=2) and so on, until the sample size is reached. If the size is not defined, we utilize the random number as the maximum size of the sample.

Snowball sampling has been widely used in sociology researches to sample hard-to-reach population (e.g. criminals, isolated people, prostitutes, HIV patients…). The technique can be also used to cope with the lack of information or population [13].

In our approach, we aim to use snowball sampling to extend user's egocentric network, for small or sparse networks, by seeking the individuals linked with the existing user's neighbors. However snowball sampling is not enough accurate as stand-alone tool. In order to obtain relevant information, we have to combine this method with other information selection technique. In this case, we propose to filter population by using a so-called "friends of friends" property. Based on [8], [9], we consider that apart from the direct contacts of user, the direct contact of contacts (i.e. friends of friends) can be the second meaningful source to describe the user.

However, we need to emphasize that the most important objective of using snowball sampling in our work is to gain more sample population in the case of the lack of information. The representativeness of the sample obtained with this technique is not always guaranteed because we have no idea of the true distribution of the population and of the sample. As we can see in the real world, it is not always certain that we know all the friens of our friends.

In such case, we need to select only the signifcant nodes. This leads us to the study of the techniques that enable to idenfity and take into account the most relevant nodes in our work. In the social network context, link prediction is one that fits well to this requirement.

2) Link prediction

Link prediction is a technique widely used in graph mining and network analysis. It consists in predicting, from a given network, which pairs of nodes are likely to link together. This technique is used to predict new links that could be formed between nodes in the future. In this case, the task can be defined as follows: given a pair of nodes n1 and n2, from a graph G, that have never connected at timestamp t1, we compute probability that these two form a link at timestamps ti (i>1). Furthermore, we can use link prediction to discover the links that couldn't be directly observed in the network but are likely to exist. This one is useful to complete and extend the network with relevant information. The link prediction has a variety of application: detecting novel relations in social network [16], identifying the collaboration in the organization, identifying suspicious links in the network [17][18].

Due to its broad applicability, a variety of approaches and algorithms are proposed [19]–[25]. The most popular ones are based on node features, probabilistic model and topological. The probabilistic approach consists in building a model using the observed social network. Examples of models in this approach are Relational Markov Networks, Relational Bayesian Networks. In the node features based approaches, we compute similarity measures by using their content or semantics. The topological based approach consists in computing similarity scores between two nodes by exploring structural pattern of the network in the analysis. The latter one is the most widespread due to its effectiveness and simplicity. That is why our proposed approach is based on this approach.

[25] studied various methods of link prediction using topological features extracted from coauthorship network including neighborhood based methods, path based methods and some height level approaches. We are interested in our work, the neighborhood methods including common neighbor, Jaccard's coefficient, Adamic/Adar. The common neighbor is a number of neighbors that are shared by a pair of nodes in the network. The Adamic/Adar measure [26] weights the importance of a common neighbor by the rarity of relationships between other nodes and this one (i.e. a node pair having a common neighbor that is not common to several nodes is considered more important). Jaccard's coefficient is a normalized measure of common neighbor by dividing with the number of total neighbors. Finally, the preferential attachment method [27] which indicates that new links are more likely to be formed with higher degree nodes. The similarity score is the product of the degrees of the two nodes.

The effectiveness of theses methods was evaluated in [25] by calculating the prediction accuracy improved over a random predictor. The results show that, in spite of its simplicity, Adamic/Adar perform well in term of accuracy. The similarity scores is presented by the following formula.

$$score(x,y) = \sum_{z \epsilon \Gamma(x) \cap \Gamma(y)} \frac{1}{\log |\Gamma(z)|} \quad (2)$$

B. Proposed approach

The aim of this work is to extend the algorithm CoBSP proposed in [28], in order to apply efficiently the community-extracting algorithm to a tiny and/or sparse network. At first, we propose to increase the size and the density of the network by adding new significant nodes. Thus, this work is located in the step 0 of the CoBSP process (i.e. the eccentric network preparation step before applying CoBSP process).

1) Network extension

Based on the snowball sampling technique, we take into account the individuals at stage k>1 in the user's egocentric network. We pretend that these individuals are directly connected with the user. Following the user's egocentric graph G(u) defined in section 2, we consider k, the distance between the user u and V . With k = 1, we consider only the individuals directly connected with the user. With k=2, we also integrate the individuals connected with the user at distance 2 (friends of friends) in the set of his neighbors and so on.

We present our user's egocentric extended graph as followed: for each user (u) we consider the undirected graph $G(u) = (V_k,E_k)$ where V_k is the set of individuals connected to u at distance $k \in N$ and E is the set of interactions between the members of V_k.. We propose in the first time to define the tiny value of k (k=2).

After extending the user's egocentric network from the first step, we obtain a graph $G_2(u) = (V_2,E_2)$, the user's egocentric network composing of his distance-1 and distance-2 neighbors. With this technique, the graph's nodes increase exponentially. We can obtain sufficient user's egocentric nodes before applying CoBSP algorithm. Note that we can also obtain, in this extended network, a large number of distance-2 neighbors (It's possible that the user's directed neigbors posssessed in his turn a lot of neigbors). And one more, it is possible that ones of them are not relevant.

To avoid taking into account the non-relevant nodes in the user profiling process, we propose to apply the link prediction method, on the extended user's egocentric network.

2) Selection of relevants individuals using link prediction technique

After studying several link prediction methods in social networks, we are interested in topological based methods. Based on the comparison of the methods of predictions of links from [25], we adopt Adamic/Adar method in our work.

We propose then, to combine the link prediction technique with our previous approach based on snowball sampling technique. Thus, we propose to apply link prediction method on the extended egocentric network $G_2(u)$. Given a graph $G_2(u) = (V_2,E_2)$, we apply to each node pair v_x, $v_y \in V_2$, the Adamic/Adar method:

$$score(Vx,Vy) := \sum_{Vz \in \Gamma(Vx) \cap \Gamma(Vy)} \frac{1}{\log |\Gamma(Vz)|} \quad (3)$$

The next step is to rank the node pairs according to their similarity score. To avoid the information overload issue as described above, we suggest limiting the number of nodes that we will take into account in the user's profile building process.

The considered relevant node pairs are computed by using the top-n elements of the rank. To start, we define the value of n = 75, based on the average of users' neighbors number in the experimentation of [1].

We obtain in the final step, for the user u, the user's egocentric network represented by the graph $G_2'(u) = (V_2',E_2')$, $|V_2'| = n$. We can then use this network in user's profile (social dimension) building process.

IV. EXPERIMENTATION

The objective of evaluation is to compare our approach resulting with one of CoBSP and ones of individual based algorithms in literary works:

- Individual based algorithm 1 (IBSP1): this approach computes the weight of an interest in the social dimension of the user by simply summing the semantic scores of this interest for each individual in the community.
- Individual based algorithm 2 (IBSP2) [4][5]: this approach use individual people (rather than communities) selected in the user's social network. Individual people are selected according to the strength of their tie with the user [5]or to their centrality values [4].

We conducted the experiment on co-authorship network. In this network nodes represent the authors. Two authors can be connected if they publish together. The interest of users from the titles of their publications was calculated. The collected data from this social network will be analyzed and integrated into the social dimension and user dimension.

The strategy of evaluation consists in looking among the algorithms CoBSP, IBSP1, IBSP2 and our approach, the one which allows to build a social dimension which is the closest to the real profile of the user (user dimension).

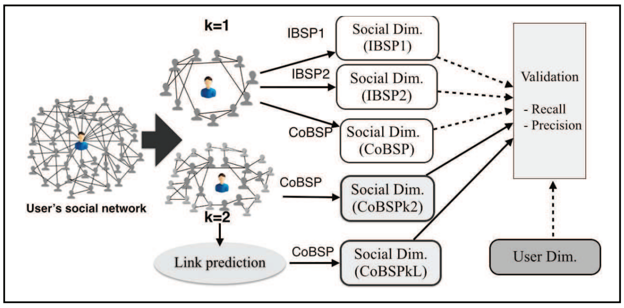

Fig. 2. Evaluation process

A. Experimentation setup

It is necessary to avoid using identical data sources in these two dimensions. In fact, the publications of an author could also exist in list of his co-authors. This could lead us to fault interpretation of the result. Thus, we have decided to adopt two distinct data sources to build the two dimensions (DBLP for the social dimension and Mendeley for the user dimension).

1) Data acquisition

We adopted in our experiment, the authors who exist in both DBLP and Mendeley and have a number sufficient of interest indicated in their Mendeley profile so that we can compare the results correctly. Subsequently, we have decided to take into account in our evaluation, the authors that have at least 6 interests explicitly indicated in their mendeley profile.

Following our contribution, the authors studied in our work have to possess the poor number of co-authors. The authors considered relevant to the experiment of [1] have at least 50 coauthors (ones who possess less co-authors could lead us to the lack of information issues as previously described). In our work, we considered the authors having less than 45 co-authors as the users possess a tiny and sparse network that we aim to improve the relevant of their dimension profile. Thus, an author corresponds to our data test set if he indicates more than 6 interests in his mendeley profile and if he possess less than 45 co-authors.

In order to study the factor of user's co-authors number, the authors studied in our test data are split into several groups according to the number of their co-authors. Our dataset is presented by several groups of authors: 50 authors with a very small network possessing less than 10 co-authors (as a community has to possess at least 2 individuals to apply the iLCD algorithm, we take into account authors having at least 3 coauthors in our work), 50 authors having 10-20 co-authors, 50 authors having 20-30 co-authors and 50 authors having 30-45 co-authors.

2) Building profiles process

We adopt in this work, the methodology process of building and evaluating authors' profiles from DBLP and Mendeley presented in [1].

a) Social dimension construction

The first stage consists in generate user's social dimension in the egocentric network with value of k=1. Interests are detected by mining texts that appear in the title of publications of communities or individuals according to the algorithm used to derive social dimension (CoBSP, IBSP1 and IBSP2).

To built social dimension of our approach, apart from the egocentric preparation step, we use the same social dimension building process as one of the existing work [1] by applying CoBSP algorithm to the prepared user egocentric network. We generate for our approach, 2 social dimensions that built after integrating distance-2 coauthors of the user. The first consider all distance-2 coauthors (COBSPk2). The second one takes in to account only the co-authors considered relevant after applying link prediction method (COBPkL).

b) User dimension construction

The user dimension in this experiment consists in representing the real interests of the user. The dimension is build by mining keywords in the list of interests he explicitly indicated in his Mendeley profile, using the same process of text mining as adopted in the social dimension construction.

c) Evaluation

To evaluate the relevance of each social dimension we use the precision and the recall measure. The precision and the recall consist in measuring the capacity of the system to compute and propose relevant items.

The precision represents the proportion of relevant founded items and the total number of items. In our experimentation context, the precision formula is presented as follow:

$$\frac{\text{Number of interest in the social dimension presented in the user dimension}}{\text{Total number of interest in the social dimension}} \quad (4)$$

The recall represents the proportion of relevant founded items compared to the total number of relevant items. It measures the capacity of the system to be restored. In our experimentation context, the precision formula is presented as follows:

$$\frac{\text{Number of interest in the social dimension presented in the user dimension}}{\text{Total number of interest in the user dimension}} \quad (5)$$

To compute the precision and the recall, we only consider the most relevant interests. The total number of interests in the user dimension top N(User's interests) + m firsts interests obtained after building the social dimension (m=5 in this experiment). For example, if the user dimension of an author's profile contains 10 interests, we will consider the social dimension as only the top 15 firsts interests computed in the social dimension.

V. RESULTS AND DISCUSSION

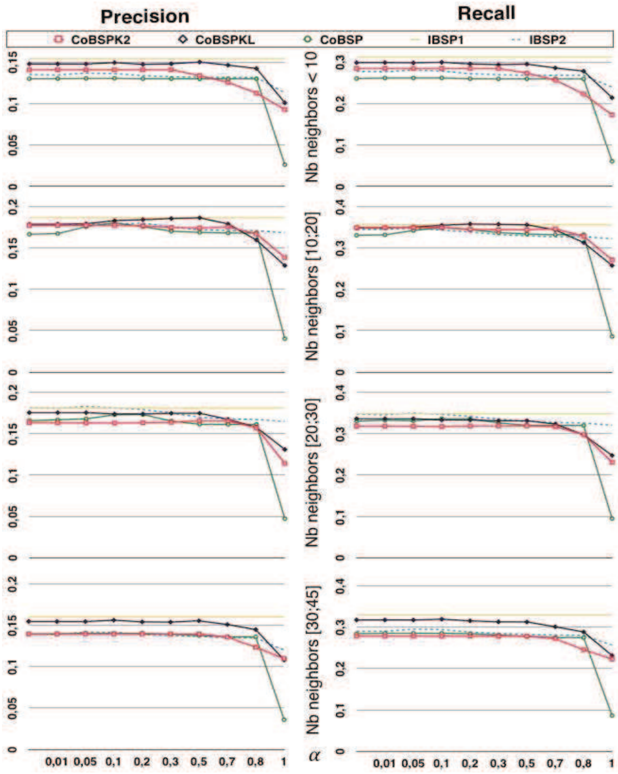

Fig. 3. Graphics comparing (precision, recall) the user dimension with the social dimensions built by algorithm CoBSP, IBSP1, IBSP2 and the social dimensions of our approach (CoBSPk2 and CoBSPkL)

The figure showed above presents comparative curves (precision, recall) of theses algorithms for all 200 authors' egocentric networks studied in this works, separated in different groups.

For the group of networks possessing few individuals especially with a number of co-authors lower than 10, CoBSPk2 outperforms CoBSP and IBSP2 in term of precision. This can be explained by the fact that in a very sparse network, there is not enough content to be mining in significant ways. After integrating more nodes into the network, we have more data source available to mining. This one demonstrates the interest of integrating more individual into user's egocentric network before applying CoBSP process, in the case of very tiny or sparse networks. However, compared to the algorithm based on the individuals IBSP1, CoBSPk2 produces less good results.

CoBSPk2 provides the worst accuracy when co-authors exceed 20. This can be explained by the fact that sometimes, by adding individuals to the distance k=2, the number of relations increases in exponential way and it could be possible that the new added nodes at such distance are not relevant (In the real world, it can be possible that the co-author of the studied author publish with another authors who are not at all in the same research field as this authors). More the author has existing co-authors more he risk to gain the non-relevant distance 2- coauthors.

After applying link prediction technique, we clearly see that CoBSPkL outperforms the CoBSP and CoBSPk2 in all groups of dataset. The link prediction technique does not provide better accuracy comparing to IBSP1. Nevertheless, the result could demonstrate the benefit of using link prediction to selection only relevant nodes into user's egocentric network before applying the user profiling process.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have presented the extension technique to improve the results of existing community-based algorithm (CoBSP) in small and sparse networks. We proposed the extension approach based on snowball sampling technique and link prediction. The empirical study proved the benefit of integrating more individuals (individuals at distance 2) into a very small/sparse network. To solve the problem remain existing in other groups of network having more than 10 neighbor nodes, we propose to apply link prediction that enabled us to provide the better performances compared to the existing work. This one also allows us to demonstrate the advantage of link prediction in social network mining for user modeling process.

Our short term perspective consist in adopting different link prediction algorithms and conducting the evaluation on other data sources such as Twitter data so that we can compare and select the prediction algorithms relevant to the type of social network. Another important perspective is to take into account the dynamic of social network in our analysis process. We could notice while working on the DBLP data that some scientists change their research field and/or their collaborations over time. The old research field should become less significant than the new one. When analyzing the interactions or information, it would be more effective if highest weight were assigned to the recent interaction and/or recent publications. The long term perspective consists in proposing a platform that extracts the information and designs the user social dimension according to the type and the specific characteristics of adopted social networks (e.g. spars, dense…) by taking into account their evolution.

REFERENCES

[1] D. Tchuente, M.-F. Canut, N. Jessel, A. Peninou, and F. Sèdes, "A community-based algorithm for deriving users' profiles from egocentrics networks: experiment on Facebook and DBLP," *Soc. Netw. Anal. Min.*, vol. 3, no. 3, pp. 667–683, Sep. 2013.

[2] D. Tchuente, M.-F. Canut, N. Baptiste-Jessel, A. Peninou, and F. Sedes, "A Community Based Algorithm for Deriving Users' Profiles from Egocentrics Networks," in Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), Washington, DC, USA, 2012, pp. 266–273.

[3] M. Mezghani, A. Péninou, C. Zayani, I. Amous, and F. Sèdes, "Detecting user's interests based on the accuracy of collaborative tagging information (short paper)," in *International Conference on Computer-Supported Cooperative Work (ECSCW), Paphos Cyprus, 21/09/2013-25/09/2013*, 2013, pp. 21–26.

[4] G. Cabanac, "Accuracy of Inter-researcher Similarity Measures Based on Topical and Social Clues," *Scientometrics*, vol. 87, no. 3, pp. 597–620, Jun. 2011.

[5] D. Carmel, N. Zwerdling, I. Guy, S. Ofek-Koifman, N. Har'el, I. Ronen, E. Uziel, S. Yogev, and S. Chernov, "Personalized Social Search Based on the User's Social Network," in *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, New York, NY, USA, 2009, pp. 1227–1236.

[6] P. Bonhard and M. A. Sasse, "'Knowing Me, Knowing You' – Using Profiles and Social Networking to Improve Recommender Systems," *BT Technol. J.*, vol. 24, no. 3, pp. 84–98, Jul. 2006.

[7] P. Massa and P. Avesani, "Trust-aware Recommender Systems," in *Proceedings of the 2007 ACM Conference on Recommender Systems*, New York, NY, USA, 2007, pp. 17–24.

[8] R. R. Sinha and K. Swearingen, "Comparing Recommendations Made by Online Systems and Friends.," in *DELOS Workshop: Personalisation and Recommender Systems in Digital Libraries*, 2002.

[9] P. Bhattacharyya, A. Garg, and S. F. Wu, "Analysis of user keyword similarity in online social networks," *Soc. Netw. Anal. Min.*, vol. 1, no. 3, pp. 143–158, Jul. 2011.

[10] R. Cazabet, M. Leguistin, and F. Amblard, "Automated Community Detection on Social Networks: Useful? Efficient? Asking the Users," in *Proceedings of the 4th International Workshop on Web Intelligence & Communities*, New York, NY, USA, 2012, pp. 6:1–6:8.

[11] L. A. Goodman, "Snowball Sampling," *Ann. Math. Stat.*, vol. 32, no. 1, pp. 148–170, Mar. 1961.

[12] M. S. Handcock and K. J. Gile, "On the Concept of Snowball Sampling," *ArXiv11080301 Stat*, Aug. 2011.

[13] P. Biernacki and D. Waldorf, "Snowball Sampling: Problems and Techniques of Chain Referral Sampling," *Sociol. Methods Res.*, vol. 10, no. 2, pp. 141–163, Nov. 1981.

[14] B. H. Erickson, "Some problems of inference from chain data," *Sociol. Methodol.*, vol. 10, no. 1, pp. 276–302, 1979.

[15] C.-N. Ziegler and J. Golbeck, "Investigating interactions of trust and interest similarity," *Decis. Support Syst.*, vol. 43, no. 2, pp. 460–475, Mar. 2007.

[16] M. Pavlov, "Finding Experts by Link Prediction in Co-authorship Networks," in *FEWS*, 2007, vol. 290, pp. 42–45.

[17] M. J. Rattigan and D. Jensen, "The Case for Anomalous Link Discovery," *SIGKDD Explor Newsl*, vol. 7, no. 2, pp. 41–47, Dec. 2005.

[18] V. Krebs, "Mapping Networks of Terrorist Cells," *CONNECTIONS*, vol. 24, no. 3, pp. 43–52, 2002.

[19] R. N. Lichtenwalter, J. T. Lussier, and N. V. Chawla, "New Perspectives and Methods in Link Prediction," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2010, pp. 243–252.

[20] L. Munasinghe and R. Ichise, "Link Prediction in Social Networks Using Information Flow via Active Links," *IEICE Trans. Inf. Syst.*, vol. E96-D, no. 7, pp. 1495–1502, Jul. 2013.

[21] B. Bringmann, M. Berlingerio, F. Bonchi, and A. Gionis, "Learning and Predicting the Evolution of Social Networks," *IEEE Intell. Syst.*, vol. 25, no. 4, pp. 26–35, Jul. 2010.

[22] M. A. Hasan, V. Chaoji, S. Salem, and M. Zaki, "Link prediction using supervised learning," in *In Proc. of SDM 06 workshop on Link Analysis, Counterterrorism and Security*, 2006.

[23] N. Benchettara, R. Kanawati, and C. Rouveirol, "Supervised Machine Learning Applied to Link Prediction in Bipartite Social Networks," in *2010 International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2010, pp. 326–330.

[24] T. Wohlfarth and R. Ichise, "Semantic and Event-Based Approach for Link Prediction," in *Practical Aspects of Knowledge Management*, T. Yamaguchi, Ed. Springer Berlin Heidelberg, 2008, pp. 50–61.

[25] D. Liben-Nowell and J. Kleinberg, "The Link Prediction Problem for Social Networks," in *Proceedings of the Twelfth International Conference on Information and Knowledge Management*, New York, NY, USA, 2003, pp. 556–559.

[26] L. A. Adamic and E. Adar, "Friends and neighbors on the Web," *Soc. Netw.*, vol. 25, no. 3, pp. 211–230, Jul. 2003.

[27] M. E. J. Newman, "Clustering and preferential attachment in growing networks," *Phys. Rev. E*, vol. 64, no. 2, Jul. 2001