TOPIC EXTENSION USING THE NETWORK EXTRACTED FROM DBLP

Alisa Babskova[(a)], Pavla Draždilová[(b)], Jan Martinovič[(c)], Kateřina Slaninová[(d)]

[(a,b,c)]VŠB - Technical University of Ostrava,
Department of Computer Science,
17. listopadu 15/2172, 708 33 Ostrava, Czech Republic
[(d)]Silesian University in Opava,
SBA in Karvina, Department of Informatics,
Univerzitní nám. 1934/3, 733 40 Karviná, Czech Republic

[(a)]alisa.babskova@vsb.cz, [(b)]pavla.drazdilova@vsb.cz, [(c)]jan.martinovic@vsb.cz, [(d)]slaninova@opf.slu.cz

ABSTRACT

This article focuses on the topic extension in an area that is initially specified by the user through the topic's keywords. The extended area of interest defined by keywords is determined by a set of terms used by the community for which the selected keywords are significant. The extracted topic by selected communities can be used to update and broaden the area of interest. This new evaluation of edges depends on terms that appear in the titles of articles of two co-authors. The newly evaluated network more accurately describes the intensity of the relationships between co-authors. This network is suitable as an input to models, which are focused on prediction of future relationships and community structures in co-author networks. Moreover, the topic extension may be used in prediction models for the extraction of expected keywords which will be used in a given community.

Keywords: e-learning, topic extraction, DBLP, subnet-works, community detection, study materials

1. INTRODUCTION

In the field of e-learning, we come across a variety of electronic learning materials that allow students to familiarize themselves with the chosen topic. There are the syllabi, lecture notes, presentations and other educational texts available and they create a collection of documents.

These materials are mostly created by educators for the selected theme and provide a comprehensive overview of selected topics and areas. Our intention is not to provide comprehensive and pre-processed materials, but to offer students the opportunity to familiarize themselves with a self-selected and an interesting topic for them in the field of computer science. A student specifies the area in which she/he is interested in (topic) with one or more terms. Our approach finds the most important community described by the specified topic (for the selected keywords) in the co-authors' network DBLP (http://dblp.uni-trier.de/). These selected keywords are often used in headlines of articles, or are common to most members of the community, or the most commonly occur in posts on the blog. Authors who have written many articles that relate to the selected terms can be determined for this community. Thereafter, we are able to find other keywords from the document collection, which can not only describe the extension of the defined set of keywords for the given topic, but also can be significant for other topics, which are on the interest of the given community of co-authors. These new words may be useful for further selection of other keywords, and for further selection of other documents, which will lead to the extension of the initial topic. Some new keywords obtained by this way can be out of the initial topic, and can therefore extend the whole scope of users' interest (not only extend the narrow topic initialised at the beginning).

Another possibility is to choose articles from the collection of documents that were written in specified time period (e.g. the latest articles in the field). These articles can provide the newest information and additional terms, which may refine or expand students' topic and thus they can be independently and proactively involved in their education process.

This type of study materials searching is suitable for doctoral students or for undergraduate students in computer science, because our system works with large database of articles focused on this sphere - DBLP. Generally, this kind of study material searching can be used in any other area with documents that have titles or are briefly described (e.g. by abstracts). In the resulting network with newly evaluated edges, there are found communities of co-authors, who published together and who have strong relations to the selected terms. Additionally, we can choose the authors who have the most publications in the topic and whose work can be beneficial and inspiring for the students.

In the publication (Monachesi, Lemnitzer and Simov 2006) is presented The European project Language Technology for eLearning (LT4eL), which aim is to improve the efficiency and the availability of the static and dynamic content created for eLearning using Learning Management Systems (LMS). This problem was solved using language technologies based on the functionality and the integration of semantic knowledge, and might facilitate the management, distribution and retrieval of study materials.

Other possible approach to gaining the extended topics is usage of time information. The authors of publication (Chen, Luesukprasert and Chou 2007) deal with retrieval of actual topics from various collections of text documents published in a given time period. The presented method consists of two steps. The first step is focused on the extraction of the actual terms and on mapping of their distribution through the given time period. The second step consists of the identification of key sentences (based on the extracted actual terms), which are clustered. The clusters then represent the actual topics defined by multidimensional vector of sentences.

The article (Schirru, Baumann, Memmel, and Dengel 2010) is focused on the automatic identification of various topics, which are a scope of interest of source sharing platforms users.

The authors of article (Sun, Barber, Gupta, Aggarwal and Han 2011) focused on the prediction of the future relations between the co-authors in the heterogeneous bibliographic network (DBLP) using the heterogeneous topological characteristics. The community evolution is a scope of interest in (Brodka, Saganowski and Kazienko 2011), in which is presented a method Group Evolution Discovery (GED). The method uses not only the size and comparison of the group members, but also takes into consideration their significance and position inside the group, to find the progress of the group in the sequential time periods. In (Patil, Liu and Gao 2013) is presented the groups evolution and their stability. The analysis (for example of DBLP) showed that it is possible to predict the group stability with the high accuracy using various attributes which describe the group composition, the activities inside the groups and the group structural aspects.

Our developed network is suitable as the input for the models, which are focused on the prediction of the future relations and the community structures in co-author networks. Moreover, the topic extension can be usable for the extraction of the future keywords used within a given community.

The rest of this paper is structured as follows: Section 2 describes the related work in the social and co-author networks, and wikis and blogs as sources of documents. In Section 3, our proposed approach is presented. We depict our idea of relations between persons on the basis of term context and describe how the ContextScore as a new edge evaluation in the network can be obtained. Then, in Section 4, we present the comparison of co-authors' network with and without term context. Our method has been tested and a topic extension is presented in Section 5. In Section 6, we summarize our findings and present ideas for the future work.

2. DBLP AND OTHER SOURCES

DBLP (Digital Bibliography Library Project) is a computer science bibliography database hosted at University of Trier, in Germany. It was started at the end of 1993 and listed more than 2.1 million publications in January 2013. These articles were published in Journals such as VLDB, the IEEE and the ACM Transactions and Conference proceedings. DBLP has been a credible resource for finding publications, its dataset has been widely investigated in a number of studies related to data mining and social networks to solve different tasks such as recommender systems, experts finding, name ambiguity, etc. Even though, DBLP dataset provides abundant information about author relationships, conferences, and scientific communities. It has a major limitation that its records provide only the paper title without the abstract and index terms.

Wiki, blogs and different sources of information are useable to create a new edge evaluation of network, in which nodes are persons created page in wiki or blog and edges represent cooperation in project.

Many experts focused on the task of finding persons with the high level of experience in a specific topic. To achieve this objective, researchers approached this task mainly in three different ways. The first group applied information retrieval techniques to solve the mentioned problem (Deng, King, Lyu 2008).- The authors of this paper proposed a weighted language model, which introduced a document prior probability to measure the importance of the document written by an expert. The second group approached this task using social network analysis metrics (Zhang, Ackerman, Adamic 2007). In this study, the Java Forum, a large online help seeking community, was analysed using social network analysis methods and a set of network-based algorithms including PageRank and HITS. The third group used a hybrid approach of information retrieval and social network analysis for finding academic experts (Zhang, Tang, Li 2007). In (Zhang, Tang, Li 2007), the authors created a local information document for each person to measure his initial level of experience on a topic using information retrieval models. Then they applied propagation on the graph of experts to update his level of expertise according to his relations with the other nodes. In the article (Drazdilova, Martinovic, Slaninova 2013), the authors focused on the detection of communities using spectral clustering. This algorithm was used in the article (Minks, Martinovic, Drazdilova, Slaninova 2011) to find the communities in subnetworks that were defined by the selected terms (from the whole DBLP).

Wiki, blogs and different sources of information are useable to create a new edge evaluation of network, in which nodes are persons created page in wiki or blog and edges represent co-operation in project.

In (Yang and Ng 2008), the authors proposed an analytical system of web forum for the analysis of the content development and for the visualisation of the social relations in web forums. Our approach creates a new evaluation of network edges (relations) and head towards the usage of the documents from DBLP, wikis or blogs to extract and develop the initial topics.

The presented approach demonstrated for DBLP can be used also for other mentioned document resources, in which the relation between the persons (co-authors) is created on the basis of the common documents and its evaluation is dependent on the extracted terms from the document titles. Table 1 presents various document resources and other relevant information. Even from such different document resources, we are able to create the co-authors' network, who participated on the creation of common documents. Therefore, as well as in DBLP, the evaluation of the relations is dependent on the extracted terms from the documents.

Source	Documents with terms	Persons	Time
DBLP	Title of Paper	Authors	Year of Publishing
Wiki	Wiki page	Editors	Last Edited
Blog	Post	Bloggers	Last Sending
Codeplex	Project Description	Developers	Last Activity

Table 1: Mapping Different Information Sources to Person and Term Context

3. RELATIONS BETWEEN PERSONS ON THE BASIS OF TERM CONTEXT

In the paper, we propose a more precise evaluation of the intensity of person's relations to ascertain the context among persons (e.g. authors, editors, bloggers, developers) and the terminology they used in documents (for example terms in article titles in DBLP, terms in Wiki pages and blogs or terms in projects description).

Wang, Mccallum and Wei (2007) present topical n-grams, a topic model that discovers topics as well as topical phrases. Another area that utilizes text information is finding of expert in DBLP bibliography data (Deng, King and Lyu 2008), or the analysis of communities based on DBLP (Biryukov and Dong 2010).

In our approach, we use terms for the evaluation of the relation between persons. We extend a standard evaluation of the relation, which is based on the number of the common articles, by a factor that represents a context between persons and term selected from the term set.

Term set is understood as a collection of all keywords, which are extracted from the document. As the source of terms were used titles of articles from the DBLP dataset. A more detailed description of the term set was presented in article (Minks, Martinovic, Drazdilova, and Slaninova 2011).

3.1. Relations between Persons

Besides the computation of evaluated term set, we can compute *association strength* between the two persons. This method is not only interesting by itself, but it is also essential for extended evaluation of the term list by selected context.

Relevancy between persons is based on the participation on the same document. This relevancy is then approximated by Jaccard coefficient (Deza and Deza 2006).

Let A be a set of all persons in dataset. We define a single person A_i. For A_i, it is evaluated the strength of association with the other persons (co-participants).

The set of co-participants of person A_i is marked as C_{A_i}. Let set P be a set of all documents (papers) and P_{A_i} be a set of all documents of person A_i.

The *association strength* between the persons A_i and A_j can be defined with Jaccard coefficient that reflects mainly the proximity of both persons from number of their common documents:

$$Q(A_i, A_j) = \frac{|P_{A_i} \cap P_{A_j}|}{|P_{A_j}| + |P_{A_i}| - |P_{A_i} \cap P_{A_j}|}. \quad (1)$$

If this method is applied to all the persons, we obtain weighted undirected graph that can be considered as a synthetic social network (with re-weighted edges between persons). This approach was inspired by (Ding 2011).

3.2. Persons and the Term Context

If we define a set T as the set of all terms in all documents and T_{A_i} as a set of all the terms that could be found in the documents of person A_i, then t_k is the term belonging to the person A_i ($t_k \in T_{A_i}$).

Thus, we define ($t_k\ in\ T_{A_i}$) as a number of occurrences of the term t_k in the documents of person A_i. Then, this number is divided by the number of occurrences of term t_k in the all project's description(t_k in T). The higher value, the less relevant term t_k becomes. In addition, a number of terms of the author A_i($|T_{A_i}|$) is added to the number of occurrences of the term t_k, because there is an assumption that T_{A_i}, which has a high cardinality, lower the importance of the individual terms, while low cardinality indicates that the author has only one subject matter. Then, we can define the *relevance of author's terms* as:

$$R(T_{A_i}, t_k) = \frac{(t_k\ in\ T_{A_i})}{(t_k\ in\ T) + |T_{A_i}| - (t_k\ in\ T_{A_i})}, \quad (2)$$

and in normalized form as:

$$R_{Norm}(T_{A_i}, t_k) = \frac{R(T_{A_i}, t_k)}{MAX(R(T_{A_i}, t_1), \ldots, R(T_{A_i}, t_{|T_{A_i}|}))}. \quad (3)$$

Because we have defined the relation between the persons and we can express the relevance of the person's terms, we can assign the best suitable co-participant to a given term. We can demonstrate the usage of significance of each co-participant as well. Our reflections were inspired by associative memory, where one is able to better recall the event, which is associated with something significant (although it was already forgotten). For a given person, it is significant the term, which associates him the best co-participant in the selected topic.

The method extension, including the person's co-participant as a context, is constructed analogically. The context is calculated for a given person according to the Formula (1). Afterwards, the persons are selected from the evaluated list of co-participants.

The $ContextScore_k$ for selected term t_k is calculated by the equation:

$$ContextScore_k(A_i, A_j, t_k) = R_{Norm}(T_{A_i}, t_k)\ R_{Norm}(T_{A_j}, t_k)\ Q(P_{A_i}, P_{A_j}) \quad (4)$$

The overall *ContextScore* between persons is given by the sum of particular $ContextScore_k$ related to the particular terms, which create the query.

$$ContextScore(A_i, A_j) = \sum_{t_k \in query} ContextScore_k(A_i, A_j, t_k) \quad (5)$$

4. COMPARISON OF CO-AUTHORS NETWORK WITH AND WITHOUT TERM CONTEXT

The selected experiments focused on the comparison of the initial edge evaluation, which represents the amount of common publications of the two authors and the resulting node evaluation in 2010 within DBLP, and the new proposed one, which takes into consideration the selected terms. By this way, we obtain two different evaluations of the authors, which allow us to sort them and to select the most important authors, which are supposed to be described by other interesting terms, which extend a given topic.

In the experiments, the set of terms was selected, which represented a simplified topic. Therefore, these input terms defined a topic of our interest towards which we will evaluate the authors in DBLP. Using these terms we have selected from the complete graph of DBLP such subgraphs, in which the relations between the persons were based on term context. Of course, such subgraphs did not contain all the initial authors from the DBLP collection, but only these who had the required terms which defined the topic in the publication title (Babskova, Drazdilova, Martinovic, Svaton, and Snasel 2013).

During the experiment we start the creation of author subgraph set for the selected year 2010 based on input terms. We have chosen two types of queries for the experiment of which the second one specifies our area of interest. The queries are “social” and “social network”. Two different evaluations of the intensity of relationships between co-authors have been used to compare the results. The first subgraph features original evaluation of the edges representing the amount of joint publications of two authors in a given year and in a given area (labelled as ‘without term context’ in the text) and the second evaluation is ContextScore (see Formula 5) and it is labelled as “with term context” in the text.

As the next step in our approach, we have calculated the weighted degree (Newman 2004) for both methods of edge weight recalculation. Following that and based on these different evaluations, we have created author lists sorted according to relevant weighted degrees that demonstrate different significance of an author depending on the used edge evaluation. Information on the subgraphs retrieved for the entered terms of “social” and “social network” are in the Table 2.

Set of terms	Nodes	Edges	Components
Social	5599	8914	1537
Social network	2418	3715	669

Table 2: Subgraphs for Terms ‘social’ and ‘social network’

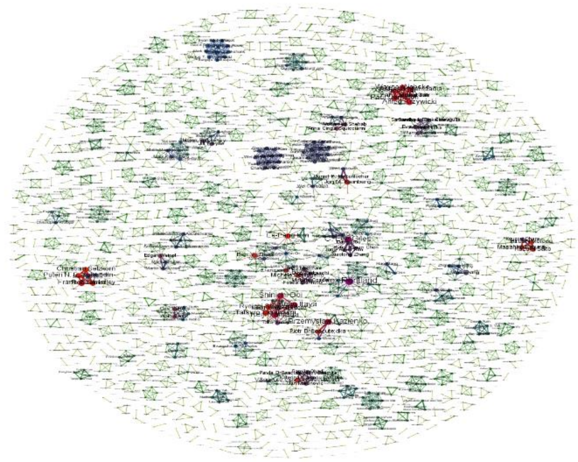

Figure 1: Subgraph (DBLP) for Term ‘social network’ – Weighted, without Term Context

Before starting to analyse the results, it is important to say that if we consider activity of an author in specific subgraph, then we always consider only the activity of an author in specific area which is specified by the input terms. The whole subgraph determined by the term “social” is shown in the Figure 1. Intensity of the relationships is determined by the original edge evaluation – i.e. by the number of joint publications in 2010. Re-evaluation of the edges by means of the term context resulted in different authors´ degrees in the subgraphs (see Figure 2).

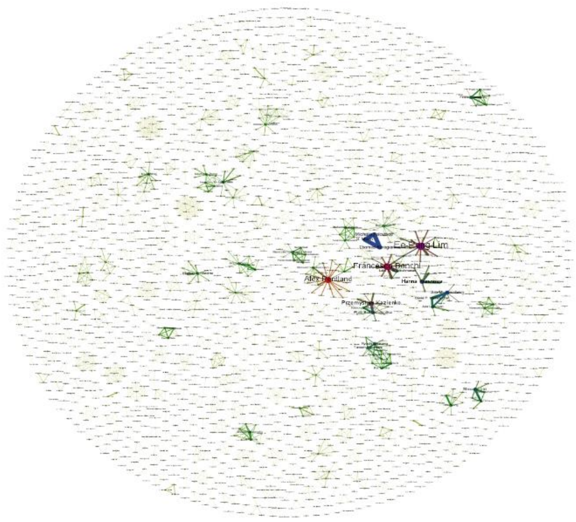

Figure 2: Subgraph (DBLP) for Term ‘social network’ – Weighted, with Term Context.

We focus on each subgraph separately in the next part of the experiment. We have calculated the weighted degrees for all the authors in the subgraph for the community with the edge evaluation both with and without term context. The examples of the authors´ weighted degree diagrams, always for both evaluation methods, are shown in Figure 3, Figure 4, Figure 5, and Figure 6.

The diagrams shown in Figure 3, Figure 4, Figure 5, and Figure 6 make it clear that the new edge evaluation has changed the node degree distribution which resulted in lower number of authors with the highest new weighted degree. That enables us to constrict the set of authors in a given area that are significant from our point of view.

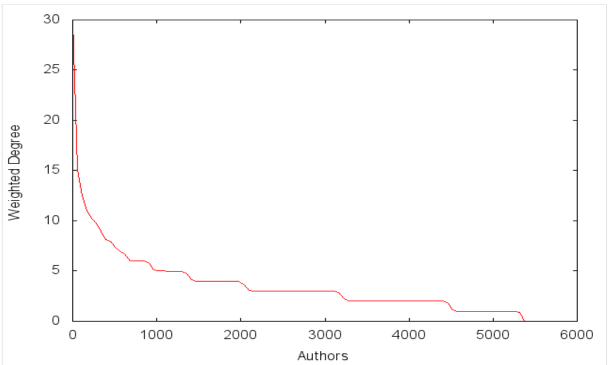

Figure 3: Histogram of Sorted Weighted Degree of Subgraph for Term ‘social’ - without Term Context

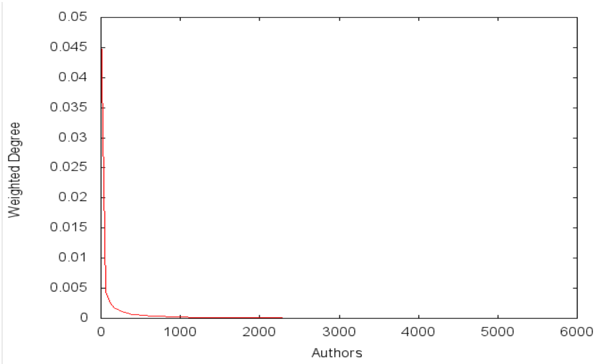

Figure 4: Histogram of Sorted Weighted Degree of Subgraph for Term ‘social’ - with Term Context

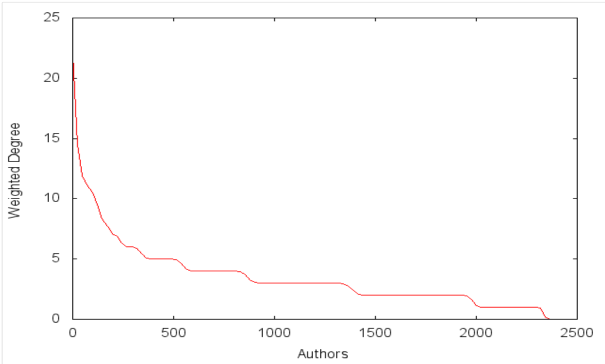

Figure 5: Histogram of Sorted Weighted Degree of Subgraph for Terms ‘social network’ - without Term Context

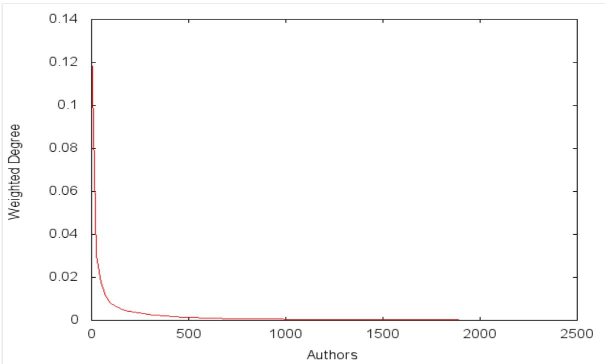

Figure 6: Histogram of Sorted Weighted Degree of Subgraph for Terms ‘social network’ - with Term Context

Further we spot the authors with relatively high degree when compared to the other authors in a given subgraph. We will find Top 10 authors with the highest degree value. Table 3 and Table 4 show Top 10 authors for the ‘social network’ and ‘social’ subgraph. These tables demonstrate that the Top 10 authors with the highest degrees, calculated on the basis of the original evaluation, mostly did not appear in Top 10 of authors with the degree calculated with term context. Only the very active or the significantly publishing authors occur in both subgraph without term context and subgraph with term context.

Subgraph, Weighted without Term Context		Subgraph, Weighted with Term Context	
ID of Author	Name of Author	ID of Author	Name of Author
49669	Alex Pentland	56679	Ee-Peng Lim
57581	C. Lee Giles	57835	Francesco Bonchi
28700	Przemyslaw Ka-zienko	49669	Alex Pentland
690549	Satoko Itaya	28700	Przemyslaw Ka-zienko
119660	Shinichi Doi	260831	Hanna Krasnova
252705	Keiji Yamada	225725	Thomas Karagi-annis
735639	Xiongcai Cai	181414	S, Moon
698479	Alfred Krzywicki	96113	Michalis Falout-sos
357960	Wayne Wobcke	38231	Jon M, Klein-berg
333282	Yang Sok Kim	57593	Shou-De Lin

Table 3: Top 10 of Authors of Subgraph 'social network' without Term Context and with Term Context

Top 10 comprises the authors with the highest degree values in the subgraph. These authors have been detected as the most active in the area defined by the input terms. It is possible to retrieve other terms with frequent occurrence in the publications of each author. As we focus on the most active authors, there is high probability that the retrieved terms by this way describe the topic defined by the input terms set in a greater depth or even expand the pre-defined topic with other areas of interest being currently explored.

Figure 7 and Figure 8 are shown parts of subgraphs, which present significance of particular authors for communities of co-authors. These graphs provide information not only about the activity of the significant members, but about the whole community.

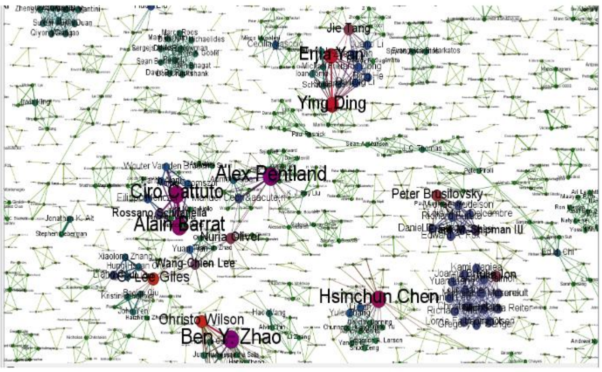

Figure 7: Part of Subgraph for Term 'social' – Weighted, without Term Context

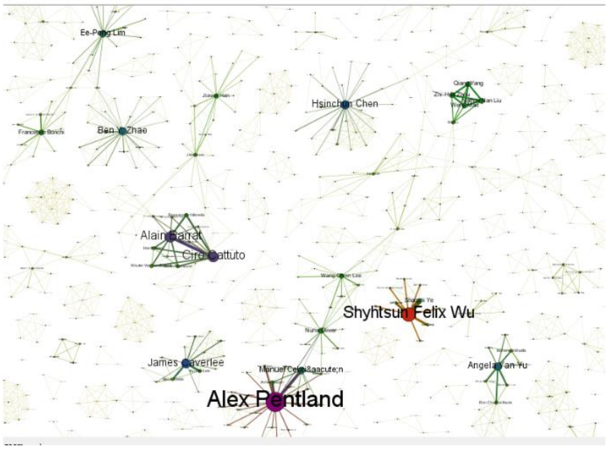

Figure 8: Part of Subgraph for Term 'social' – Weighted, with Term Context

Subgraph, Weighted with-out Term Context		Subgraph, Weighted with Term Context	
ID of Author	Name of Author	ID of Au-thor	Name of Author
49669	Alex Pentland	49669	Alex Pentland
111861	Ben Y. Zhao	120967	Shyhtsun Felix Wu
182741	Alain Barrat	182741	Alain Barrat
360307	Ciro Cattuto	360307	Ciro Cattuto
46738	Hsinchun Chen	57006	James Caverlee
45474	Ying Ding	46738	Hsinchun Chen
45635	Erjia Yan	170122	Angela Yan Yu
57581	C. Lee Giles	111861	Ben Y. Zhao
543984	Christo Wilson	57593	Shou-De Lin
57006	James Caverlee	252892	Munmun De Choudhury

Table 4: Top 10 of Authors of Subgraph 'social' without Term Context and with Term Context

Table 55 and Table 66 for 'social' and 'social network' subgraphs show other frequently occurring terms for Top 10 users, retrieved by means of weighted degree without and with term context. These terms allowed us other possibilities, to which we can concern. For example, term 'social' has the most frequently occurred other term 'network'. Other extended terms are 'signal', 'online', 'movement', 'ontology', 'spam', 'socialtrust', etc.

Subgraph, Weighted without Term Context		Subgraph, Weighted with Term Context	
ID of Au-thor	Other Terms	ID of Author	Other Terms
49669	social;network; interaction;signal; processing;sensing	49669	social;networks; interation;signal; processing;sensing
111861	social;graph;online; networks;detecting; characterizing	120967	social;online; networks;systems; estimating
182741 360307	social;link; live; semantics;creation; profile;network	182741 360307	social;link; live; semantics;creation; profile;network
46738	social;movement; network;cyber; research;web	57006	social;socialtrust; spammers;informa-tion;communities
45474	social;tagging;tag; ontology;integrating	46738	social;movement; network;cyber;

45635		170122	social;learning; online;networking;knowledge
57581	social;network; ranking;document	111861	social;graph;online; networks;detecting; characterizing
543984	social;detecting; spam;campaigns	57593	social;information; network;egocentric
57006	social;socialtrust; spammers;information;communities	252892	social;information; media; communication

Table 5: Other Terms for Top 10 of Authors of Subgraph 'social'

Subgraph weighted without term context		Subgraph weighted with term context	
ID of author	**Other terms**	**ID of author**	**Other terms**
49669	sensor; time; face; composite; predicting; mobile; apps; installation; dynamics	56679	modeling; mining; dynamic; spatio; temporal; discovery; link; formation; visualizing; semantic
57581	based; analysis; ranking; temporal; scientific; document; sndocrank; video; snakdd; mining; report	57835	influence; mining; analysis; learning; propagation; data; perspective
28700	multi; layered; based; analysis; extraction; group; evolution; online; label; dependent; feature; node; method; complex; discovery; email	49669	sensor; time; face; composite; predicting; mobile; apps; installation
690549	incentive; rewarding; services; analysis; communication	28700	multi; layered; based; analysis; extraction; group; evolution; online; label; dependent; feature; node; method; complex; discovery; email
119660		260831	privacy; sites; online; calculus; germany; empirical; usa; trust; study
252705		96113	online
735639	people; recommendation; collaborative; filtering	225725	
698479		181414	
357960		38231	predicting; problem; directed; closure; process; hybrid; analysis
333282		57593	

Table 6: Other Terms for Top 10 of Authors of Subgraph 'social network'

5. DISCUSSION AND CONCLUSION

Our experiments proved that there is benefit in creating subgraphs based on the context, specified by a topic (term set). It is also obvious that by using the context, we retrieve other terms to expand the topic which enables new ways of searching for other articles or capable authors.

In our future work, we want to implement the outlined approach into the recommended system described by the Figure 9 in order to make the processes as automated as possible. Let's assume that the results presented in this article can be further implemented into the modelling instruments to provide them with the graph evaluated in a different way based on which they can make predictions.

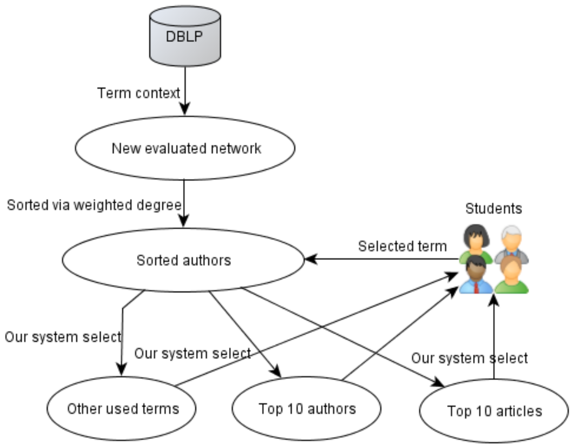

Figure 9: Description of Proposed Approach

ACKNOWLEDGMENTS

This work was supported by the SGS, VŠB–Technical University of Ostrava, Czech Republic, under the grant No. SP2013/167 'Analysis of Users' Behaviour in Complex Networks' and by grant of Silesian University no. SGS/6/2013 'Advanced Modeling and Simulation of Economic Systems'.

REFERENCES

Babskova, A., Drazdilova, P., Martinovic, J., Svaton, V., Snasel, V., 2013. Evolution of co-authors communities formed by terms on DBLP. In *Proceedings of the Databases, Texts, Specifications and Objects 2013,* pp. 109-118.

Biryukov, M. and Dong, C., 2010. Analysis of computer science communities based on DBLP. *Research and Advanced Technology for Digital Libraries*, pp. 6273-6279.

Brodka, P., Saganowski, S., Kazienko, P., 2011. Group Evolution Discovery in Social Networks. *2011 International Conference on Advances in Social Networks Analysis and Mining*, pp.247-253.

Chen, K.-Y., Luesukprasert, L., Chou, S.-C.T., 2007. Hot Topic Extraction Based on Timeline Analysis and Multidimensional Sentence Modeling. *IEEE Transactions on Knowledge and Data Engineering*, 19(8), pp.1016-1025.

Deng, H., King, I., Lyu, M.R., 2008. Formal models for expert finding on DBLP bibliography data. *Eighth IEEE International Conference on Data Mining*, pp. 163–172.

Deza, M. M. and Deza, E., 2006. Dictionary of Distances. Elsevier Science, Amsterdam, The Netherlands.

Ding, Y., 2011. Scientific collaboration and endorsement: Network analysis of co-authorship and citation networks. *Journal of informetrics*, pp. 187-203.

Drazdilova, P., Martinovic, J., Slaninova, K., 2013. Spectral Clustering: Left-Right-Oscillate algorithm for detecting communities. In *Lecture Notes in Computer Science: Computer Information Systems and Industrial Management,* pp. 278-289.

Minks, S., Martinovic, J., Drazdilova, P., Slaninova, K., 2011. Author cooperation based on terms of article titles from DBLP. *Advances in Intelligent Systems and Computing*, pp. 281–290.

Monachesi, P., Lemnitzer, L., Simov, K., 2006. Language technology for eLearaing. In *Lecture Notes in Computer Science including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*. pp. 667-672.

Newman, M.E.J., 2004. Analysis of weighted networks. *Physical Review E*, 70(5), p.9.

Patil, A., Liu, J., Gao, J., 2013. Predicting group stability in online social networks. In *Proceedings of the 22nd international conference on World Wide Web* (WWW '13). International World Wide Web Conferences Steering Committee, pp. 1021-1030.

Schirru, R., Baumann, S., Memmel, M., Dengel, A., 2010. Extraction of Contextualized User Interest Profiles in Social Sharing Platforms. *Journal Of Universal Computer Science*, pp.2196-2213.

Sun, Y., Barber, R., Gupta, M., Aggarwal,Ch. C., Han, J., 2011. Co-author Relationship Prediction in Heterogeneous Bibliographic Networks. *2011 International Conference on Advances in Social Networks Analysis and Mining*, pp.121-128.

Wang, X., Mccallum, A., Wei, X., 2007. Topical n-grams: Phrase and topic discovery, with an application to information retrieval. *Seventh IEEE International Conference on Data Mining ICDM 2007*.

Yang, C.C. and Ng, T.D., 2008. Analyzing content development and visualizing social interactions in web forum. In *IEEE International Conference on Intelligence and Security Informatics 2008 IEEE ISI 2008*. pp. 25-30.

Zhang, J., Ackerman, M.S., Adamic, L., 2007. Expertise Networks in Online Communities: Structure and Algorithms. W. W. W. C. Committee, ed. *Forum American Bar Association*, pp. 221-230.

Zhang, J., Tang, J., Li, J., 2007. Expert Finding in a Social Network. *Network*, LNCS 4443, pp.1066-1069.