Scientometrics (2011) 89:569–590
DOI 10.1007/s11192-011-0475-x

Multi-modal social networks for modeling scientific fields

Georg Groh · Christoph Fuchs

Received: 15 March 2011 / Published online: 25 August 2011

Abstract This paper analyzes whether methods from social network analysis can be adopted for the modeling of scientific fields in order to obtain a better understanding of the respective scientific area. The approach proposed is based on articles published within the respective scientific field and certain types of nodes deduced from these papers, such as authors, journals, conferences and organizations. As a proof of concept, the techniques discussed here are applied to the field of 'Mobile Social Networking'. For this purpose, a tool was developed to create a large data collection representing the aforementioned field. The paper analyzes various views on the complete network and discusses these on the basis of the data collected on Mobile Social Networking. The authors demonstrate that the analysis of particular subgraphs derived from the data collection allows the identification of important authors as well as separate sub-disciplines such as classic network analysis and sensor networks and also contributes to the classification of the field of 'Mobile Social Networking' within the greater context of computer science, applied mathematics and social sciences. Based on these results, the authors propose a set of concrete services which could be offered by such a network and which could help the user to deal with the scientific information process. The paper concludes with an outlook upon further possible research topics.

Keywords Social networks · Mobile Social Networking · Modeling of a scientific domain · Co-authorship networks · Person-organization networks · Author co-citation networks · Journal-person networks · Conference-person networks

Introduction

The publishing process can be seen as the backbone of scientific work—it allows an author to participate in the scientific discourse. With the help of publications, researchers

G. Groh (✉) · C. Fuchs
Department of Informatics, TU München, Boltzmannstr. 3, 85478 Garching, Germany
e-mail: grohg@in.tum.de

C. Fuchs
e-mail: fuchsc@in.tum.de

distribute and protect their results. The huge amount of existing literature and the vast number of ways to publish work (e.g. books, journals, conferences, workshops, symposia, e-print archives or regular web pages) makes it difficult to stay in touch with current research results. When students or PhD candidates have first contact to a field of scientific research, acquiring a sound and comprehensive overview can be a huge challenge. In most cases, researchers develop a special kind of feeling for the forging events, persons, journals and research centers in their scientific fields over time. This (subjective) valuation is often backed by citation indices or other indicators, e.g. the journal impact factor. The idea behind such indicators is a supposed correlation between the quality of a publication and the number of documents citing it. If a document is cited more often, its indicator gets a higher valuation that leads to an increasing reputation of the original author within the scientific community. Similar indicators exist for journals.

These indicators only allow an assessment of individual persons or journals—a comparison between two items is possible, but it does not expose much about the deeper reasons. This work proposes an approach to model a scientific field entirely based on social network analysis and graph theory for human overview. We examine possibilities to render the time consuming process of researching and evaluating of scientific literature more efficient and traceable. Misjudgments resulting from subjective decisions should be avoided via a structured analysis of the collected data. A collection of publications, authors, journals, conferences and organizations from the scientific field 'Mobile Social Networking' was collected with the help of a specifically developed web application. This data-set forms the model of the scientific field. Based on this model, a set of services is developed and discussed in the next step.

The paper is divided in several parts: first, an overview of related work is given, explicitly with respect to co-authorship networks, co-citation analysis and domain analysis. Afterwards, elements of the new approach are introduced and discussed for the collected Mobile Social Networking dataset. Later, a number of possible use cases is presented and an outlook upon further research topics is given.

Related work

Social Network Analysis (SNA) has been suggested as a valuable tool in the scientific fields concerned with methodologies and services for monitoring, searching, analyzing and structuring information in general and especially scientific contributions (journal or conference articles, books etc.) in particular such as infometrics, bibliometrics and scientometrics (see Otte and Rousseau 2002). The analysis of multi-modal networks involving and relating authors, articles, conferences and other modes with the help of e.g. centrality measures (Koschützki et al. 2005), analysis of dense sub-networks (Clauset et al. 2004; Gaertler 2005; Kosub 2005) or role analysis (Lerner 2005) or network visualization techniques (Groh et al. 2009; Kaufmann and Wagner 2001) has become an important approach in these disciplines. Classical link analysis has adopted and incorporated methods from SNA (see e.g. Thelwall 2004, chapter 22) for studying e.g. author collaboration networks and other social modes of the mentioned multimodal networks, so that all fields concerned with network analysis make use of similar methods from abstract network analysis (see Björneborn 2004; Park 2003; Thelwall et al. 2005). We will now briefly discuss key sub-fields that are immediately relevant for our study.

Co-authorship networks

When two authors publish their work together, both authors are treated as nodes and are connected by an edge (their publications) in the co-authorship graph. The edge can be weighted e.g. to reflect the number of papers published jointly or to illustrate temporal aspects. Co-authorship networks are—as typical representatives for social networks—scale-free and conform to the small world phenomenon (Barabasi et al. 2002; Porter et al. 2009).

An often cited example for a co-authorship network is the Erdős number proposed by Goffman (1969). This number reflects the length of the shortest path from an author's node to the node representing Paul Erdős in the co-authorship network. Newman described his approach to analyze the co-authorship network formed by articles from physics, biomedicine and computer science in Newman (2001a) and presented a measure to weight the collaboration ties in Newman (2001b), Lu and Feng (2009) used Salton's measure instead (Luukkonen et al. 1993).

In most cases, the co-authorship graph will consist of several single components. The common procedure is to narrow the analysis to the biggest component (in the studies of Newman (2001c) and Yan et al. (2009) the biggest component contains 57–90% of all authors).

Co-citation networks

The assumption for co-citation analysis is that two documents, authors, journals or other objects which get cited jointly by a (later) third document have—at least from the perspective of the citing author—some coincidence in terms of content. The more frequently two objects get cited together the more this similarity is stressed. This technique was presented for documents (*document co-citation*) by Small (1973) and for authors (*author co-citation*) by White and Griffith (1981) and is often used to create a semi-automatic overview of the literature of a scientific field. Traditionally, ISI's Science Citation Index of Thomson Reuters is used as a data pool. Both Small et al. (1973–1985) and White/McCain (1981, 1986–1990, 1998) describe their respective method and offer a lot of concrete examples. A discussion of the advantages/disadvantages of document co-citation vs. author co-citation is given by White in (1990). By now, many author co-citation studies have been performed for many different fields (e.g. Chen and Carr 1999; McCain et al. 1990; Tsay et al. 2003; White and McCain 1998; Zhao and Strotmann 2008). Current approaches use Pathfinder Networks (Buzydlowski 2003; Chen and Morris 2003; Chen and Hsieh 2007; Lin et al. 2003; McCain et al. 1990; White 2003b) or self-organized maps (Buzydlowski 2003; Lin et al. 2003).

Due to the lasting discussion whether the often used Pearson's R measure is the right measurement (Bensman 2004; van Eck and Waltman 2008; Egghe and Leydesdorff 2009; Leydesdorff 2008; Leydesdorff and Vaughan 2006; White 2003a) a new approach by Wallace and Gingras (2009) uses the raw co-citation network without any further computation.

Domain analysis

Birger Hjørland and Hanne Albrechtsen proposed the term "domain analysis" for the development of a deeper understanding of a scientific field from the perspective of library science (Hjørland and Albrechtsen 1995). Their approach lists a set of methods which allow a researcher of library science to gain technical knowledge in a different scientific field (Hjørland 2002).

Our approach

Data set

It was the goal to collect a data set of scientific articles of all types (including conference-, journal, or review articles) and associated 'items' (authors, conferences, large scientific projects, cited articles etc.) that 'covers' the scientific domain 'Mobile Social Networking' with the help of a suitable tool in order to construct multi-modal social networks from this data set for later analysis. In order to more precisely define 'to cover' one basically has to answer three questions/define three sub-concepts of 'to cover':

- Define criteria whether an article belongs to the domain in question
- Define criteria when the data-set is sufficiently large to map the relevant structures in the corresponding networks.
- Decide upon the set of meta-data recorded for the construction of the networks

The first question/challenge corresponds to maximizing Information Retrieval (IR) precision with respect to searching for respective articles in the domain. In order to maximize precision we employed the following process: First we identified a set of search services/databases for scientific articles in the field of computer science on the basis of an informal assessment of their reputation and potential coverage of the field in question through limited interviews among experts and comparison of lists of such services on the Web. This step can be formalized by refining measures for reputation and coverage of the corresponding database through a bootstrapping process: reputation can be assessed by interviews conducted among domain experts, which are found by network analysis. Coverage can also be decided based on network analysis. This type of bootstrapping alternation between analysis-steps and refinement of analysis objects (in this case the set of scientific databases or the set of retrieved articles) is characteristic for the type of problem investigated. For this study, we restricted ourselves to three sources (Google Scholar, arXiv.org and Springer-Link). In the IR systems associated with the respective databases, we employed a keyword based bootstrapping process: we started with the terms *mobile social networks*, *mobile social networking* and *location based services*. From the titles and abstracts of the found articles we subjectively decided upon their relevance for the domain in question. Again this decision step can be refined by intermediate network analysis steps. If relevant we added the article and all of its associated items (via accessible meta-data) to our database with the help of our meta search tool. We then expanded our keyword list by adding keywords manually extracted from titles and abstracts. Searching, relevance decision and keyword-list expansion were iterated until the keyword list did not change significantly any more. After that, as an optional step, the set of retrieved associated items and their relations was refined by a manual Web-search using the author's names and articles titles in order to identify e.g. collaboration in scientific groups, scientific projects, or further author to author relations (such as PhD-supervisor-PhD-candidate relations) which can optionally be included into the data-set.

The second question or challenge posed above corresponds to maximizing a structural flavor of IR recall: Structural recall is considered maximal if adding an item o (article, author, project etc.) or a set of items $o_1, \ldots o_n$ to the network G, resulting in a new network

G' does not alter the network significantly, measured by a threshold T on a suitably chosen similarity measure $sim(G, G') \leq T$. Numerous alternatives for *sim* exist (Baur and Benkert 2005), e.g. relating $|G|$ and $|G'|$ to their maximal isomorphic sub-graph (see Corneil and Gotlieb 1970; Ullmann 1976), via a weighted minimal set of operations for transforming G into G' (Messmer and Bunke 1998) or via a portfolio of general statistical measures characterizing the network (centrality statistics, distance statistics etc.) (Brinkmeier and Schank 2005). However, we considered the systematic application of such techniques beyond the scope of our study and relied on a subjective assessment of a sufficient structural recall.

Concerning the third question, we decided upon a set of associated meta data/items to consider for the later construction of the networks we want to analyze, which are described in more detail in the following sections. The data-model of items considered encompasses persons (authors and researchers), documents (articles), journals (and journal issues), conferences (and conference instances) and projects (which is an abstraction of scientific projects, working groups and other target-oriented organizations of persons) and free optional tags for every item. The relations between the items e.g. general affiliations (relating items with items), citations or person-to-person-affiliations described in the following sections in more detail. We furthermore decided to construct a specialized tool, which allowed to interleave search steps and semi-automatic update any of the modes/items considered.

The resulting data set which was collected in July and August 2009 consists of 933 articles and their associated items/objects from the scientific (mostly computer science-related) domain 'Mobile Social Networking', forming a multi-modal network.

Among a number of less interesting minor problems, the collection of the data raises two problems which need to be considered:

- A single physical object can be depicted by multiple different representations (*synonym*), e.g. the names “J. E. Katz”, “James Katz” or “James E. Katz” can all refer to the same person. Synonyms can be avoided to a big extend by manual intervention: Therefore, a list of all author’s surnames in alphabetical order is generated and checked for identical surnames (these are the candidates for potential synonyms). For a pair of author names (a_1, a_2) being a possible synonym the associated set of documents D_1 respectively D_2 can be determined (with all documents in D_1 being written by author a_1 and all documents in D_2 being written by author a_2). It is advisable to search for titles of documents in D_1 and the full name a_2 (and vice versa) using web search engines. If a reliable source confirms a connection between D_1 and a_2 resp. between D_2 and a_1, a_1 and a_2 belong to the same equivalence class and are synonym.
- The denotation of multiple objects with a single name is called *homograph*. This is the case when a database entry for “Paul Smith” depicts two different persons, both being called “Paul Smith”. The solution for this problem is to add at least one additional criterion (e.g. the date of birth) to increase the probability to get a unique identifier for each person. Due to the huge effort we refrained from this solution within this work.

Aging of ties

Nearly all graph analysis procedures have the problem that with increasing amount of data a newly added edge has subsiding influence on the graph. The influence of a new edge is small when a large number of other edges in the graph already exist. If one is interested in short-term trends by analyzing the graph’s dynamics, the statistical dominance of these numerous old edges might cover up the the significance of new edges, which might represent a appreciable paradigm shift.

Therefore, certain assumptions should be made:

- There is a direct relation between the scientific field and the deduced graph: changes which take place in the real world can be traced in the graph. Any change in the graph is caused by a respective transition in the real world and vice versa. Changes in the graph consist of adding or dropping nodes (persons, organizations, etc.) and edges (relations).
- The intensity of the relations between objects can be modeled by the weight of the relations (e.g. in a co-authorship network the number of jointly published articles is proportional to the intensity of the relation in real life).
- Relations between nodes become less important when they were not activated for a long time.

Based on these assumptions, we use the following approach to model the aging of edges in a graph based on three parameters *m*, *g* and *l*:

1. If a past relation between two nodes could be identified, the minimal possible edge weight is *m* (i.e. there is no further aging if the weight is already *m*).
2. Time is seen as a discrete value, the unit of measurement is 1 year. Each year in which two nodes have a proven relation leads to an additional edge weight of $g \cdot (ln(cnt) + 1)$ where *cnt* is the number of contacts between the two nodes in the respective year. A single contact within 1 year leads to an increase of the edge weight of *g*. The marginal improvement drops with every further contact in the same year (that's why the *ln* function is used).
3. For each year since the first contact the edge gets "older", i.e. the weight drops with *l*. If the edge weight falls back to *m*, the aging process is stopped until the weight becomes larger than *m* again.

For a given pair of nodes (*a*, *b*) the edge weight can be calculated like that:

1. Get the years where the two nodes had contact, save the years and the number of contacts for each year.
2. Calculate the positive change of the weight (*YearsWithContact* is the set of years where *a* and *b* had contact, cnt_j represents the number of contacts in a given year *j*):

$$p := g \cdot \Sigma_{j \in YearsWithContact}(ln(cnt_j) + 1)$$

3. Calculate the negative change of the value ($Year_{firstContact}$ is the year where both had contact for the first time):

$$n := (Year_{now} - Year_{firstContact} - |YearsWithContact|) \cdot l$$

4. The edge weight between two nodes *a*, *b* can be calculated thus:

$$weight(a, b) := max(m, p - n).$$

Methodology of interpretation

In chapter 5 the results are discussed using the techniques explained now. At first, different topological graph features are discussed and compared (e.g. number of components, clusters, centrality measures etc.). In order to get an idea of the content of the underlying documents, a vector space model is used. Therefore, the titles and abstracts of the documents of each entity form the text corpus for the respective entity. To get better results, Porter's stemming algorithm (Porter 1980) is used to get fewer equivalence classes. Based on this representation, each group of documents form a vector in a vector space with the different word stems forging the different dimensions of the vector space. These vectors can be easily compared using the cosine measure.

Based on the description gained by this means, the most important stems for each document group (component or cluster) determined by *tfidf* (Manning et al. 2008) are presented.

Discussion of the results

Analysis of the co-authorship network

The nodes represent authors. Two nodes are connected with an (undirected) edge if they published at least one document together. The co-authorship network derived from the collected data consists of 1687 nodes and 2926 edges. The graph resolves into 538 components with the biggest component containing 200 nodes (approx. 11.86%). This proportion is a lot lower than the results observed in other studies like (Newman 2001c) or (Yan et al. 2009) but one has to keep in mind that these studies try to consider all works by an author whereas the present study only takes into consideration works based on their individual content. Table 1 shows the biggest components of the co-authorship network. Converting the data into a vector space model and comparing the components' vectors as explained above yields the results shown in Table 2. All components have the biggest similarity with component #3 (this is not surprising since it is the one with >11% of all nodes).

Using $tf - idf$, a list of describing word stems can be identified for each component (as listed in Table 3). Component #3 focuses on application, mobility, social relations and usefulness, component #8 deals with sensor networks and component #2 attends to navigation. Component #51 concentrates on data acquisition for mobile devices, #49 is rather concerned with data storage topics. Component #1 can be located near the medical realm, scale-free networks are used to analyze the spread of diseases. The articles in component #40 address topics from the security and privacy area in a mobile context. In component #25, the topics range from virtual communities to the role of the internet in social relations.

Table 1 The biggest components of the co-authorship network

id	# Nodes	# Edges	Density	Diameter	Avg. path length	Glob. cl. coeff.
3	200	785	0.0394	13	5.2102	0.6330
8	63	276	0.1413	6	3.0266	0.8034
2	23	37	0.1462	5	2.7866	0.4960
51	23	84	0.3320	4	2.0158	0.8502
49	20	52	0.2737	3	1.7789	0.5718
1	19	41	0.2398	5	2.5673	0.6774
40	18	39	0.2549	5	2.4510	0.7500
25	17	33	0.2426	3	1.8456	0.4888
$\sum$	383	1347				

Table 2 Cosine between the vectors of the biggest components of the co-authorship graph—the highest value is highlighted

id	3	8	2	51	49	1	40	25
3	1.000	***0.427***	***0.425***	***0.414***	***0.307***	***0.346***	***0.330***	***0.305***
8	***0.427***	1.000	0.208	0.375	0.189	0.180	0.174	0.164
2	0.425	0.208	1.000	0.221	0.175	0.121	0.180	0.240
51	0.414	0.375	0.221	1.000	0.199	0.153	0.160	0.153
49	0.307	0.189	0.175	0.199	1.000	0.124	0.288	0.095
1	0.346	0.180	0.121	0.153	0.124	1.000	0.093	0.163
40	0.330	0.174	0.180	0.160	0.288	0.093	1.000	0.092
25	0.305	0.164	0.240	0.153	0.095	0.163	0.092	1.000

Table 3 Word stems with high *tfidf* results for each component of the co-authorship network

id	stems
3	network, applic(*-ation, -able*), comput(*-ational,-er, -ation, -ing*), human, mobil(*-e, -izing, -ity*), system, social(*-, -ize, -ly*) undus(*-e, -ed, -eful, -es, -efulness, -ing*)
8	sensor, network(*-, -ed*), node, chord, cricket, pothol(*-e*), capac(*-ity,-ities*), sens(*-ing*), applic(*-ation,-able*), road
2	social, navig(*-ation*), space, activ(*-e,-ity*), voice-mail, privaci (*=privacy*), capit(*-al*), workspac(*-e*), studi (*=study, studying*), new
51	sens(*-ing, -e, -ed*), people-centr(*-ic*), sensor, metrosens(*-e*), bikenet, micro-mobl (*=micro-mobility*), network(*-, -ing, -ed*), opportunist(*-ic*),data, model(*-, -led*)
49	queri (*=query, querying*), data, multidimension(*-al*), imprecis(*-e, -ion*), pre-aggreg(*-ation*), cube, olap, move(*-, -ing*), type, dimens(*-ion*)
1	epidem(*-ic*), scale-fre(*-e*), network, pandem(*-ic*), spread(*-,-ing*), fit(*-ness*), wordwid(*-e*), global, influenza, attach(*-ment*)
40	locat(*-ion*), servic(*-e*), cloak(*-ing, -ed*), locaction-bas(*-ed*), cach(*-e, -ed, -ing*), anonym(*-ity,-izer, -ous*), server, queri (*=query*)
25	internet, cssn(*-s*), media, commun(*-ication,-icate, -ity*), ti(*-es, -ed*), onlin(*-e*), social, network(*-, -ed*), capit(*-al*), contact(*-, -ed*)

The application of several centrality measures on the nodes in component #3 shows the most central authors in the co-authorship graph (Table 4). It is remarkable that e.g. Alex Pentland or Mike Y. Chen just published with comparatively few colleagues (because both have a low degree centrality) but have a considerable high closeness centrality. When eigenvector centrality is used and an aging mechanism is taken into account (as considered above with $m = 1, g = 2, l = \frac{2}{3}$) the centrality vectors are nearly the same (Pearson's R $\sim$0,97). This can be an indication that the data is quite up to date.

Analysis of the person-organization network

The Person-Organization Network is the graph which results from loooking at the relations between authors and their affiliations (companies, universities, research centers, etc.) retrieved from the articles in the database. The bipartite graph consists of 393 components (1194 author nodes and 544 nodes standing for organizations). To increase the probability

Table 4 Most central authors of the biggest component in the co-authorship graph (without considering edge weights)

Degree		Closeness		Betweenness	
Anthony LaMarca	0.171	Tanzeem Khalid Choudhury	0.310	Tanzeem Khalid Choudhury	0.486
Gaetano Borriello	0.166	Anthony LaMarca	0.292	Alex (Sandy) Pentland	0.480
Sunny Consolvo	0.161	Sunny Consolvo	0.291	David Lazer	0.368
Albert-László Barabási	0.161	Jeffrey Hightower	0.290	Albert-László Barabási	0.350
Jeffrey Hightower	0.151	James A. Landay	0.290	James A. Landay	0.300
Ian E. Smith	0.146	Ian E. Smith	0.289	Jason I. Hong	0.269
Tanzeem Khalid Choudhury	0.126	Alex (Sandy) Pentland	0.288	Nathan Eagle	0.190
James A. Landay	0.126	Gaetano Borriello	0.285	Bill N. Schilit	0.152
Bill N. Schilit	0.111	Timothy Sohn	0.274	Eyal de Lara	0.141
Timothy Sohn	0.111	Mike Y. Chen	0.272	Hannu Toivonen	0.105
Jason I. Hong	0.111	Dirk Haehnel	0.271	Roy Want	0.105
Dirk Haehnel	0.111	Johnathan Lester	0.270	Antti Oulasvirta	0.104
James Scott	0.101	Beverly Harrison	0.267	James Scott	0.098
Eyal de Lara	0.101	Bruce Hemingway	0.267	Gaetano Borriello	0.088
Jonathan Lester	0.090	Louis LeGrand	0.267	Heikki Mannila	0.086
Arithmetic mean	0.039	Arithmetic mean	0.201	Arithmetic mean	0.021
Median	0.025	Median	0.197	Median	0.000
Std. Dev.	0.033	Std. Dev.	0.041	Std. Dev.	0.070

that two persons connected to the same organization really know each other, an organization in this context is defined by the combination of "company" and "department" (resp. "university" and "faculty") —e.g. "Intel Labs Seattle" and "Intel Research Berkeley" are treated as different organizations, although the company is the same.

The biggest component contains 330 nodes ($\sim 15\%$) and is composed of 277 authors and 53 organizations. The average degree of an organization node is 6.62 (median 4.0) and the standard deviation is 7.76. For the nodes representing a person the average degree is 1.27 (median 1.0) with a standard deviation of 0.56. This implies that a big part of the authors in the data set maintain relations to only one organization and that many authors concentrate on a few organizations (17 organizations only have one assigned author each, whereas just two organizations— Carnegie Mellon University and MIT Media -Lab—are connected to more than 30 authors). The network is shown in Fig. 1. The edges are weighted using the aging mechanism mentioned above with parameters $m = 1$, $g = 2$ and $l = \frac{2}{3}$. Nodes with a high degree are the computer department of Carnegie Mellon University, Media Laboratory and Laboratory for Computer Science at MIT, the Department of Computer Science at University of Washington (and University of Washington without any mentioned faculty), the Computer Science Department at UC Berkeley, Intel Research, AT&T's Research Labs, University of Toronto and the College of Computing at Georgia Institute of Technology.

According to the edge weights, some organizations like the Computer Science Department at Carnegie Mellon University, Intel Research or MIT Media Laboratory either publish more regularly or just published current studies (UC Berkeley, for example,

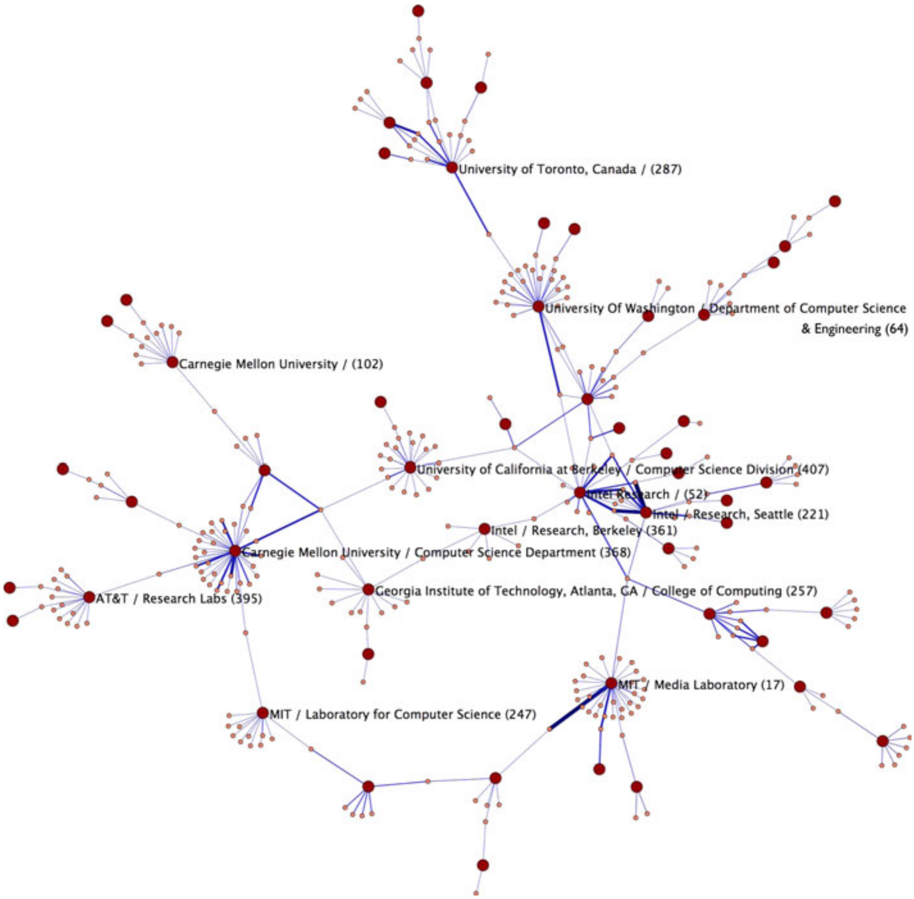
University of Toronto, Canada / (287)
University Of Washington / Department of Computer Science & Engineering (64)
Carnegie Mellon University / (102)
University of California at Berkeley / Computer Science Division (407)
Intel Research / (52)
Intel / Research, Seattle (221)
Intel / Research, Berkeley (361)
Carnegie Mellon University / Computer Science Department (368)
AT&T / Research Labs (395)
Georgia Institute of Technology, Atlanta, GA / College of Computing (257)
MIT / Media Laboratory (17)
MIT / Laboratory for Computer Science (247)

Fig. 1 Largest component of the person-organization network, the width of the edges represent their weight

appears as a different case—the data set probably does not contain publications on a regular basis or current publications).

Analysis of the author co-citation network

An author co-citation analysis can be done at least in two ways: The traditional way ensuing McCain's paper (McCain 1990) uses a vector model to compare the co-citation profiles of authors. For each author, a vector is calculated which contains the author co-citation count for each other author of the data set. Afterwards, the analysis compares the author vectors using a measure like cosine (van Eck and Waltman 2008; Egghe and Leydesdorff 2009) or Pearson correlation (McCain 1990). A link between two authors does not necessarily mean that both got co-cited, it just tells something about the similarity of their co-citation vectors (i.e. how they get co-cited with all authors).

The other approach used in studies like White (2003b) works on the raw data: a link between two authors expresses that these two authors got co-cited (and does not reveal something about their relationship to any other author).

Since this study focuses on social networks, the second approach is discussed in detail. Contrary to many other studies this work does not limit the author co-citation analysis to the first author but comprises all authors of an article. The reason for the first author limit is traditionally the better availability of the data—in our case this is not significant since this study uses a self-maintained data set. In the classification scheme of (Rousseau and Zuccala 2004) a "pure author co-citation" is performed, i.e. all co-authors are taken into account, but cited documents where two authors are co-authors do not count as co-citation counts for this specific authors in order to avoid an overlaying co-authorship network.

The data set contains 1687 authors who make up 52928 co-citation pairs. 1490 authors ($\sim$88.3%) form the biggest component with 51171 edges. The remaining 197 authors are distributed among 148 additional components which are noticeably smaller (and therefore not considered in this article). The diameter of the biggest component is 7, the density is 0.0475, the average path length is 2.8672, the global clustering coefficient is 0.7 and the average local clustering coefficient is 0.86.

Table 5 gives an overview of the most central authors where centrality is determined with different centrality measures. Degree-, closeness- and betweenness-centrality ignore edge weights. Eigenvector centrality was applied with absolute co-citation counts as edge weights and with aged edge weights with parameters $m = 1$, $g = 2$ and $l = \frac{2}{3}$.

One can see that the team of Intel Research around Anthony LaMarca and their co-authors get co-cited with a lot of other authors (Ian E. Smith, James A. Landay, Sunny Consolvo, Jeffrey Hightower, Gaetano Borriello, Bill N. Schilit, Jeff Hughes, James Scott, John Krumm, William G. Griswold, Mike Y. Chen, Yatin Chawathe). The only researcher in the top-10 list of the degree centrality who does not have a direct link to LaMarca's Intel research team is Hari Balakrishnan from MIT. Looking at the list of nodes with a high closeness centrality one can notice the central position of Alex Pentland (as already seen in the analysis of the co-authorship network) working at MIT Media Lab. Nathan Eagle works at MIT Media Lab, too. Antti Oulasvirta, Mika Raento and Hannu Toivonen work at University of Helsinki on the "Context" project (Contextproject. URL http://www.cs.helsinki.fi/group/context/ 2009). Richard S. Ling works at Telenor Research and published in the late 1990s/early 2000s several research articles concerning the mobile communication in Norway. Albert-László Barabási works on the theory of scale-free networks and classic social network analysis. Christian S. Jensen works on data management in mobile applications at Aalborg University in Denmark. Caroline Haythornthwaite's research topics include topics like internet and society, mobile social networking and e-learning. David W. McDonald works at the University of Washington in the areas of computer supported cooperative work and human computer interaction. Shravan Gaonkar deals in the collected data set with distributed publishing with the help of mobile phones. Romit Roy Choudhury is (just like Landon P. Cox) a co-author of Shravan Gaonkar and works together with Gaonkar at the University of Illinois (Urbana-Champaign).

It is prominent that the eigenvector centrality vectors with and without the aging mechanism are quite similar (Pearson's correlation is 0.945, cosine is 0.954). This phenomenon can also be seen in the co-authorship analysis above and might be caused by the fact that the collected documents are quite recent and one might argue that the scientific field is quite young.

The next question is whether it is possible to split the big component into several clusters with different research topics within the mobile social networking community. Thus the big component was clustered using a clustering method based on modularity by

Table 5 Central authors of the largest component of the author co-citation network (based on absolute values)

Degree		Closeness		Betweenness		Eigenvector (abs.)		Eigenvector (aged)	
Ian E. Smith	0.266	Ian E. Smith	0.514	Richard S. Ling	0.148	Ian E. Smith	1.000	James A. Landay	1.000
James A. Landay	0.238	Alex (Sandy) Pentland	0.512	Alex (Sandy) Pentland	0.075	Jeffrey Hightower	0.984	John Krumm	0.976
Sunny Consolvo	0.233	Nathan Eagle	0.501	Albert-László Barabási	0.061	Sunny Consolvo	0.983	Hari Balakrishnan	0.963
Jeffrey Hightower	0.233	Richard S. Ling	0.501	Christian S. Jensen	0.053	Anthony LaMarca	0.974	Jeffrey Hightower	0.913
Anthony LaMarca	0.230	Antti Oulasvirta	0.488	Caroline Haythornthwaite	0.046	Gaetano Borriello	0.945	Gaetano Borriello	0.912
Gaetano Borriello	0.216	Mika Raento	0.485	David W. McDonald	0.046	James A. Landay	0.931	William G. Griswold	0.909
Hari Balakrishnan	0.209	Sunny Consolvo	0.480	James A. Landay	0.043	Shravan Gaonkar	0.821	James Scott	0.909
Bill N. Schilit	0.207	Jeffrey Hightower	0.480	Nathan Eagle	0.043	Romit Roy Choudhury	0.821	Sunny Consolvo	0.908
Jeff Hughes	0.197	Anthony LaMarca	0.479	Hannu Toivonen	0.038	Michel Goraczko	0.820	Mike Y. Chen	0.905
James Scott	0.197	Gaetano Borriello	0.475	Ian E. Smith	0.035	Landon P. Cox	0.820	Yatin Chawathe	0.888
Arithmetic mean	0.048	Arithmetic mean	0.355	Arithmetic mean	0.001	Arithmetic mean	0.098	Arithmetic mean	0.098
Median	0.040	Median	0.358	Median	5×10^{-6}	Median	0.006	Median	0.003
Standard deviation	0.038	Standard deviation	0.046	Standard deviation	0.006	Standard deviation	0.195	Standard deviation	0.220

Clauset, Newman and Moore (Clauset et al. 2004). To make the interpretation a bit easier, we restricted the graph to edges with minimal weight of 2 (to avoid random co-citations). Such a threshold is an accepted way to limit the study to the most connected authors (McCain 1990). Thus, the number of edges is reduced from 52928 to 1547. Again we get a large component with 211 nodes (the second and third largest components contain five and four nodes, all other components just contain one node). The modularity reached with the clustering mechanism is 0.616 and relatively high. The word stems with high $tf - idf$ ranking are listed in Table 6 for each cluster separately.

A Pathfinder Network (Schvaneveldt et al. 1989) of the largest component can be seen in Fig. 2.

Cluster #1 contains documents which deal with different use cases for context sensitive applications. The example applications cover e.g. tourist information systems and the fast development of prototypes for mobile, context sensitive applications. Central authors are Victoria Bellotti, Adrian Friday, Keith Mitchell and Keith Cheverst. In Cluster #2 the theory of scale-free networks is the main topic, important authors are Réka Albert, Albert-László Barabási, Duncan J. Watts and Mark Granovetter. Cluster #3 highlights security and data privacy, the main authors are Marco Gruteser, Jason I. Hong, Paul Dourish und Anind K. Dey. The authors in cluster #4 write about ubiquitous computing, mobile applications with social, local and contextual reference. Important persons are Alex Pentland, Ian E. Smith, Roy Want and Nathan Eagle. In cluster #5 the influence of new forms of communication and the general technological development for our society are examined. Besides Richard S. Ling, Nicola Green, Barry Wellman and Pirjo Rautiainen are the essential authors here. The topics in cluster #6 are sensor networks based on mobile phones. The most important authors are Dirk Haehnel, Beverly Harrison, Landon P. Cox and Shravan Gaonkar. Compared with the other clusters, cluster #7 is rather hardware-centric and deals with delay tolerant networks. The respective authors are Ashvin Goel, Sushant Jain, Rabin Patra, Kevin Fall and Jing Su.

Analysis of the journal-person network

This part of the article discusses the derived journal-person network. The bipartite graph contains 151 components with more than one node and consists of 861 authors and 242 journals. In this part, the largest component with 443 authors and 69 journals is discussed. For this case, we decided to emphasize the journals, so the journal-network will be analyzed since more methods for dealing with one-mode networks than methods for the analysis bipartite networks are known. The nodes represent journals and two journals are connected by an edge when at least one person exists who published in both journals. The weight of an edge correlates to the number of authors who published in both of the connected journals. Thus, the decision on how similar two journals are is left to the authors, the users of the journals, themselves.

The density of the network is 0.084, it consists of 69 nodes (journals) and 196 edges. The Pathfinder Network shown in Fig. 3 reaches from the area of human computer interaction (upper left) to the social sciences (upper right): starting at ACM Transactions on Computer-Human Interaction (without label in the figure: Wireless Networks, ACM Transactions on Information Systems, Human Computer Interaction), proceeding to rather general IT journals (Communications of the ACM) to IEEE Pervasive Computer (beneath without label: among others IEEE MultiMedia, Personal and Ubiquitous Computing, MIT

Table 6 Words with high tf-idf rating for each cluster of the largest component in the author co-citation network

cluster	words/stems
1	hoard (-ing, -), guid (-e), tourist (-), visitor (-s), disadvantag (-es), web (-), context-awar (-e), wan (-s), scalabl (-e), protocol (-s), portabl (-e), outdoor (-), lancast (-er), flexibl (-e), context-sensit (-ive), citi (city), deploi (deployed, deploying), represent (-ations), transfer (-ring), suffer (-)
2	scale-fre (-e), evolut (-ion), epidem (-ics, -ic), graph(-s, -), complex (-ity, -), bursti (bursty, -ness), weak (-), ti (-es, -ed), small-world (-), acquaint (-ance, -ances), random (-ness, -), strength (-, -s), rang (-e, -ing), phenomena (-), overlap (-s, -ping), media (-), phase (-), memori (memory), mathemat (-ical, -ics), asset(-)
3	protect (-ing, -, -ion), k-anonym (-ity), releas (-e, -ed, -ing, -es), holder (-s, -), suppress (-ion), guarante (-es, -eing, -e), anonym (-ity, -ous), re-identifi (re-identify, -ed), entiti (-es), datafli (datafly), confab (-), combin (-es, -e, -ed, -ing), cannot(-), -argu (-s), record (-s, -), identifi (-ers, identify, identifying), safeguard (-ing, -s), resolut (-ion), remov (-ed, -e), person-specif(-ic)
4	beacon (-s, -), wearabl (-e), sensor (-s, -), smartphon(-e, -es), cricket (-), awar (-eness, -e), gp (-s), automat (-ic, -ically), strategi (strategy), parasit (-ically, -ic), learn (-ed, -ing, -), wearcom (-), social-mobil (-e), mode (-), manual (-, -ly), listen (-er, -ers, -ing), lab (-), hummingbird (-s, -), give (-, -s), contextphon (-e)
5	youth (-), telephoni (telephony), relationship (-s, -), accept (-ance, -), space (-s, -), eas (-e), young (-), rheingold(-, -s), perceiv (-ed, -ing), cultur (-e, -al), teenag (-ers), rhythm(-s), adolesc (-ents, -ent), telephon (-e, -es), life (-), voice-mail(-), perspect (-ives, -ive), norwegian (-), japanes (-e), home (-)
6	sensor (-s, -), people-centr (-ic), opportunist (-ic), metrosens (-e), bikenet (-), satir (-e), participatori (participatory), cyclist (-), archiv (-ing, -e, -al), mobiscop (-es), etc (-), monitor(-ing), upload (-ing), testb (-eds), small-scal (-e), multihop (-), mote (-s), microphon (-e), last (-), everi (every)
7	packet (-s), delay-toler (-ant), end-to-end (-), radio (-), toler (-ance, -ant), delai (delay), possibl (possibilities, -e), algorithm (-ic, -s, -), wireless-en (-abled), usersth (-e), time-vari(time-varying), thepervas (-ive), thattend (-), thatrout (-es), thatit(-), sucha (-), student-net (-), situationlimit (-s), short-rang (-e), severalalgorithm (-s)

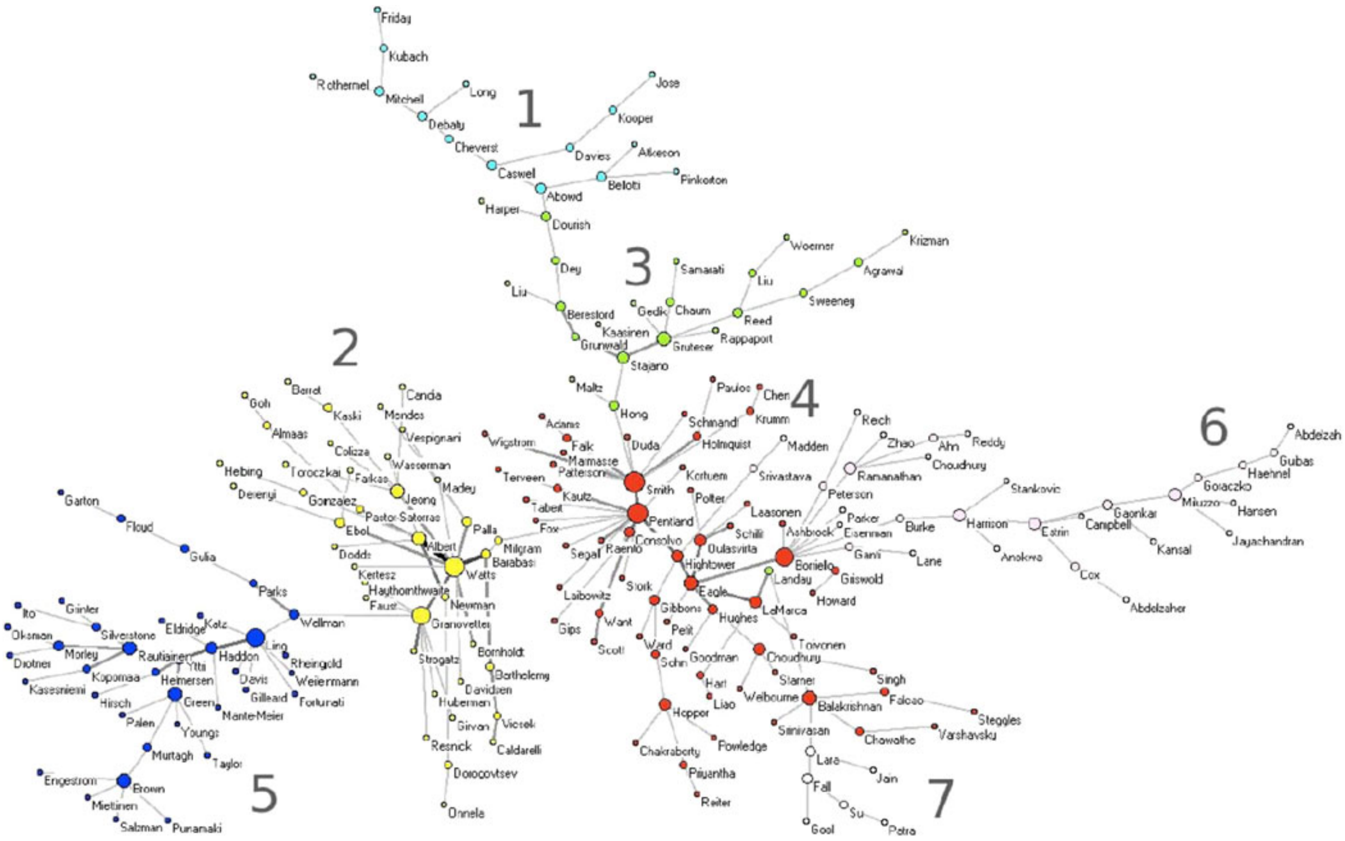

Fig. 2 Pathfinder Network of the largest component of the author co-citation network

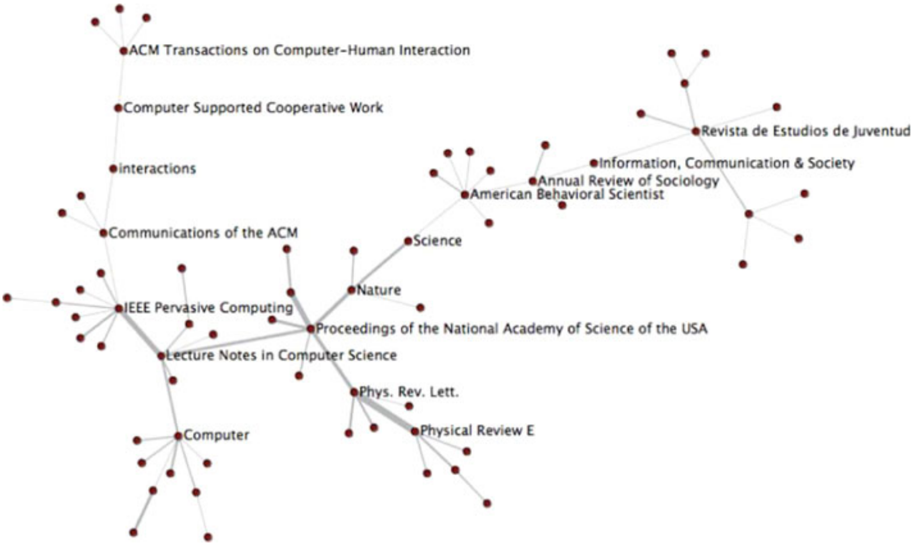
ACM Transactions on Computer-Human Interaction
Computer Supported Cooperative Work
interactions
Communications of the ACM
IEEE Pervasive Computing
Lecture Notes in Computer Science
Computer
Proceedings of the National Academy of Science of the USA
Nature
Science
Phys. Rev. Lett.
Physical Review E
American Behavioral Scientist
Annual Review of Sociology
Information, Communication & Society
Revista de Estudios de Juventud

Fig. 3 Pathfinder Network of the largest component of the journal-person network

Sloan Management Review, Human Computer Interaction with Mobile Devices & Services) further proceeding to other general computer topics (Lecture Notes of Computer Science, below without label: Computer, Mobile Networks and Applications, Mobile Computing and Networking, ACM Transactions on Database Systems, Information Systems) and to journals from physics (Proceedings of the National Academy of Science in the USA, Physical Review Letters, Advances in Complex Systems, Journal of Statistical Mechanics, Complexity, New Journal of Physics, The European Physical Journal) as well as to more general journals like Nature and Science. Starting from here, one reaches the area of journals characterized as belonging to the social sciences like American Behavioral Scientist (below and without label City & Community, Sociological perspectives, etc.), Annual Review of Sociology, Information, Communication and Society und Revista de Estudios de Juventud.

Not all nodes strictly follow this line, e.g. Sociological Methods & Research was placed beside IEEE Pervasive Computing.

As far as one assumes that the data set contains all relevant documents, authors and journals (or a significant part thereof) this form of presentation provides the opportunity to identify different research areas which might influence the scientific field of interest.

Analysis of the conference-person network

The network consisting of conferences and persons contains 121 components with more than one node. A component has 7.66 nodes on average (with a standard deviation of 17.29). The largest component includes 185 nodes (see Fig. 4). The most central conferences within this component are the International Conference on Ubiquitous Computing (Ubicomp 2003, 2005, 2007), the Conference on Human Computer Interaction (CHI 2005,

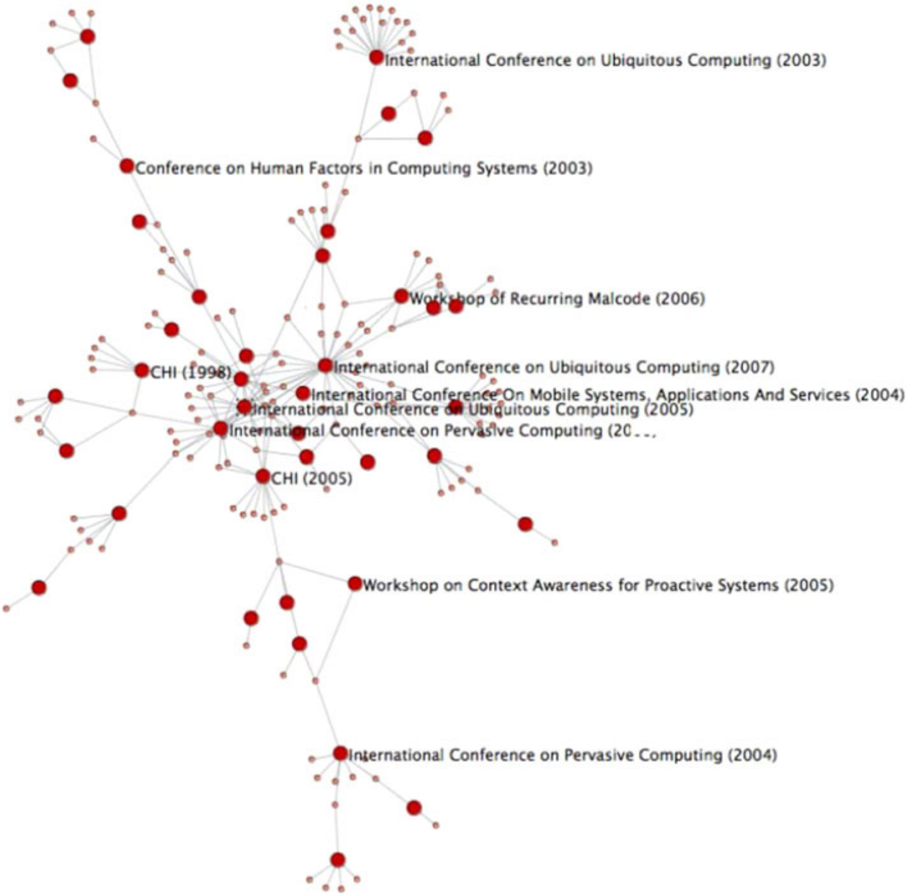
International Conference on Ubiquitous Computing (2003)
Conference on Human Factors in Computing Systems (2003)
Workshop of Recurring Malcode (2006)
CHI (1998)
International Conference on Ubiquitous Computing (2007)
International Conference On Mobile Systems, Applications And Services (2004)
International Conference on Ubiquitous Computing (2005)
International Conference on Pervasive Computing (20...
CHI (2005)
Workshop on Context Awareness for Proactive Systems (2005)
International Conference on Pervasive Computing (2004)

Fig. 4 Largest component of the conference-person network

2006), the conference Computer Supported Cooperative Work (CSCW 1996) and the International Conference on Pervasive Computing (Pervasive 2004).

Possible applications

Many feasible scenarios exist where a model like the one presented in this study can be useful. We suggest three different kinds of services:

- *overview services*: the primary goal is to get a comprehensive overview of the modeled scientific area
- *tracking services*: one or more nodes of the graph are separately tracked together with a history of their dynamics over a specific time-period. This type of service can unleash its full potential not until the update process of the model can be done automatically, because frequent regular manual model updates are uncomfortable and resource consuming for the user.

- *evolution services*: these kind of services try to explain the development in the model starting from a given point in time up to the current situation

The single types are discussed in detail now.

Overview services

Especially newcomers in a scientific field can have problems acquiring a broad overview of the area. The approach presented here can help to analyze the scientific area more in detail. In view of that, several steps can be taken:

- The author co-citation network can be used to locate the authors concerning their research interests. The representation as Pathfinder Network is—besides the application of a clustering mechanism—useful to get easily interpretable results. The author co-citation qualifies because the decision concerning the similarity of authors is done by the authors themselves. A disadvantage is that the picture drawn with author co-citation analysis only shows the past and not the present since it takes a while since a newly published paper won't get cited immediately.
- The collaboration of authors can be visualized using the co-authorship network. This graph can be used to identify definable schools within the scientific field (to get results which can be interpreted easily a clustering mechanism and a Pathfinder Network rendering can help).

A concrete service could allow the user to identify different sub-areas within the scientific field. To get an insight in the content of the sub-areas, the documents which are associated with the clusters can be analyzed using text analysis techniques (e.g. detect the specific words for each section using a vector space model and $tf - idf$). Using the eigenvector centrality, the group of relevant authors can be derived from the graph. The user has the option to focus his research on interesting sections only (ignoring the other subareas) or to do a more macro-orientated analysis and work on the whole graph.

The analysis of graphs with journals or conferences offers an insight in the communication patterns within the scene: from the perspective of the newcomer, this view on the graph can reveal interesting literature, the more experienced researcher can use this view to identify the best fitting journals for his/her articles. Apart from that, one can identify the related scientific fields which may influence the area of interest as demonstrated with the journal network above.

Using the citation graph it is easy to identify papers which cite a respectable amount of literature—the chance is high that these papers are review papers. Review papers are very helpful because they subsume the scientific discussion of a longer period and discuss the results in a bigger context.

The network of organizations can help active researchers to identify interesting organizations as career options: if an author specialized in a specific topic it would be favorable to work at an organization which already has influence in the specific area. For the topic *Mobile Social Networking*, e.g. Carnegie Mellon University, the MIT Media Lab or the group around Anthony LaMarca at Intel Research would be very attractive. For organizations or institutes, the organization graph or a derived coloring of the author co-citation network or co-authorship graph could be helpful for getting an impression of the role that the own organization plays in the respective scientific field. This kind of benchmark could be used to strengthen one's own position when negotiating financial research grants.

Beyond that, it is possible to identify other institutes in the same field, perhaps dealing with the same problems which could be interested in a cooperation.

Tracking services

If the underlying network can be updated automatically, tracking services can be implemented effectively. These services observe a set of nodes and document their development over time. Checks can be performed on a regular (e.g. daily) basis and could generate reports (which can be delivered to the user via e-mail). One could track which publications are presented on the important conferences or in which journals or on which conferences one's colleagues/friends/rivals are publishing/presenting.

Due to flexible parameter settings, upcoming top-researchers could be identified early by searching for authors who are new in the network but published a remarkable amount of articles which get cited heavily. How such metrics have to be designed in detail must be evaluated empirically.

Evolution-based services

Sometimes one might be interested in the development of the modeled scientific field as a whole. Again, for this type of services the network needs to be updated on a regular basis (manually or, preferably, in an automatic way). Such a service can be useful for people who have been working in the observed field for a certain time but had to interrupt their work for a longer period of time (e.g. for other projects, parental leave, sabbatical, etc.). The respective person actually can be assumed to possess a certain basic overview of the field, but this overview is not up to date (because the development of the last *x* years/months has not been observed). Potential views on the data are the author co-citation network and the co-authorship network when one is interested in a classification of persons/topics. A representation with a graph visualization tool like SoNIA (McFarland and Bender-deMoll 2009) might be helpful too, in order to get a high level overview of the changes in the area. SoNIA displays changes in a graph as an animation which documents the dynamics of the network on a step by step basis.

Summary and conclusions

For this contribution, a large dataset of articles from the domain Mobile Social Networking was collected and analyzed as a multi modal network. Different views on the network have been discussed, like the co-authorship network, the author co-citation network, the network between organizations and persons, journals and persons and conferences and persons. An approach for modeling the aging process of social relations was presented and evaluated. After that, a set of possible services were presented which could make the life of active researchers or graduate students easier.

This results lead to new questions, which should be answered in future studies. An important, but very time-intensive process is the collection of the articles which form the model of the regarded domain. This work has a very high demand on data quality, so all articles where collected manually using a specifically developed data base front end. It would be interesting if an automatic approach with lower data quality leads to similar results. Therefore, respective procedures and benchmark indicators have to be defined which allow a comparison of the results. If the results are similar to those presented in this article, an automatic update process of the data could be established and a lot work could be saved. Another interesting aspect is the influence on the data set caused by the user's search behavior: what happens, if the user forgets a specific search term? Does the network still converge to a comprehensive model for the domain? Therefore, the setup of the study has to be changed so that the origin of each article is saved. Afterwards, one can simulate the effect of missing one or more specific search terms.

This article attempts to perform a domain analysis based on social network analysis and hopefully contributes to the process of simplifying the exploration process of scientific fields.

References

Barabasi, A., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A, 311*, 590–614.

Baur, M., & Benkert, M. (2005). Network comparison. In U. Brandes, & T. Erlebach (Eds.), *Network analysis* (pp. 318–340). Berlin: Springer.

Bensman, S. J. (2004). Pearson's r and author cocitation analysis: A commentary on the controversy. *Journal of the American Society for Information Science and Technology, 55*(10), 935.

Björneborn, L. (2004). *Small-world link structures across an academic web space—A library and information science approach*. PhD Thesis, Royal School of Library and Information Science, Copenhagen, Denmark.

Brinkmeier, M., & Schank, T. (2005). Network statistics. In U. Brandes, & T. Erlebach (Eds.), *Network analysis* (pp. 293–317). Berlin: Springer.

Buzydlowski, J. W. (2003). *A comparison of self-organizing maps and pathfinder networks for the mapping of co-cited authors*. Ph.D. thesis, Drexel University.

Chen, C., & Carr, L. (1999). Trailblazing the literature of hyptertext: Author co-citation analysis (1989–1998). *Proceedings of the 10th ACM conference on hypertext and hypermedia.*

Chen, C., & Morris, S. (2003). Visualizing evolving networks: Minimum spanning trees versus pathfinder networks. In *Proceedings of IEEE symposium on information visualization* (pp. 67–74). IEEE Computer Society Press.

Chen, T. T., & Hsieh, L. C. (2007). On visualization of cocitation networks. *Proceedings of the 11th international conference information visualization* (pp. 470–475).

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E, 70* (066111).

Contextproject. URL http://www.cs.helsinki.fi/group/context/. [Online, 18. September 2009].

Corneil, D. G., & Gotlieb, C. C. (1970). An efficient algorithm for graph isomorphism. *Journal of the ACM, 17*(1), 51–64.

Egghe, L., & Leydesdorff, L. (2009). The relation between Pearson's correlation coefficient r and Salton's cosine measure. *Journal of the American Society for Information Science and Technology, 60*(5), 1027–1036.

Gaertler, M. (2005). Clustering. In U. Brandes, & T. Erlebach (Eds.), *Network analysis* (pp. 178–215).

Goffman, C. (1969). And what is your Erdös number? *The American Mathematical Monthly, 76*(7).

Groh, G., Hanstein, H., & Wörndl, W. (2009). Interactively visualizing dynamic social networks with dyson. In *Proceedings of the IUI'09 workshop on visual interfaces to the social and the semantic Web* (Vol. 2). Citeseer.

Hjørland, B. (2002). Domain analysis in information science—Eleven approaches—Traditional as well as innovative. *Journal of Documentation, 58*(4), 422–462.

Hjørland, B., & Albrechtsen, H. (1995). Toward a new horizon in information science: Domain-analysis. *Journal of the American Society for Information Science, 46*(6), 400–425.

Kaufmann, M., & Wagner, D. (2001). *Drawing graphs: Methods and models* (Vol. 2025). Springer.

Koschützki, D., Lehmann, K., Peeters, L., Richter, S., Tenfelde-Podehl, D., & Zlotowski, O. (2005). Centrality indices. In U. Brandes, & T. Erlebach (Eds.), *Network analysis* (pp. 16–61).

Kosub, S. (2005). Local density. In U. Brandes, & T. Erlebach (Eds.), *Network analysis* (pp. 112–142).

Lerner, J. (2005). Role assignments. In U. Brandes & T. Erlebach (Eds.), *Network analysis* (pp. 216–252).

Leydesdorff, L. (2008). On the normalization and visualization of author co-citation data: Salton's cosine versus the jaccard index. *Journal of the American Society for Information Science and Technology, 59*(1), 77–85.

Leydesdorff, L., & Vaughan, L. (2006). Co-occurrence matrices and their applications in information science: Extending aca to the web environment. *Journal of the American Society for Information Science and Technology, 57*(12), 1616–1628.

Lin, X., White, H. D., & Buzydlowski, J. (2003). Real-time author co-citation mapping for online searching. *Information Processing and Management, 39*(5), 689–706.

Lu, H., & Feng, Y. (2009). A measure of authors' centrality in co-authorship networks based on the distribution of collaborative relationships. *Scientometrics*, Online First.

Luukkonen, T., Tussen, R. J. W., Persson, O., & Sivertsen, G. (1993). The measurement of international scientific collaboration. *Scientometrics, 28*(1), 15–36.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval.

McCain, K. W. (1986). Co-cited author mapping as a valid representation of intellectual structure. *Journal of the American Society for Information Science, 37*(3), 111–122.

McCain, K. W. (1989). Mapping authors in intellectual space: Population genetics in the 1980s. *Communication Research, 16*, 667–681.

McCain, K. W. (1990). Mapping authors in intellectual space: A technical overview. *Journal of the American Society for Information Science, 41*(6), 433–443.

McCain, K. W., Verner, J. M., Hislop, G. W., Evanco, W., & Cole, V. (2005). The use of bibliometric and knowledge elicitation techniques to map a knowledge domain: Software engineering in the 1990s. *Scientometrics, 65*(1), 131–144.

McFarland, D., & Bender-deMoll, S. (2009). Sonia—Social network image animator URL http://www.stanford.edu/group/sonia/. [Online, 03. Oktober 2009].

Messmer, B. T., & Bunke, H. (1998). A new algorithm for error-tolerant subgraph isomorphism detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(5), 493–504.

Newman, M. E. J. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E, 64* (016131).

Newman, M. E. J. (2001b). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E, 64* (016132).

Newman, M. E. J. (2001c). The structure of scientific collaboration networks. In *Proceedings of the National Academy of Science of the USA, 98*, 404–409.

Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science, 28*(6), 441.

Park, H. W. (2003). Hyperlink network analysis: A new method for the study of social structure on the web. *Connections, 25*(1), 49–61.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program, 14*(3), 130–137.

Porter, M. A., Onnela, J. P., & Mucha, P. J. (2009). Communities in networks. URL http://ssrn.com/abstract=1357925. [Online, 13. Oktober 2009]

Rousseau, R., & Zuccala, A. (2004). A classification of author co-citations: Definitions and search strategies. *Journal of the American Society for Information Science and Technology, 55*(6), 513–529.

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network structures in proximity data. *The Psychology of Learning and Motivation: Advances in Research and Theory, 24*, 249–284.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265–269.

Small, H. (1981). The relationship of information science to the social sciences: A co-citation analysis. *Information Processing and Management, 17*(1), 39–50.

Small, H., & Sweeney, E. (1985). Clustering the science citation index using co-citations. I. A comparison of methods. *Scientometrics, 7*, 391–409.

Small, H., Sweeney, E., & Greenlee, E. (1985). Clustering the science citation index using co-citations. II. Mapping science. *Scientometrics, 8*, 321–340.

Thelwall, M. (2004). Social network analysis. In M. Thelwall, (Ed.), *Link analysis: An information science approach*. (See chapter 22 in URL http://linkanalysis.wlv.ac.uk/index.html), (pp. 213–217). Emerald Group Publishing Limited.

Thelwall, M., Vaughan, L., & Björneborn, L. (2005). Webometrics. *Annual Review of Information Science and Technology, 39*(1), 81–135.

Tsay, M. Y., Xu, H., & Wu, C. W. (2003). Author co-citation analysis of semiconductor literature. *Scientometrics, 58*(3), 529–545.

Ullmann, J. R. (1976). An algorithm for subgraph isomorphism. *Journal of the ACM, 23*(1), 31–42.

van Eck, N. J., & Waltman, L. (2008). Appropriate similarity measures for author co-citation analysis. *Journal of the American Society for Information Science and Technology, 59*(10), 1653–1661.

Wallace, M. L., & Gingras, Y. (2009). A new approach for detecting scientific specialties from raw cocitation networks. *Journal of the American Society for Information Science and Technology, 60*(2), 240–246.

White, H. D. (1990). Author co-citation analysis: Overview and defense. In C. L. Borgman (Ed.), *Scholarly communication and bibliometrics* (p. 85).

White, H. D. (2003a). Author cocitation and Pearson's r. *Journal of the American Society for Information Science and Technology, 54*(13), 1250–1259.

White, H. D. (2003b). Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists. *Journal of the American Society for Information Science, 54*(5), 423–434.

White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science, 32*(3), 163.

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science, 49*(4), 327–355.

Yan, E., Ding, Y., & Zhu, Q. (2009). Mapping library and information science in china: A coauthorship network analysis. *Scientometrics*, Online First, 157.

Zhao, D., & Strotmann, A. (2008). Information science during the first decade of the web: An enriched author cocitation analysis. *Journal of the American Society for Information Science and Technology, 59*(6), 916–937.