Visualizing Opportunities of Collaboration in Large Research Organizations

Mohammad Amin Yazdi[1], André Calero Valdez[2], Leonhard Lichtschlag[3], Martina Ziefle[2], and Jan Borchers[3]

[1]RWTH Aachen University, Germany
{amin.yazdi}@rwth-aachen.de

[2]Human-Computer Interaction Center, RWTH Aachen University, Germany
{calero-valdez,ziefle}@comm.rwth-aachen.de

[3]Media Computing Group, RWTH Aachen University, Germany
{lichtschlag,borchers}@cs.rwth-aachen.de

Abstract. In order to support interdisciplinary collaboration in a large organization, providing opportunities to meet new collaborators is essential. Besides offline approaches (e.g., conferences, colloquia, etc.) data driven and online approaches can be considered. Using the publication data and the additional profile information of researchers on a scientific portal, we try to support the process of uncovering opportunities for collaboration. For this purpose we develop a visualization that focuses on revealing potential co-authors that are a good fit according to track-record and profile information. In a design study we present the result of an iterative user-centered design process – a novel prototype and its evaluation. Overall, our visualization was able to inform researchers about valid collaboration opportunities while at the same time effectively conveying organizational information. Our prototype showed a high usability and loyalty score (SUS=82.5, NPS=40).

Keywords: Design study, Interdisciplinarity, Visualization Collaboration, Recommender System.

1 Introduction

Interdisciplinary collaboration is considered both boon and bane of scientific advancement in recent years. Funding organizations like the NSF have shifted capacities to interdisciplinary research efforts [1]. Interdisciplinary research is considered to be an effective solution for large scale complex problems overarching the limits of disciplinary boundaries. In spite of its promises, interdisciplinary teams face several challenges in their collaboration [2]. Differences between disciplinary cultures (e.g., language, methodology, scientific performance evaluation) and individuals, in combination with shorter project run-times, inhibit effective collaboration, which requires a mutual understanding of the topics and the team

itself [3]. The more experienced researchers are in interdisciplinary research, the more successfully they collaborate[4].

Larger research clusters (over 100 researchers) are part of the German strategy for scientific excellence (forty funded research clusters in Germany. Whether these clusters surpass simple smaller research projects heavily depends on the effort to interlink researchers within such a cluster. In order to address the staff volatility and sheer size of such a research cluster, as one measure we devised the "Scientific Cooperation Portal" (SCP). [5]. The SCP is a web-based social portal that serves as a means to centralize communication, file-exchange, member profiles, and offers interdisciplinary collaboration support and output tracking of the individual researchers. One part of the SCP is track-keeping of publications generated in the cluster to enable steering. The publications of the researchers are visualized to assist both the researchers themselves as well as the cluster administration to assess the interdisciplinary collaboration. [6]. In this paper we use this data to construct visualizations that help facilitate collaboration.

2   Related work

In order to understand how effort (i.e., money) is spent effectively some form of performance evaluation is necessary. For this purpose bibliometric methods are used (often with a smattering of knowledge) to evaluate performance of individual researchers. Certain criteria can be measured relatively directly from publication data. Citation data is often used to evaluate institutions but is badly suited for automated researcher evaluation due to problems like insufficient database coverage, citation lag, disciplinary differences and bad interpretability [7].

Co-authorship analysis [8] reveals who has published with whom, and thus collaborated successfully (in the widest sense of the word). It is also used to identify who could collaborate on what topics [9] and when analyzing the content of communally published documents. Using text-mining approaches like document clustering enables identifying topics and relevant keywords [10]. Both co-authorship analyses [11] and document clustering approaches have been used to visualize the status quo, but not in the scope of recommending possible collaborators. Wu et al. [12] even visualized the change in research topic per individual researcher over the path of their careers.

Yu et al. [13] have developed a system to find collaborators in the PubMed database using a controlled vocabulary for the medical sciences (UMLS) and evaluated its usability with 26 experts. However, suggestions of collaborators were not based on prior collaboration but only on shared research interests. Chaiwanarom et al. [14] proposed a method for finding collaborators within the author's co-author networks and based on keyword similarity. Using a prediction test their method could find approx. 89% of all actual collaborators. Suggestions were then shown as a list.

Visualizing suggestions for collaborators has not been attempted to our knowledge. Ehrlich et al. [15] propose such a solution, but (also) rely on analyzing email content to find collaborators. This approach is quite unthinkable

in a research cluster of independent research groups in a German cultural background that values data privacy highly. Loep et al. [16] presented a recommendation system for movies based on previous choices and showed its superiority over manual search in lists. Visualizing recommendations increased trust in them and revealed sufficiently novel information. Suggesting collaborators goes beyond a simple expert search [17] attempted by using social network analysis methods such as HITS. It requires finding a person willing to collaborate, thus sharing similar work ethics, procedures and methods.

When analyzing co-author relationships for reasons of their successful collaboration two types of relationships are dominant. Successful researchers are either similar ("birds of a feather flock together") in their co-authorship network and publication output or complementary ("opposites attract") [18, 19]. In general inferring interests from social relationships can be very successful when done adequately [20].

Scientific social networks and analytic sites like ResearchGate, Academia.edu, ArnetMiner, ResearcherId, etc. address understanding researcher profiles. ResearchGate and Academia.edu are Social Networking Sites for scientists that incorporate research interests, discussion boards but among others also present citation and activity based metrics. Nonetheless, they do not address the task of finding or even suggesting collaborators with a specialized visualization. ArnetMiner does provide various visualization in order to understand research foci's of scientists (mostly from computer science). From our experience data coverage is highly insufficient in order to suggest collaborators effectively.

In a research cluster with over 200 researchers from different disciplines, making interdisciplinary collaboration in the cluster [3] is hard work.

Initially we visualized existing collaboration by visualizing publication behavior. This visualization was seen to be beneficial in the cluster[21] and can be used for analyzing the degree of interdisciplinarity [6]. Still the requirement to actively suggest collaborators was considered necessary. An approach to do this was to model the suggestions on more than one variable – keyword similarity and a common social network.

3   Research Questions

In our design study, we try to apply the findings from related work to visualize opportunities of possible collaboration. Regarding this visualization we investigate the following research questions:

RQ1 What are user's expectations of a visualization tool to enhance collaboration and organizational knowledge?

RQ2 How can a visualization approach be used to suggest collaborators?

RQ3 Does the visualization at the same time inform members how the organization is structured?

4   Method

Using a user-centered approach, we established *user requirements* first addressing RQ1. For this purpose, we conducted semi-structured interviews, which generated a list of requirements. These requirements were then used to develop several paper prototypes. The design elements of the prototypes were selected in accordance with criteria of visual ergonomics.

Two of these prototypes were selected for data-driven evaluation. This evaluation was based on a speak-aloud scenario-based user test addressing both RQ2 and RQ3. Prototypes were improved in each iteration by immediate feedback evaluation from the researchers.

4.1   Participants

At a local university an integrative interdisciplinary research cluster addresses research in production technology. Currently there are 209 researchers in the cluster in 21 institutes with over 30 faculty. Interdisciplinary collaboration (ranging from material sciences to logistics) is highly important for the given topic and strongly encouraged.

We identified three different user categories, which we refer to as beginners (2 or less publications), intermediates (3-9 publications) and experts (10 or more publications). From this population we selected 40 participants for our studies by randomly selecting researchers from the three experience levels. Thirteen participants from seven different institutes agreed to take part in the study (see Table 1).

Category	Sample	Interviews	User Study
Beginner	4	2	3
Intermediate	5	1	4
Expert	4	2	3
Total	13	5	10

Table 1. Selection of participants from different experience levels for both studies

5   Requirements Analysis – Interview Method

For requirement analysis we conducted five semi-structured interviews (see Table 1). The interviews were divided in three sections. First, questions regarding the participants' background knowledge were asked (i.e. role within the research organization, level of expertise as in published scientific articles, self-evaluation

in regard to scientific impact, interdisciplinary experience, software usage, interdisciplinary motivation).

The second part dealt with the process of publishing scientific articles (i.e. track record, publishing frequency, interdisciplinary publications, favorite publications, literature study process, collaboration and publication practice, joys and frustrations of publishing). This particularly included questions that directly addressed the process of writing and finding co-authors that possibly have required knowledge. It also included the perceived importance of choosing good and relevant keywords.

The last part of the interview related to publishing in the cluster specifically, in particular whether finding co-authors from within the cluster is necessary and whether other members of the cluster show a willingness to collaborate. Interviews took less then one hour and audio was recorded.

5.1 Results from the Interviews

From the transcription of these semi-structured interviews we derived a total of six requirements by categorization (given in italics). For this purpose interviews were transcribed and evaluated according to Mayring [22]. We determined that researchers would like to *form a mental model* (i.e. a structural representation $R1$) of the cluster, the institutes, and the connections between researchers to improve the understanding of the main organizational research interests and orientation of the cluster as a whole ($R2$). Members are willing to *present their own research interests* to others through keywords in order to identify each researcher's expertise and skills. Here they referred to *similarities of keywords* between two researchers as a satisfying indication of relatedness between two researchers ($R3$). We found that members of the cluster often face the challenge of *discovering new co-authors or experts* in a specific field from another discipline that also match their research interests. Some authors have left the cluster but are still considered for consultation, but they should be identifiable clearly ($R4$). Interviewees referred to *willingness to collaborate and motivation* as key factors for identifying possible candidates that want to get involved in interdisciplinary collaboration ($R5$). However, they also struggle to determine a common research method prior to initiating research. It is necessary to *acknowledge current and preceding research interests* to evaluate a possible collaboration ($R6$).

The results from this requirement analysis adequately address RQ1 and were used to generate the visualizations described in the next section.

6 Visualization Prototypes

We observed that our participants were struggling to comprehend the functionality of our prototype using medium fidelity prototypes with imaginary data, hence we decided to take our prototype into high fidelity using real data. For this purpose, we acquired a database of publications from the research cluster from 2012 to early 2014. Furthermore, we extracted authors and keywords from

6   Yazdi et al.

the titles of the papers. Additionally, we identified authors that were no longer in the cluster. Slight improvements were integrated between trials to incorporate user feedback.
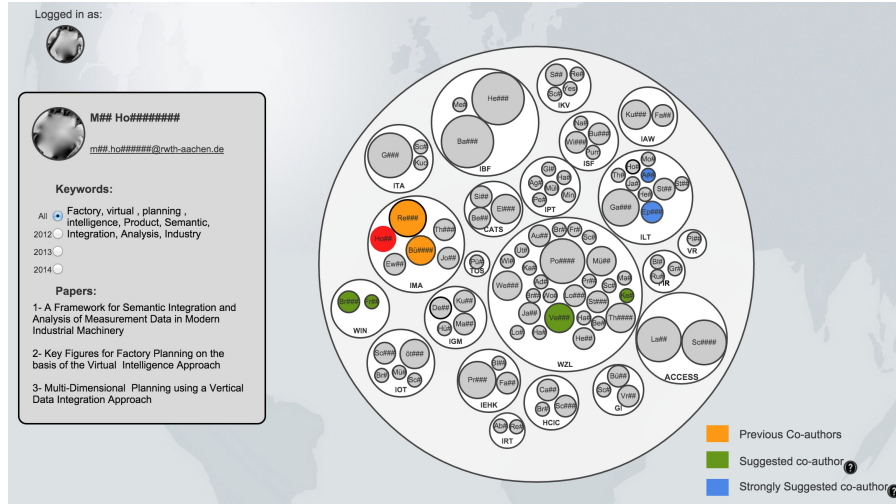

Fig. 1. Prototype 1 showing all members of the cluster. Orange bubbles are used for previous co-authors, green bubbles indicate having at least two similar keywords, and blue bubbles imply two common co-authors that also have at least two similar keywords. The user itself is highlighted in red. By clicking on a bubble the respective colors are overlaid on the suggested collaborators. Names are blurred for reasons of privacy.

The interactive visualization is a bubble graph. Authors are represented as bubbles. Institutes are represented as bubble bags, containing all authors from the respective institute. Bubble size is determined by publication output and increases linearly with increasing publications (see Fig. 1, addressing $R5$). The position of the each author is fixed to a relative location by using the name as a hash for its positioning within its institute. Institute bubbles contain the acronym of the institute. These design choices were made to allow users to visually explore and interrogate the structure of the cluster by visualizing the relevant dimensions of data (addressing $R1 - 2$). Interactive bubble-bag visualizations allow encoding of multiple dimensions (e.g. numbers of papers, keywords, institute, previous/possible connections, etc.), which were indicated as relevant by the users. Bubbles are furthermore spatially efficient and their shape naturally encodes the behavior of transient grouping [23]. Additionally and most importantly participants stated, that their mental image of the cluster was indeed bubble shaped (instead of hierarchically as a triangle for instance).

We used two types of parameters to find new collaborators. We used heuristics to determine possible co-authors according to the "birds of a feather flock together" rationale [19]. Similarity according to keywords and a shared co-

authorship network were used to find suggestions for new collaborators (addressing $R3, 5-6$). In our initial stage of our prototype we found that having only one similar keyword is not a sufficient indication of similarities in research interests according to the users. Validity of extracted keywords was assessed by asking the respective interviewees. Recommendations are given by hovering of author nodes. Relevant recommendations are shown by highlighting recommended co-authors. By color-coding the degree of recommendation additional information is given. This allows not only finding relevant authors for the user himself but also finding relevant connections between different colleagues (addressing $R1-2$). Thus fostering the creation of a mental model of the organizational structure and organizational knowledge. In both prototypes clicking on a bubble opens a panel that reveals the authors name, picture, and email-address. Additionally the list of keywords and publications are shown, which can be filtered according to their years (addressing $R3$).

Our second prototype focused on highlighting only the recommendations for the user by leaving out all non-suggested co-authors (see Fig. 2). This should reduce cognitive load and direct the users attention. Suggestions are placed in orbits according to their suggestion as a co-author. Previous co-authors that are not in the cluster are placed outside of the bubble, addressing the requirement of also showing but at the same time identifying external collaborators (addressing $R4$). Suggested co-authors are placed in the medium orbit. Placement of bubbles within orbits is done using a force-based layout. Authors from the same institute attract each other, while others repel.

Both prototypes can be seen in a short video online[1].

7   Prototype Evaluation – User Study

We tested the developed prototypes, which were based on our requirements analysis, with two participants from the interview study and eight additional users ($N = 10$, see also Table 1). We evaluated it using a scenario-based speak-aloud procedure. Both final visualizations were tested in all trials. We randomized the ordering of the visualization between subjects.

Participants were first asked to interpret the visualization without any interaction. In a second step participants were asked to interact with the visualization and speak about the changes in the visualization. In a third step, finding a possible co-author was given as a task and an evaluation of the suggestion was asked for. Lastly, the participants should freely comment on the visualizations and compare both for suitability. The visualizations were then assessed using the system usability scale (SUS) and the net promoter score (NPS). Both are scales that can be used to quickly judge a tool as a whole for usability and loyalty. They do not provide insights into details of usability problems.

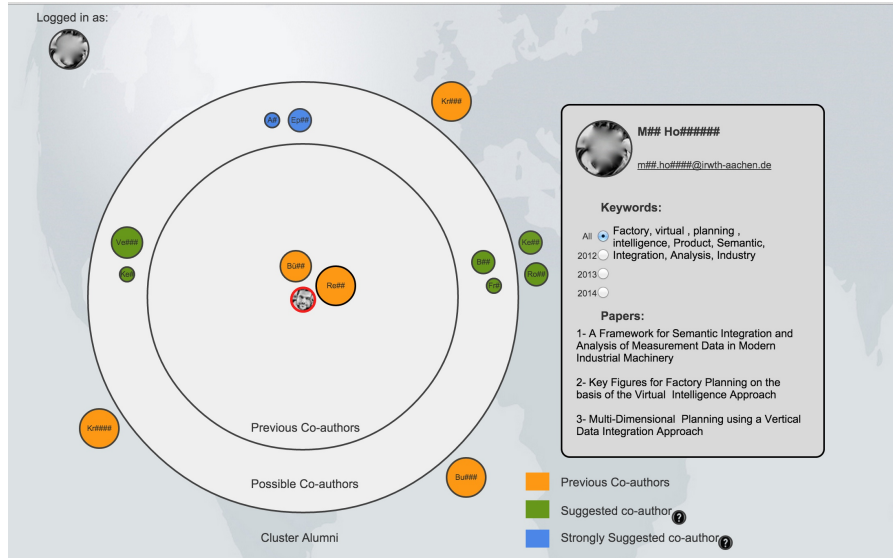[1] A short video demonstration can be found at: `https://vimeo.com/120483587`

Fig. 2. Prototype 2 showing only recommended co-authors. The rings indicate the level of recommendation (inner ring = previous co-authors, outer ring = similar topics, common co-authors)

7.1 User Study Results and Conclusions

As there are similarities and differences between the two visualizations, we decided to split our results into five sections, first describing both common and specific results separately. The evaluation then investigates the validity of our approach and possible applications. All findings relate to two prototypes from the last iteration of our participatory design process.

7.2 General Findings

As interviewees compared publication efforts of their colleagues to the size of the bubble, all immediately concluded that the size of the bubble is proportional to number of papers per person and that larger bubble represent more active and experienced researchers. Users tried to understand our suggestion system by analyzing and comparing their own work, keywords, and papers with previous coauthors to those of each suggested person from the visualization.

All users understood the meaning of colors by hovering over the legend, which explained the reasoning for the different colors. Users found a notification system that informed them about changes in their graph helpful and necessary for long term use. Overall, interviewees preferred to have both visualizations side by side to map necessary information more easily and quickly.

Quantitatively the SUS showed a mean of $M=82.5$ $(SD=24.4)$ indicating a high acceptance of the prototype. The NPS analysis yields 4 Promoters, 6

Passives and 0 Detractors. The overall NPS is 40 indicating good usability and possible loyalty.

Reflections on Prototype 1. This prototype supports the process of decision making by locating key players, their publication effort and connections at institutional level.

Self-awareness, which is another key issue in large organizations, is now partly resolved by being able to consciously track who does what, when and where. By hovering over a group of people connections and topics that over-arch institutional collaboration become visible.

Our visualization also gives an opportunity for exploring possibilities of collaboration between researchers who already know each other. Some participants mentioned that the visualization contained more information about them than they previously knew. During the speak-aloud scenarios utterances like *"oh he works there?"* or *"I didn't know she is also interested in . . . "* occured.

Over all, it became clear that users did not follow a specific pattern to rate or rank suggested collaborators. All preferred to use their own instinct and background knowledge to investigate and choose between suggestions.

Reflections on Prototype 2. This type of visualization enhanced information delivery by removing all unrelated researchers. Participants were much quicker in finding possible co-authors but lacked insights on organizational structure. The closeness of authors, caused by the force-layout, was understood by all users. The benefit of showing external collaborators was well received by the participants. This visualization caused most participants to state that both visualizations should be combined or presented next to each other.

7.3   Validity of the Approach

From our video transcription we extracted all statements that relate to the usefulness of our system. We grouped them into three categories: *Confirmation*, *discovery of new knowledge*, and *problem solving* (see Table 2). Each had 5, 6 and 3 distinct statements respectively. From these statements we derive that our approach successfully addresses RQ2 and RQ3. Our approach is a valid type of visualizing collaboration in a large research organization, which allows finding collaborators and provides a means of creating organization awareness.

7.4   Possible Applications

In addition to finding co-authors through our visualization, interviewees suggested that they could also apply the system to solve other challenges such as finding literature (n=2), discovering experts (n=3), locating people with access to particular facilities or hardware (n=1) and also simplify the process of developing proposals for research grants (n=1).

Category	Transcripts
Confirmation	*"Oh, I have met this person at a conference recently and we have agreed to write a paper together."*
Discovery of new knowledge	*"I do not know the person but it seems like what he does really fits good to my work. I think I can work with him together."* *"Now I know which person I could contact that has related work in this institute for an interdisciplinary publication."*
Problem solving	User hovers over a suggested co-author: *"This visualization could help us having a publication from multiple disciplines."*

Table 2. Example transcripts from the interviews for the three result categories

From our point of view similar visualizations could be used on an institutional level to visualize topics addressed by various institutions, revealing institutes that address similar topics. These could be used in competitor analyses or collaboration scenarios.

8   Limitations and Future work

For our visualizations, we performed both a requirements analysis and a user study in an iterative participatory design process. As future work we would like to include some of the features that were suggested to optimize user fit in the next iteration. As an example, we want to give users the ability to accept or reject a suggested collaborator after evaluation of their relevance. This feedback should be integrated into the recommendation algorithm. Furthermore recommendations could be generated by using text-mining procedures instead of keyword analysis (although this design study did not focus on data generation).

Another example is to display the keyword similarities between the user and suggested co-authors or the capability of viewing co-authors of each particular paper. By extending the scope to suggesting particular papers instead of authors, we could allow the user to judge the relative importance of a certain keyword for the researcher in question.

Furthermore the approach should be extended to include collaborators that have not published yet. This would require new researchers to fill a profile indicating research interests using keywords. Also finding a way of visualizing a missing track record without breaking the natural mapping of size and track record should be considered.

A limitation is the specific sample from one research cluster. To generalize our approach we could map our visualization to other contexts. The bubbles

could also reflect institutes from an entire department or school in order to understand collaboration in a university as a whole. Whether the visualization will effectively scale is yet to be answered. Whether the approach can be used in non-academic scenarios also warrants investigation. The choice of bubbles might be effective only because a research cluster is a loosely coupled organization. In more structred enterprises other forms of representation might be more accurate.

In our approach we assume a relative homogeneous user group. Since regional, organizational and disciplinary cultural differences can lead to a very heterogeneous user group, factors of user diversity must be considered when dealing with data of employees. In addition finding an expert still leaves the task of starting collaboration. Knowledge sharing is social process and requires more than simple tool assistance.

Only titles were used for the extraction of keywords. Using full texts or abstracts should reveal better keywords in the long run as would manual keyword selection by users. Furthermore, no disambiguation of keywords or synonym detection was applied. Particularly in interdisciplinary settings this is a strong requirement. Thus, in this regard our system does not help overcome disciplinary language barriers.

The sample for this study was relatively small (approx. 5% of the research cluster). For a better quantitative evaluation more participants should be considered. Publication data was only selected from 2012 to early 2014, limiting the insights from senior researchers and very recent publications.

References

1. Jacobs, J.A., Frickel, S.: Interdisciplinarity: A critical assessment. Annual Review of Sociology **35** (2009) 43–65
2. Repko, A.F.: Interdisciplinary research: Process and theory. (2011)
3. Marzano, M., Carss, D.N., Bell, S.: Working to make interdisciplinarity work: investing in communication and interpersonal relationships. Journal of Agricultural Economics **57**(2) (2006) 185–197
4. Cummings, J.N., Kiesler, S.: Who collaborates successfully? prior experience reduces collaboration barriers in distributed interdisciplinary research. In: Proceedings of the 2008 ACM conference on CSCW, ACM (2008) 437–446
5. Calero Valdez, A., Schaar, A.K., Ziefle, M., Holzinger, A.: Enhancing interdisciplinary cooperation by social platforms. In Yamamoto, S., ed.: HIMI. Information and Knowledge Design and Evaluation. Volume 8521 of Lecture Notes in Computer Science. Springer International Publishing (2014) 298–309
6. Calero Valdez, A., Schaar, A.K., Ziefle, M., Holzinger, A., Jeschke, S., Brecher, C.: Using mixed node publication network graphs for analyzing success in interdisciplinary teams. AMT (2012) 606–617

 7. Agasisti, T., Catalano, G., Landoni, P., Verganti, R.: Evaluating the performance of
    academic departments: an analysis of research-related output efficiency. Research
    Evaluation **21**(1) (2012) 2–14
 8. Liu, X., Bollen, J., Nelson, M.L., Van de Sompel, H.: Co-authorship networks
    in the digital library research community. Information processing & management
    41(6) (2005) 1462–1480
 9. Morel, C.M., Serruya, S.J., Penna, G.O., Guimarães, R.: Co-authorship network
    analysis: a powerful tool for strategic planning of research, development and ca-
    pacity building programs on neglected diseases. PLoS neglected tropical diseases
    3(8) (2009) e501
10. Liu, Y., Goncalves, J., Ferreira, D., Xiao, B., Hosio, S., Kostakos, V.: Chi 1994-
    2013: mapping two decades of intellectual progress through co-word analysis. In:
    Proceedings of the 32nd annual ACM conference on Human factors in computing
    systems, ACM (2014)
11. Huang, T.H., Huang, M.L.: Analysis and visualization of co-authorship networks
    for understanding academic collaboration and knowledge domain of individual re-
    searchers. In: Computer Graphics, Imaging and Visualisation, 2006 International
    Conference on, IEEE (2006) 18–23
12. Wu, M.Q.Y., Faris, R., Ma, K.L.: Visual exploration of academic career paths.
    In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in
    Social Networks Analysis and Mining, ACM (2013) 779–786
13. Yu, W., Yesupriya, A., Wulf, A., Qu, J., Khoury, M.J., Gwinn, M.: An open
    source infrastructure for managing knowledge and finding potential collaborators
    in a domain-specific subset of pubmed, with an example from human genome
    epidemiology. BMC bioinformatics **8**(1) (2007) 436
14. Chaiwanarom, P., Ichise, R., Lursinsap, C.: Finding potential research collabora-
    tors in four degrees of separation. In: Advanced Data Mining and Applications.
    Springer (2010) 399–410
15. Ehrlich, K., Lin, C.Y., Griffiths-Fisher, V.: Searching for experts in the enter-
    prise: combining text and social network analysis. In: Proceedings of the 2007
    international ACM conference on Supporting group work, ACM (2007) 117–126
16. Loepp, B., Hussein, T., Ziegler, J.: Choice-based preference elicitation for collab-
    orative filtering recommender systems. In: Proceedings of the 32nd annual ACM
    conference on Human factors in computing systems, ACM (2014) 3085–3094
17. Zhang, J., Ackerman, M.S., Adamic, L.: Expertise networks in online communities:
    structure and algorithms. In: Proceedings of the 16th international conference on
    World Wide Web, ACM (2007) 221–230
18. Kretschmer, H.: A new model of scientific colloboration part 1. theoretical ap-
    proach. Scientometrics **46**(3) (1999)
19. Settles, B., Dow, S.: Let's Get Together : The Formation and Success of Online
    Creative Collaborations. In: Proc. CHI 2013. (2013) 2009–2018
20. Wen, Z., Lin, C.Y.: On the quality of inferring interests from social neighbors. In:
    Proceedings of the 16th ACM SIGKDD, ACM (2010) 373–382
21. Schaar, A.K., Calero Valdez, A., Ziefle, M.: Publication network visualization as
    an approach for interdisciplinary innovation management. In: Professional Com-
    munication Conference (IPCC), 2013 IEEE International, IEEE (2013) 1–8
22. Mayring, P.: Qualitative Inhaltsanalyse. Springer (2010)
23. Watanabe, N., Washida, M., Igarashi, T.: Bubble clusters: an interface for manip-
    ulating spatial aggregation of graphical objects. In: Proceedings of the 20th annual
    ACM symposium on User interface software and technology, ACM (2007) 173–182