Publication Analysis of the Formal Concept Analysis Community

Stephan Doerfel, Robert Jäschke, Gerd Stumme

Knowledge & Data Engineering Group,
University of Kassel, Wilhelmshöher Allee 73, 34121 Kassel, Germany
http://www.kde.cs.uni-kassel.de/

Abstract. We present an analysis of the publication and citation networks of all previous editions of the three conferences most relevant to the FCA community: ICFCA, ICCS and CLA. Using data mining methods from FCA and graph analysis, we investigate patterns and communities among authors, we identify and visualize influential publications and authors, and we give a statistical summary of the conferences' history.

Keywords: bibliometrics, citation analysis, community, data mining, influence

1 Introduction

On the occasion of the 10th anniversary of the *International Conference on Formal Concept Analysis (ICFCA)* we are presenting a quantitative and qualitative analysis of all papers published at the previous editions of ICFCA. Additionally, we included the two related conference series *International Conference on Conceptual Structures (ICCS)* and *Concept Lattices and their Applications (CLA)* to extend the range of analyzed publications relevant to Formal Concept Analysis.

Being active members of the FCA community, our intention for this analysis was to gain more insights into the structure of our community and its relationship to closely related disciplines. We will address questions that every researcher is asking himself from time to time, such as

- Which are the most influential authors, papers, and conferences?
- Who is cooperating with whom on which topics?
- Who is citing whom?

We will target these and other questions on three different levels: on the conference level, the author level, and the paper level.

This paper will allow long-term participants of one or more of these conference series to gauge their perception about their community. It may also allow newcomers a faster access to the community by being pointed to the must-read papers and to the different schools of thought that are attending these conferences. Last but not least, we intend to spark further research about our community's structure. To this end, we publicly provide the dataset which is underlying this paper's analysis at http://www.kde.cs.uni-kassel.de/datasets/.

The structure of this paper is as follows: In the next section, we discuss related work. Section 3 describes the dataset of publications in detail. In Section 4, we briefly introduce the various analysis methods that we used. Section 5 provides the results of the analysis – this is the main contribution of this paper. Finally, in Section 6, we briefly address future work.

2 Related Work

The field of research we are dealing with in this paper is *bibliometrics*, the science of analyzing (scientific) literature. Subjects of analysis are, among others, the statistical and structural properties of citation or collaboration networks and measures of influence and impact of publications, authors, journals or conferences. Given the multitude of bibliometric publications it is difficult to provide the most relevant pointers. A good starting point are dedicated journals, e.g., the *Scientometrics* journal.

Some recent analyses with a focus on (parts of) computer science include [8] and [1]. In the latter the authors discuss graph properties like connectivity and degree distributions in the citation graph of a publication corpus. An analysis of collaboration networks including the discussion of community structure and the small-world phenomenon is given in [8]. Tilley and Eklund use FCA for a qualitative analysis of 47 publications from software engineering in [15]. They relate publications to software-related activities and classify them by the lines of code of a particular programming language, applied in the publications.

Poelmans et al. combine text mining and FCA to provide a survey on the FCA literature related to knowledge discovery [10] (140 publications) and information retrieval [11] (103 publications). Using a thesaurus of relevant terms, the retrieved papers are classified and visualized using a concept lattice. In the sequel the focus of both papers is a detailed survey of some of the publications under study. An early practical application of FCA to the management of literature is presented in [12], where meta data of publications is used to search and visualize a given publication corpus.

In contrast to these previous papers we neither focus on a detailed analysis of a small publication corpus, nor on a rough statistical analysis of a large scale corpus. The medium size of our corpus (954 publications with 17121 citations) still allows us to look at specific authors or publications. We provide the first analysis of the three conference series, in particular the first analysis with a focus on FCA that is applied next to such diverse methods as graph partitioning and ranking.

3 Dataset

We first describe how we collected the publication corpus and then define the data structures upon which our analysis is based.

Table 1. Venues of the three conference series.

ICCS	1993: Quebec City (CA), 1994: College Park (US), 1995: Santa Cruz (US), 1996: Sydney (AU), 1997: Seattle (US), 1998: Montpellier (FR), 1999: Blacksburg (US), 2000: Darmstadt (DE), 2001: Stanford (US), 2002: Borovets (BG), 2003: Dresden (DE), 2004: Huntsville (US), 2005: Kassel (DE), 2006: Aalborg (DK), 2007: Sheffield (UK), 2008: Toulouse (FR), 2009: Moscow (RU), 2010: Kuching (MY), 2011: Derby (UK)
ICFCA	2003: Darmstadt (DE), 2004: Sydney (AU), 2005: Lens (FR), 2006: Dresden (DE), 2007: Clermont-Ferrand (FR), 2008: Montreal (CA), 2009: Darmstadt (DE), 2010: Agadir (MA), 2011: Nicosia (CY)
CLA	2004: Ostrava (CZ), 2005: Olomouc (CZ), 2006: Hammamet (TN), 2007: Montpellier (FR), 2008: Olomouc (CZ), 2010: Sevilla (ES), 2011: Nancy (FR)

3.1 Gathering and Preprocessing

For our analysis we gathered meta data for all papers published at any of the past editions (up to 2011) of the three conference series ICCS, ICFCA, and CLA, i.e., 19 editions of ICCS, 9 editions of ICFCA, and 7 editions of CLA,[1] see Table 1. ICCS began as a conference on Conceptual Graphs (CG), with first FCA papers in 1995, and a balanced contribution of CG and FCA papers a few years later; while both ICFCA and CLA focus on FCA topics.

We collected data like paper titles, authors and their cited references from the publisher website SpringerLink[2] (ICCS and ICFCA) or extracted them from the paper's PDFs of CLA's website.[3] In our dataset, invited talks, regular and short papers are treated the same; poster sessions, satellite workshops as well as separate 'contributions' proceedings were not considered.

To gain knowledge about publications citing any of the conference papers, we retrieved citations from Microsoft Academic Search.[4] Note that these citations only roughly reflect the real number of citations a publication received, since this search engine relies on citation data that is available on the web and can only to a certain extent remove errors and correctly match different citation variants.

Our preprocessing included the extraction of authors, titles, years, and references from HTML and PDF files using regular expressions and manual work. Further, we implemented several normalization and completion steps for the titles and author names to allow matching and duplicate detection and an extensive manual error correction. Therefore, we employed the normalization steps described in [16] with an additional removal of diacritics (e.g., 'ä' and 'á' were replaced by 'a'). We used different heuristics, e.g., the Levenshtein distance, to

[1] The first edition of the CLA 2002 in Horní Bečva was a small seminar with four talks and hence no published proceedings exist.

[2] `http://www.springerlink.com/`

[3] `http://cla.inf.upol.cz/papers.html`

[4] `http://academic.research.microsoft.com/`

find errors in author names and titles. All references without authors (often encountered for cited web pages) were removed from the dataset.

Since many publications were cited as different editions or prior to their publication ('to appear'), we normalized the publication year by dating back different editions to the earliest mentioned date of publication. For example, the collected papers of Charles S. Peirce [47] were cited with different publication years (1931, 1935, 1953, 1958, 1966) which we normalized to 1931.

For the first ICFCA 2003 in Darmstadt no proceedings were published. Thus, we used the book from 2005 [33] which contains contributions from the participants of the first ICFCA on the state of the art on FCA and its applications.

Finally, we would like to point out that – since the focus of our analysis is on the three conference series – many publications related to FCA (in particular journal articles) have not been included in the dataset. The results presented in this paper should be interpreted with this fact in mind.

3.2 Notations and Derived Data Structures

From the collected data we derived several structures (graphs and formal contexts) that are described in detail in the following. All structures that use the references were created after removing self-citations (cited publications where one of the authors is also an author of the citing paper).

We denote the set of all authors that published at any of the three conferences by A and the set of all papers published at any of the conferences by P.

Authorship. The formal context $\mathbb{K}_{\text{pa}} = (P, A, I_{\text{pa}})$, with $(p, a) \in I_{\text{pa}}$ iff a is an author of paper p, describes who authored which publication.

The graph of co-authorship $\mathfrak{G}_{\text{coa}}$ is an undirected, weighted graph with A as node set. Two authors are connected, iff they published together and their edge's weight is the number of co-authored publications at the conferences.

In Section 5.2, we cluster (partition) $\mathfrak{G}_{\text{coa}}$ and use these clusters as attributes of formal contexts. We denote by $C_n(\mathfrak{G}_{\text{coa}})$ the set containing the n clusters with the highest cardinality.

Citations. The directed, weighted graph $\mathfrak{G}_{\text{cit}}$ again has the authors in A as nodes. An edge (a, b) with weight w indicates that in all considered publications, w times, some publication of b was referenced by a.

Conferences. To analyze the distribution of all authors over the three conference series, we use $\mathbb{K}_{\text{conf}} = (A, \{ICCS, ICFCA, CLA\}, \mathbb{N}, I_{\text{conf}})$, a many-valued context where $(a, c, n) \in I_{\text{conf}}$, iff a published exactly n papers at conference c.

4 Definitions and Methodology

In this section, we give a brief overview of the different algorithms and methods we use in our analysis. Most of the FCA notions are explained in great detail in

the textbook [5]. In Section 5.2, we discuss the extents of an *iceberg lattice* of a context, i.e., an ordered subset of the concept lattice containing only concepts with extents larger (w.r.t cardinality) than a given threshold (minimum support). Iceberg lattices and a construction algorithm are explained in [13].

In the same section, we analyze *communities* of co-authorship. Intuitively, communities are certain subsets of some larger set of entities, such that the members of a subset are somewhat more related or similar to each other than they are to others. There is, however, no generally accepted formal definition of the notion of a "community". Here, by communities we mean the classes of a partitioning on the node set of a given graph. To create such a partitioning and its visualization for the co-authorship graph $\mathfrak{G}_{\text{coa}}$, we laid out the graph using the force directed graph visualization provided by Graphviz [4]. Then the *GMap* algorithm (again Graphviz) based on [9] was applied to discover communities of collaborators. GMap optimizes its output clustering w.r.t. *modularity*, which is a community quality measure that compares the number of co-author edges within each community to the expected value for this number in an equivalent random graph. Finally, *Voronoi diagrams* are used to draw the 'borders' between the different 'countries'.

In Section 5.2, we also apply different node centrality measures which indicate the importance of nodes within the citation graph $\mathfrak{G}_{\text{cit}}$. Next to the simple measures *in-degree* (number of edges pointing towards a node) and *in-strength* (sum of the weights of all edges pointing towards a node), we use *PageRank* [2] to rank authors of the conferences. PageRank is an eigenvector-based measure that was originally developed to measure the importance of web pages according to the link structure of the World Wide Web. To assign a score to each node in a graph, a linear equation system is solved which integrates the adjacency matrix of the graph and a probabilistic component. The main idea of the ranking is that important nodes are pointed to by other (important) nodes. In our scenario of citations, an author is considered important (i.e., has a high PageRank), if he or she is cited by many other important authors.

Based on a similar idea, the (also eigenvector-based) *HITS algorithm* [6] determines *hubs* and *authorities* in a graph. Roughly speaking, hubs are nodes that point to many good authorities in the graph. Authorities are those nodes that are referenced by many good hubs. In the citation graph, an author is a good hub, if he or she references many authors that have high values as authorities (e.g., authors of survey papers). Of interest for us, however, are the authorities, i.e., authors that have been cited by authors with high hub values.

5 Results

Now, we present the results of our analysis along the three dimensions of conferences (Section 5.1), authors (Section 5.2), and publications (Section 5.3).

5.1 Conferences

We start the section on conferences by some basic statistics (cf. Tables 2 and 3) that give an overview of the conference history. The two lower blocks of Table 2

Table 2. The history of the three conference series in numbers.

	ICCS	ICFCA	CLA	total
editions	19	9	7	35
publications	567	208	179	954
avg. publications per edition	29.84	23.11	25.57	27.26
authors	542	218	269	872
avg. publications per author	2.04	1.94	1.62	2.25
'outgoing' citations (publications that have been cited by the conferences' papers)				
citations	10131	4328	2662	17121
cited authors	5871	2655	2027	8513
cited publications	6079	2406	1668	8813
self-citations	2255 (≈22 %)	965 (≈22 %)	529 (≈20 %)	3749 (≈21 %)
'incoming' citations (conference papers that have been cited)				
citations	3202	1322	153	4677
citing publications	1776	985	134	2522
cited publications	404 (≈71 %)	128 (≈62 %)	47 (≈26 %)	579 (≈61 %)

show statistics for two types of citations: 'outgoing', i.e., citations we extracted from the conference papers, and 'incoming', i.e., publications that cite one of the papers published at one of the conferences. The fraction of 20–22 % self-citations is comparable to or lower than prior results (e.g., [14] reports 38 % for mathematical publications). The lower fraction of publications at ICFCA and CLA that have been cited (last row) can partly be explained by the young age of these two conferences.

Table 3. The top five contributing authors of each conference. In case of a tie all authors with the same number of publications are listed.

ICCS	ICFCA	CLA	total
R. Wille (24)	R. Wille (14)	S. Ben Yahia (13)	R. Wille (42)
G.W. Mineau (19)	P. Eklund (11)	R. Bělohlávek (11)	S.O. Kuznetsov (27)
J.F. Sowa (14)	P. Valtchev (10)	A. Napoli (10)	P. Eklund (26)
S.O. Kuznetsov (13)	B. Ganter (10)	E. Mephu Nguifo (8)	B. Ganter (24)
M. Keeler (13)	S.O. Kuznetsov (8)	V. Vychodil (7)	P. Valtchev (20)
	S. Ferré (8)	M. Huchard (7)	G.W. Mineau (20)
	L. Nourine (8)	J. Outrata (7)	

Publication Habits. To gain insights into the publication habits we consider the many-valued context $\mathbb{K}_{\text{conf}}$. Through conceptual scaling this context is transformed into the single-valued context

$$\mathbb{K}_{\text{freq}} = (A, \{CLA, ICCS, ICFCA, 3 \times CLA, 3 \times ICCS, 3 \times ICFCA\}, I_{\text{freq}})$$

where each author coincides with a conference if he or she published there at least once. An author incides with one of the other three attributes if he or she published at the corresponding conference at least three times. The threshold of three was selected since publishing three times at the same conference series indicates already a certain commitment to it. On the other hand, we did not set a higher value, since especially CLA and ICFCA are young conferences (seven and nine editions, resp.). The line diagram of the context's concept lattice is depicted in Figure 1, where the values below each concept count the number of authors in the concept extent (support values). Exemplarily, the top contributing authors from Table 3 are annotated at their object concepts. To interpret the lattice, one

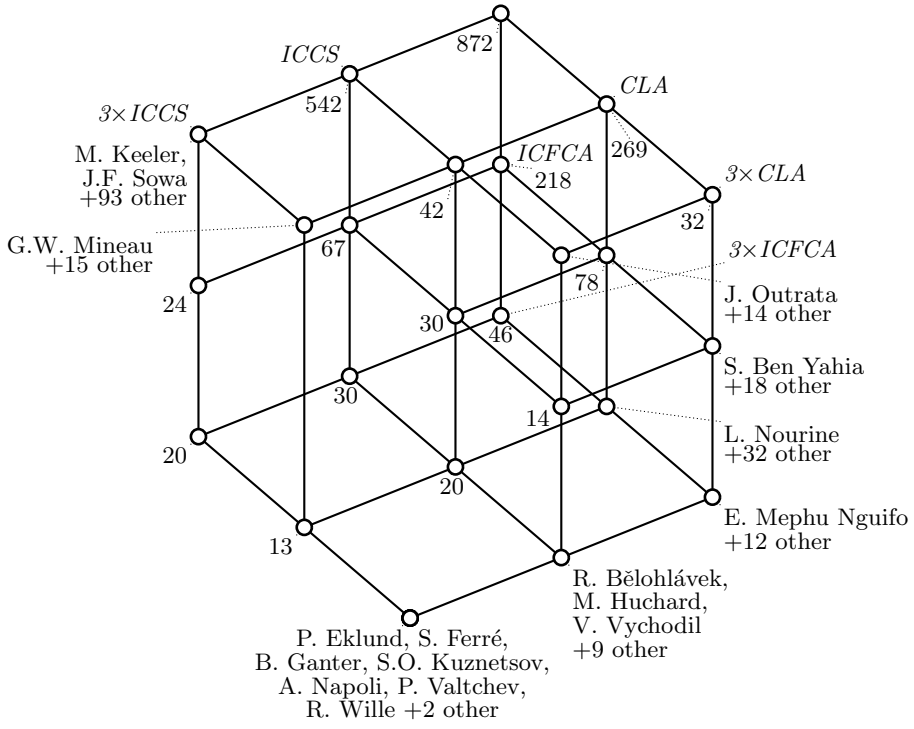


Fig. 1. The concept lattice for the author-conference context $\mathbb{K}_{\text{freq}}$, annotated with support values and the top contributing authors mentioned in Table 3.

has to keep in mind that ICCS runs more than twice as long as the other two conference series, naturally resulting in higher author participation: 542 authors

vs. 218 (ICFCA) and 269 (CLA). Of the 872 authors, 127 (14.6 %) published at least at two and 30 (3.4 %) of them at all three conference series.

The Duquenne-Guigues base of implications contains – aside from the trivial rules resulting from the choice of scales – only two rules:

1. 3×ICCS and 3×CLA $\Longrightarrow$ 3×ICFCA
2. 3×ICCS and ICFCA and CLA $\Longrightarrow$ 3×ICFCA.

The first rule states that any author who frequently published at both *ICCS* and *CLA* also frequently published at ICFCA. Similar rules do not hold for the other combinations of conferences. However, several association rules with high confidence further confirm the bonds between the three conferences. The following list contains those rules with a confidence greater or equal to 80% (each given with its absolute support and confidence):

1. 3×CLA and ICCS $\Longrightarrow$ ICFCA (15/93 %)
2. 3×CLA and 3×ICFCA $\Longrightarrow$ ICCS (13/92 %)
3. 3×CLA and ICCS and ICFCA $\Longrightarrow$ 3×ICFCA (14/86 %)
4. 3×ICCS and ICFCA $\Longrightarrow$ 3×ICFCA (24/83 %)
5. 3×ICCS and CLA $\Longrightarrow$ 3×ICFCA (16/81 %).

Roughly speaking, these rules express the fact that many authors who frequently published a paper at ICCS or CLA also (frequently) published a paper at ICFCA.

Author Fluctuation. Now, we want to answer the question, *How many new authors can the conferences attract each year?* Therefore, we investigate for each year which fraction of authors of all accepted publications is 'new', i.e., has never before published a paper at the corresponding conference. As can be seen in Figure 2, for the first edition of each conference this fraction naturally is equal to 1 and has a decreasing trend for the immediately following years. On the contrary, the fraction of authors that appeared at a conference for the 'last' time (negative bars) naturally increases to -1 for last year's conferences. Therefore, we omitted the first (last) two editions of each conference for the calculation of the mean first (last) fractions. For all three conferences, on average, over half of the authors never published before at the conference. We conclude that the conferences are able to attract new authors each year. Similarly, on average, half of the authors did not publish again. Thus, there is a considerable exchange of authors and possibly ideas. For CLA, both values are considerably higher, meaning that this young conference still has a high fluctuation rate. Another observation is the steady increase of newcomers in the years from 2003 to 2007 for ICCS, followed by a sharp drop in 2008. This is also reflected by the absolute counts (not shown here) that drop from 58 'newcomers' in 2007 to only 15 in 2008 and the similar behaviour for those years with the 'last' authors. One explanation is given by the absolute numbers of authors for these years: 90 (2007) and 47 (2008), i.e., a decrease by a factor of two. Nevertheless, this might not be the only explanation, since in the following year 2009 only 40 authors published at ICCS but both the fraction of 'newcomers' and 'lasttimers' increases. We could not find a convincing explanation for this phenomenon, but plan to specifically compare the collaboration graphs of these years.

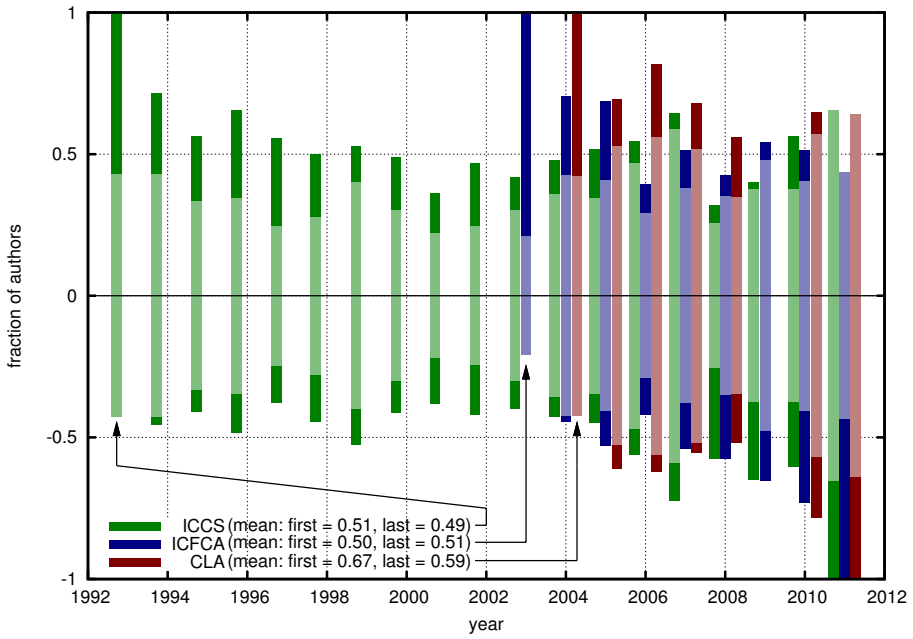


Fig. 2. Fluctuation of authors for each conference. The dark positive (negative) bars depict the fraction of authors that submitted a paper to the corresponding conference for the first (last) time in that year. The light bars in front of them depict the fraction of authors for which that year was also the only year (up to now) they submitted a paper (note that this measure is symmetric with respect to 'first' and 'last'.) For the calculation of the mean values for first (last), the first (last) two editions of each conference were omitted.

5.2 Authors

We analyze collaboration and influence between the authors of the conferences.

The Structure of the Community. First, we take a look at the co-authorship structure of the conferences. The most frequent collaborators can be read off from an iceberg lattice (frequent closed itemsets) of the publication-author-context $\mathbb{K}_{\text{pa}}$. Setting for instance the minimum support (minimum number of publications) to six, the following ten pairs[5] constitute the only (non singleton) intents of the iceberg lattice (given with their absolute support):[6] R. Bělohlávek/V. Vychodil (10), S. Ferré/O. Ridoux (9), J. Ducrou/P. Eklund (8), M.R. Hacene/ P. Valtchev (8), P. Øhrstrøm/H. Schärfe (8), R. Godin/P. Valtchev (7), E. Mephu

[5] The fact that only pairs show up indicates that there were no teams of three or more authors who published more than six papers together.

[6] We do not show the iceberg lattice, due to space restrictions, and to the fact that it is structurally just an anti-chain.

Nguifo/S. Ben Yahia (7), M. Ducassé/S. Ferré (6), B. Ganter/S.O. Kuznetsov (6) and T. Hamrouni/S. Ben Yahia (6). Using a lower minimum support threshold of 4 yields another 12 concepts with 5 publications and 8 concepts with 4 publications in the extent. Among them are three concepts with intents containing more than just two authors: P. Cellier/M. Ducassé/S. Ferré (5), T. Hamrouni/E. Mephu Nguifo/S. Ben Yahia (5) and M.R. Hacene/M. Huchard/P. Valtchev (4).

The co-author graph $\mathfrak{G}_{\text{coa}}$ reveals interesting patterns of collaboration within and between the FCA and CG (Conceptual Graphs) communities. The map in Figure 3 shows a clustering created by GMap [3]. Connected components that contain less than four authors or that are based on less than four papers have been omitted for the sake of legibility. The width of the edges between two co-authors reflects the number of publications they have written together at any of the three conferences; similarly, the size of the author names depicts the number of published papers.

The giant connected component (GCC) of the graph is divided into 13 clusters (1–13) and contains 314 of the 482 authors shown on the map. The second largest component (clusters 14 and 15) contains the second largest cluster (14) with 52 members mostly belonging to the Conceptual Graph (CG) subcommunity that is based in France. The remaining five large clusters (with more than ten members) are not connected. Based on our knowledge of the community they can roughly be classified to belong to the CG community (clusters 17–19) and to the FCA community (clusters 16 and 20). Adepts of the conferences can discover many further interesting aspects in this collaboration graph. Due to space restrictions we only want to outline that the CG community forms more separate clusters than the FCA community. Besides the five mentioned separate clusters, we consider only three of the 13 clusters of the GCC to be part of the core CG community (clusters 4, 5, and 9). Except for cluster 10 (the Description Logics community) all remaining clusters of the GCC belong to the FCA community. Finally, we would like to point out the remarkable role of G. Mineau (cluster 5) as a bridge between two CG clusters and the FCA community.

Topics of the Clusters. To get an idea about the topics that the authors of single clusters deal with, we visualize their citations of the most often cited publications and authors in two concept lattices (Figure 4). For legibility, we restrict this analysis to the set C_8 of the eight largest clusters (each containing more then 24 authors, while the others contain at most 14 authors), i.e., the clusters 1–7 and 14. Many different ways of choosing attribute sets and incidence relations are conceivable and it would be interesting to observe the influence of these choices. In this paper, we choose the following two examples for visualization: We construct the contexts $\mathbb{K}_{\text{cp}} = (C_8, P_{20}, I_{\text{cp}})$ and $\mathbb{K}_{\text{ca}} = (C_8, A_{C5}, I_{\text{ca}})$. Hereby, the set P_{20} contains the 20 most often cited publications of the corpus. In contrast to that, the set A_{C5} contains for each of the eight clusters the top five authors w.r.t to the number of papers – with at least one author from the cluster – that reference them. A cluster c is set in relation with a publication p

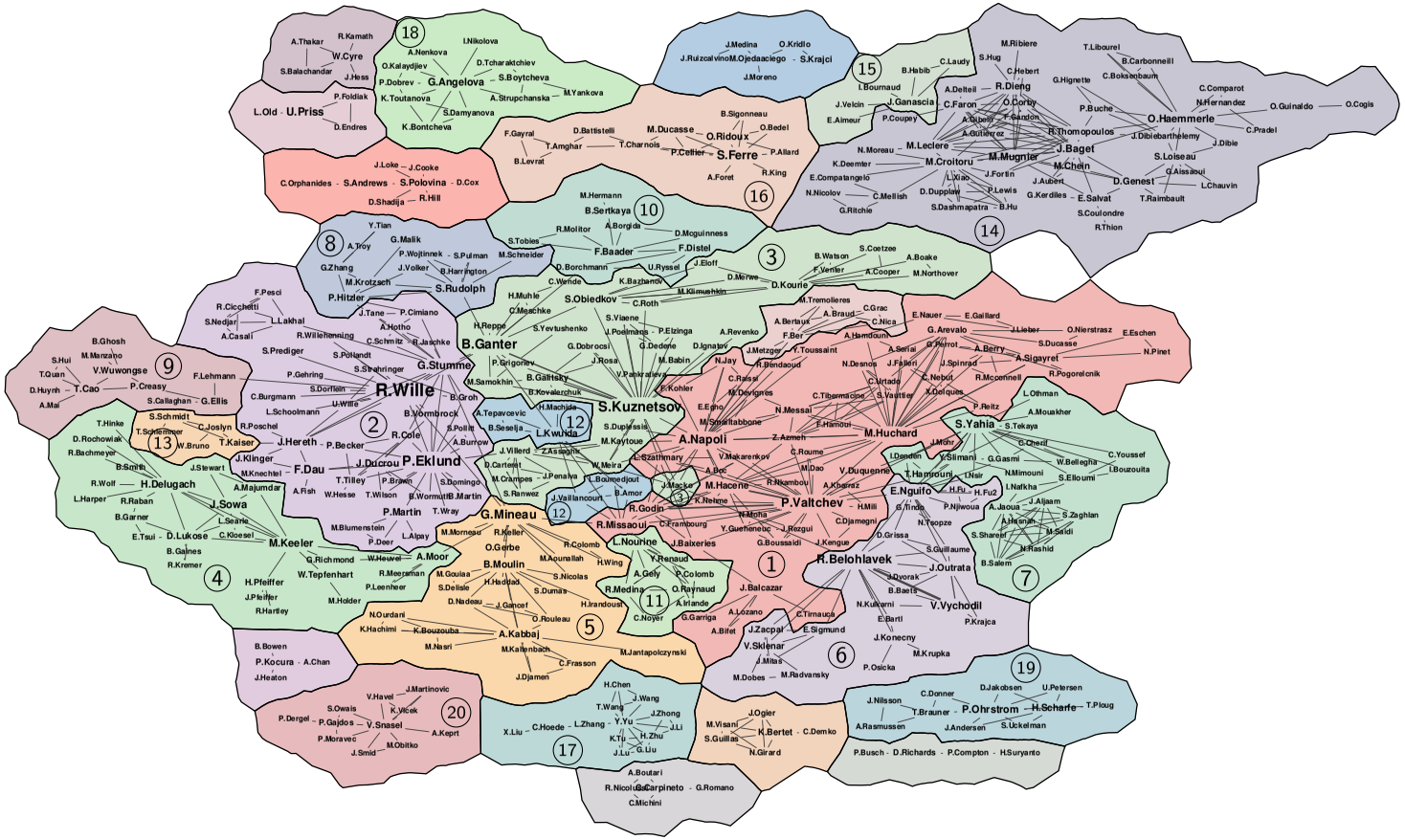


Fig. 3. A map of the co-author graph. Isolated 'islands' with less than four publications or less than four authors have been removed.

(an author a), if p (a) is cited by at least three (five) papers from c. Figures 4(a) and 4(b) show the resulting lattice diagrams.

Both lattices seem to reflect the two main schools of the considered conferences: FCA and CG. Each cluster cites one of their cornerstone-publications ([60] and [54]) and their creators (R. Wille and J.F. Sowa). Clearly, clusters 1, 6 and 7 belong to the FCA community and clusters 4 and 14 to the CG community, while 2, 3 and 5 cite publications and authors from both. The philosophical foundations of C.S. Peirce are important for clusters 2 and 4. In the FCA community, we can see the high impact of the foundations book [5] by B. Ganter and R. Wille and of papers on implications and association rules. The topics of the papers further suggest that clusters 2 and 4 might be more interested in mathematical and philosophical foundations while clusters 1, 6 and 7 often cite important algorithmic publications.

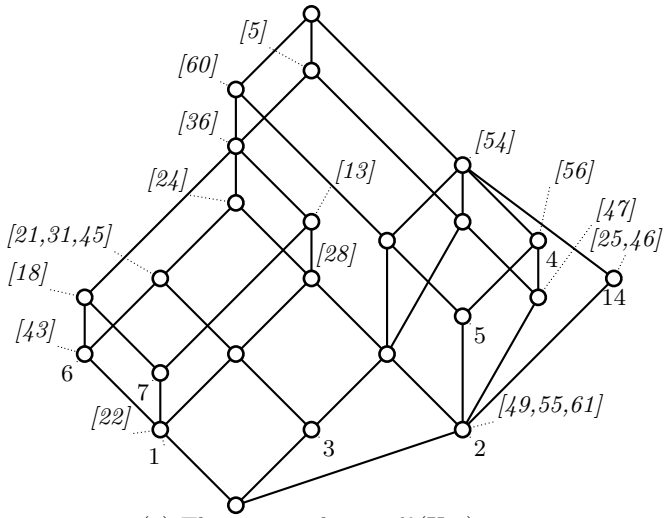
Table 4. Top ten rankings for the network analysis measures in-degree, in-strength, PageRank and authority (HITS, cf. Section 3.2) in $\mathfrak{G}_{\text{cit}}$.

	in-degree		in-strength		PageRank		authority	
1	R. Wille	443	R. Wille	1877	J.F. Sowa	.101	R. Wille	.161
2	B. Ganter	424	B. Ganter	1322	R. Wille	.068	B. Ganter	.087
3	J.F. Sowa	307	J.F. Sowa	1033	B. Ganter	.043	G. Stumme	.042
4	G. Stumme	211	G. Stumme	570	M.-L. Mugnier	.021	L. Lakhal	.031
5	R. Godin	156	M.-L. Mugnier	427	M. Chein	.020	J.F. Sowa	.030
6	S.O. Kuznetsov	151	L. Lakhal	412	G. Ellis	.017	S. Prediger	.023
7	R. Missaoui	134	R. Godin	374	G. Stumme	.014	M.J. Zaki	.019
8	G.W. Mineau	128	M. Chein	360	O. Gerbé	.014	R. Godin	.019
9	L. Lakhal	127	S.O. Kuznetsov	349	S. Prediger	.013	S.O. Kuznetsov	.018
10	P. Eklund	124	C. Carpineto	264	G.W. Mineau	.011	C. Carpineto	.017

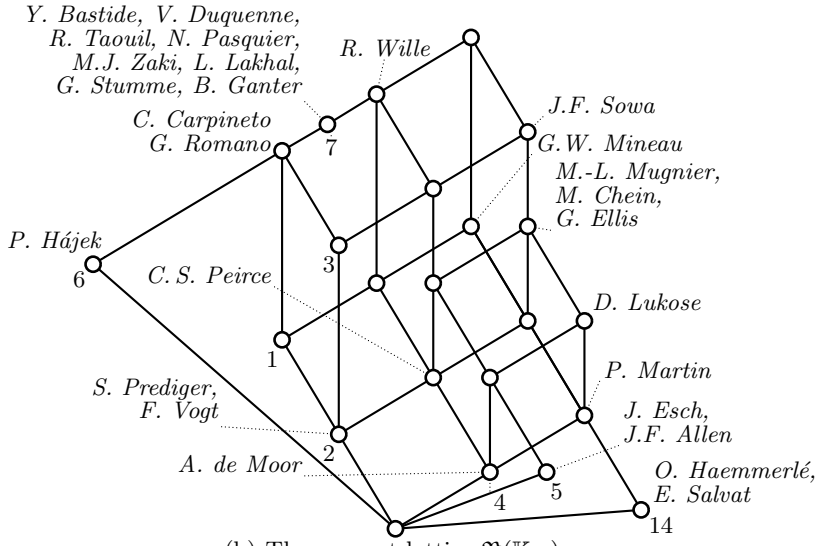
Influence. Finally, we use the author-citation graph $\mathfrak{G}_{\text{cit}}$ to identify key players, i.e., authors that are the most influential or the most central in the graph. Several centrality measures have been proposed (see, e.g., [7]). In Table 4 we present four rankings according to the different measures described in Section 4. One can observe that the different measures show a strong agreement. Note, that the scores are only valid within the investigated community of the three conferences, since we did only consider citations from papers published there. Thus, these figures do not make a general statement about the importance of the authors.

5.3 Publications

In this section, we take a closer look on individual publications and their citations. For each conference the first four rows of Table 5 list cited publications



(a) The concept lattice $\underline{\mathfrak{B}}(\mathbb{K}_{\mathrm{cp}})$.



(b) The concept lattice $\underline{\mathfrak{B}}(\mathbb{K}_{\mathrm{ca}})$.

Fig. 4. The two lattices relate the eight largest clusters from Figure 3 as objects to the most often (in conference papers) cited publications and authors as attributes. The eight clusters are: 1 (P. Valtchev, A. Napoli, A.M.R. Hacene, ...), 2 (R. Wille, P. Eklund, F. Dau, ...), 3 (S.O. Kuznetsov, B. Ganter, S. Obiedkov, ...), 4 (J.F. Sowa, H.S. Delugach, M. Keeler, ...), 5 (G.W. Mineau, B. Moulin, A. Kabbaj, ...), 6 (R. Bělohlávek, V. Vychodil, E. Mephu Nguifo, ...), 7 (S. Ben Yahia, T. Hamrouni, Y. Slimani, ...) and 14 (J.-F. Baget, O. Haemmerlé, M.-L. Mugnier, ...).

and citation counts and the top most cited publications for each conference and for the set of all sources other than the three conference series.

The most often cited paper of ICCS at ICCS [61] paved the way for a connection of the two schools of research that are the foundation of ICCS, namely *Formal Concept Analysis* and *Conceptual Graphs*. As a general observation, the most often cited papers from ICCS are theory-minded, the most important papers from ICFCA equally present theory and applications of and for FCA. The most often cited papers from other sources include publications belonging to the foundations of the disciplines FCA [5,36,60] and CG [54].

While the first four rows of the table reveal the most important publications of and from each community, we take a closer look at the theoretical foundations of the conferences in its last row. It contains the most cited publications only from authors that never attended any of the conferences. Naturally, this excludes the well-known foundation papers of Ganter, Wille, Sowa, etc., but it reveals onto which (other) theories the conferences' main results are built. We can see a clear agreement between CLA and ICFCA about the most important foundational publication for both conferences, namely the book by Birkhoff [21]. Furthermore, *association rule mining* was an important topic at both conferences. For the ICCS – as one would assume – three publications of Peirce are the most often cited 'external' publications. Interestingly, the paper that laid the foundation for the *Semantic Web* [20] is the third most important paper in this category. This shows the influence of the Semantic Web community on the ICCS community.

6 Future Work

In this paper, we have analyzed the citation and collaboration behaviour of authors of the three FCA-related conferences ICCS, ICFCA, and CLA. The picture of the FCA community could be completed by adding further publications from journals and books. Finding relevant publications and retrieving their metadata and citations is clearly a first step for future work.

Since we intended to give a broad overview of many different aspects of the community, we naturally chose not to go into too much detail with only one specific aspect of the performed analyses. Each analysis could be extended to a comparison of different settings or methods, e.g., one might try different clustering algorithms to validate the communities found in Section 5.2. Therefore, with respect to space and time constraints, we did only deal with some of the questions relevant for the community and for newcomers. For example, the highly interesting structure of the FCA community that can be read off the co-author graph presented in Section 5.2 could be investigated further. Which kind of subcommunities exist? Which authors are bridges between different communities? Can roles like student, supervisor, etc. be identified? We also plan to validate our ad-hoc assignment of community labels by analyzing the titles and abstracts of the authors' papers. Thereby, it would be possible to explicitly assign authors to topics and thus get a clearer picture of how the community is constituted.

Table 5. The most often cited papers of a certain conference by papers of another conference. The first line of each cell reflects the number of cited papers and the number of citations. The following lines point to the top three citations, the first is given with title.

papers from this conference were cited	papers have been cited at this conference: ICCS	ICFCA	CLA
ICCS	249 publications in 737 citations 26× *Conceptual graphs and formal concept analysis* [61] 19× [53], 16× [50]	66 publications in 192 citations 11× *Pattern structures and their projections* [32] 11× *Conceptual graphs and formal concept analysis* [61] 11× *Boolean concept logic* [62]	33 publications in 51 citations 5× *Pattern structures and their projections* [32] 4× [42], 4× [44]
ICFCA	38 publications in 60 citations 6× *The ToscanaJ suite for implementing conceptual information systems* [19] 5× [59], 4× [27], 4× [26]	60 publications in 120 citations 6× *The ToscanaJ suite for implementing conceptual information systems* [19] 6× *Formal concept analysis for knowledge discovery and data mining: the new challenges* [58] 5× [27], 5× [35], 5× [41], 5× [30], 5× [57]	63 publications in 100 citations 5× *Machine learning and formal concept analysis* [40] 5× *Formal concept analysis for knowledge discovery and data mining: the new challenges* [58] 5× *Analysis of social communities with iceberg and stability-based concept lattices* [37]
CLA	10 publications in 10 citations (at most one citation per paper)	11 publications in 13 citations 2× *What is a fuzzy concept lattice?* [23] 2× *Camelis: Organizing and browsing a personal photo collection with a logical information system* [29]	19 publications in 31 citations 3× *What is a fuzzy concept lattice?* [23] 3× *Towards concise representation for taxonomies of epistemic communities* [52] 3× *The basic theorem on generalized concept lattice* [39] 3× *Parallel recursive algorithm for FCA* [38]
other	4686 publications in 7069 citations 284× *Conceptual structures: information processing in mind and machine* [54] 100× [5], 65× [56]	1877 publications in 3038 citations 139× *Formal concept analysis: mathematical foundations* [5] 32× [60], 26× [36]	1218 publications in 1951 citations 124× *Formal concept analysis: mathematical foundations* [5] 30× [60], 24× [36]
external	3674 publications in 4708 citations 43× *Collected papers* [47] 19× [51], 15× [20], 15× [48]	1304 publications in 1741 citations 25× *Lattice theory* [21] 14× [45], 12× [17]	825 publications in 1041 citations 16× *Lattice theory* [21] 12× [22], 11× [17]

A dimension we could not analyze in the scope of this paper is *time*. Such an analysis would reveal developments and trends of the conferences. It could also allow us to judge the vitality of the communities in the co-author graph.

We would like to invite interested researchers to collectively tackle the above-mentioned challenges. The dataset is freely available,[7] extensions and error corrections are welcome and will be added to the dataset's web page. The metadata of all publications referenced in this paper is available in BibSonomy at `http://www.bibsonomy.org/group/kde/citedBy:doerfel2012publication`.

Acknowledgement. Part of this research was funded by the DFG in the project "Info 2.0 – Informationelle Selbstbestimmung im Web 2.0".

References

1. Y. An, J. Janssen, and E. E. Milios. Characterizing and mining the citation graph of the computer science literature. *Knowledge and Information Systems*, 6(6):664–678, Nov. 2004.
2. S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117, 1998.
3. E. R. Gansner, Y. Hu, and S. G. Kobourov. GMap: Drawing graphs as maps. *cs.CG*, arXiv:0907.2585v1, July 2009.
4. E. R. Gansner and S. C. North. An open graph visualization system and its applications to software engineering. *Software Practice & Experience*, 30(11):1203–1233, Sept. 2000.
5. B. Ganter and R. Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer, Berlin/Heidelberg, 1999.
6. J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *J. ACM*, 46:604–632, September 1999.
7. D. Koschützki, K. Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl, and O. Zlotowski. Centrality indices. In U. Brandes and T. Erlebach, editors, *Network Analysis*, volume 3418 of *LNCS*, pages 16–61. Springer, Berlin/Heidelberg, 2005.
8. M. E. J. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404–409, 2001.
9. M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.
10. J. Poelmans, P. Elzinga, S. Viaene, and G. Dedene. Formal concept analysis in knowledge discovery: A survey. In M. Croitoru, S. Ferré, and D. Lukose, editors, *Conceptual Structures: From Information to Intelligence*, volume 6208 of *LNCS*, pages 139–153. Springer, Berlin/Heidelberg, 2010.
11. J. Poelmans, P. Elzinga, S. Viaene, G. Dedene, and S. O. Kuznetsov. Text mining scientific papers: a survey on FCA-based information retrieval research. In P. Perner, editor, *Industrial Conference on Data Mining - Poster and Industry Proceedings*, pages 82–96. IBaI Publishing, 2011.
12. T. Rock and R. Wille. Ein TOSCANA–Erkundungssystem zur Literatursuche. FB4-Preprint 1901, TH Darmstadt, 1997.

[7] `http://www.kde.cs.uni-kassel.de/datasets/`

13. G. Stumme, R. Taouil, Y. Bastide, N. Pasquier, and L. Lakhal. Computing iceberg concept lattices with titanic. *Data & Knowledge Engineering*, 42(2):189–222, 2002.
14. B. Thijs and W. Glänzel. The influence of author self-citations on bibliometric meso-indicators. the case of european universities. *Scientometrics*, 66(1):71–80, 2006.
15. T. Tilley and P. Eklund. Citation analysis using formal concept analysis: A case study in software engineering. In *18th International Workshop on Database and Expert Systems Applications (DEXA)*, pages 545–550. IEEE Computer Society, Sept. 2007.
16. J. Voss, A. Hotho, and R. Jäschke. Mapping bibliographic records with bibliographic hash keys. In R. Kuhlen, editor, *Information: Droge, Ware oder Commons?*, Proceedings of the ISI. Hochschulverband Informationswissenschaft, Verlag Werner Hülsbusch, 2009.

References of the Analyzed Publications

17. R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th International Conference on Very Large Data Bases*, pages 487–499, San Francisco, 1994. Morgan Kaufmann Publishers Inc.
18. M. Barbut and B. Monjardet. *Ordre et classification: algèbre et combinatoire*. Hachette, Paris, 1970.
19. P. Becker and J. Hereth Correia. The ToscanaJ suite for implementing conceptual information systems. In B. Ganter, G. Stumme, and R. Wille, editors, *Formal Concept Analysis: Foundations and Applications*, volume 3626 of *LNCS*, pages 324–348. Springer, Berlin/Heidelberg, 2005.
20. T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 284(5):34–43, 2001.
21. G. Birkhoff. *Lattice Theory*. American Mathematical Society, Providence, 3rd edition, 1967.
22. J. P. Bordat. Calcul pratique du treillis de galois d'une correspondance. *Informatiques et Sciences Humaines*, 96:31–47, 1986.
23. R. Bělohlávek and V. Vychodil. What is a fuzzy concept lattice? In *CLA 2005, Proceedings of the 3rd International Workshop*, volume 162, pages 34–45, Olomouc, 2005. CEUR-WS.org.
24. C. Carpineto and G. Romano. *Concept Data Analysis: Theory and Applications*. John Wiley & Sons, Chichester, England, 2004.
25. M. Chein and M.-L. Mugnier. Conceptual graphs: fundamental notions. *Revue d'Intelligence Artificielle*, 6(4):365–406, 1992.
26. P. Cimiano, A. Hotho, G. Stumme, and J. Tane. Conceptual knowledge processing with formal concept analysis and ontologies. In P. Eklund, editor, *Concept Lattices*, volume 2961 of *LNCS*, pages 189–207. Springer, Berlin/Heidelberg, 2004.
27. F. Dau and J. Klinger. From formal concept analysis to contextual logic. In B. Ganter, G. Stumme, and R. Wille, editors, *Formal Concept Analysis: Foundations and Applications*, volume 3626 of *LNCS*, pages 81–100. Springer, Berlin/Heidelberg, 2005.
28. B. A. Davey and H. A. Priestley. *Introduction to lattices and order*. Cambridge University Press, Cambridge, 1990.
29. S. Ferré. Camelis: Organizing and browsing a personal photo collection with a logical information system. In *Proceedings of the Fifth International Conference on*

Concept Lattices and Their Applications, volume 331, pages 112–123, Montpellier, 2007. CEUR-WS.org.
30. R. Freese. Automated lattice drawing. In P. Eklund, editor, *Concept Lattices*, volume 2961 of *LNCS*, pages 112–127. Springer, Berlin/Heidelberg, 2004.
31. B. Ganter. Two basic algorithms in concept analysis. FB4-Preprint 831, TH Darmstadt, 1984.
32. B. Ganter and S. Kuznetsov. Pattern structures and their projections. In H. Delugach and G. Stumme, editors, *Conceptual Structures: Broadening the Base*, volume 2120 of *LNCS*, pages 129–142. Springer, Berlin/Heidelberg, 2001.
33. B. Ganter, G. Stumme, and R. Wille, editors. *Formal Concept Analysis: Foundations and Applications*, volume 3626 of *LNCS*. Springer, Berlin/Heidelberg, 2005.
34. B. Ganter and R. Wille. *Formal Concept Analysis: Mathematical Foundations*. Springer, Berlin/Heidelberg, 1999.
35. R. Godin and P. Valtchev. Formal concept analysis-based class hierarchy design in object-oriented software development. In B. Ganter, G. Stumme, and R. Wille, editors, *Formal Concept Analysis: Foundations and Applications*, volume 3626 of *LNCS*, pages 304–323. Springer, Berlin/Heidelberg, 2005.
36. J.-L. Guigues and V. Duquenne. Familles minimales d'implications informatives résultant d'un tableau de données binaires. *Mathématiques et Sciences Humaines*, 95:5–18, 1986.
37. N. Jay, F. Kohler, and A. Napoli. Analysis of social communities with iceberg and stability-based concept lattices. In R. Medina and S. Obiedkov, editors, *Formal Concept Analysis*, volume 4933 of *LNCS*, pages 258–272. Springer, Berlin/Heidelberg, 2008.
38. P. Krajca, J. Outrata, and V. Vychodil. Parallel recursive algorithm for FCA. In *Proceedings of the Sixth International Conference on Concept Lattices and Their Applications*, volume 433, pages 71–82, Olomouc, 2008. CEUR-WS.org.
39. S. Krajči. The basic theorem on generalized concept lattice. In *CLA 2004, Proceedings of the 2nd International Workshop*, pages 25–33, Ostrava, 2004.
40. S. Kuznetsov. Machine learning and formal concept analysis. In P. Eklund, editor, *Concept Lattices*, volume 2961 of *LNCS*, pages 287–312. Springer, Berlin/Heidelberg, 2004.
41. S. Kuznetsov and S. Obiedkov. Counting pseudo-intents and #P-completeness. In R. Missaoui and J. Schmidt, editors, *Formal Concept Analysis*, volume 3874 of *LNCS*, pages 306–308. Springer, Berlin/Heidelberg, 2006.
42. S. Kuznetsov, S. Obiedkov, and C. Roth. Reducing the representation complexity of lattice-based taxonomies. In U. Priss, S. Polovina, and R. Hill, editors, *Conceptual Structures: Knowledge Architectures for Smart Applications*, volume 4604 of *LNCS*, pages 241–254. Springer, Berlin/Heidelberg, 2007.
43. S. O. Kuznetsov and S. A. Obiedkov. Comparing performance of algorithms for generating concept lattices. *Journal of Experimental & Theoretical Artificial Intelligence*, 14(2-3):189–216, 2002.
44. F. Lehmann and R. Wille. A triadic approach to formal concept analysis. In G. Ellis, R. Levinson, W. Rich, and J. Sowa, editors, *Conceptual Structures: Applications, Implementation and Theory*, volume 954 of *LNCS*, pages 32–43. Springer, Berlin/Heidelberg, 1995.
45. D. Maier. *The Theory of Relational Databases*. Computer Science Press, Rockville, 1983.
46. M.-L. Mugnier and M. Chein. Représenter des connaissances et raisonner avec des graphes. *Revue d'Intelligence Artificielle*, 10(1):7–56, 1996.

47. C. S. Peirce. *Collected Papers*. Harvard University Press, Cambridge, 1931–1935.
48. C. S. Peirce. *Reasoning and the Logic of Things*. Harvard University Press, Cambridge, 1992.
49. S. Prediger. *Kontextuelle Urteilslogik mit Begriffsgraphen: ein Beitrag zur Restrukturierung der mathematischen Logik*. Shaker, Aachen, 1998.
50. S. Prediger. Simple concept graphs: A logic approach. In M.-L. Mugnier and M. Chein, editors, *Conceptual Structures: Theory, Tools and Applications*, volume 1453 of *LNCS*, pages 225–239. Springer, Berlin/Heidelberg, 1998.
51. D. Roberts. *The Existential Graphs of Charles S. Peirce*. Mouton, The Hague, 1973.
52. C. Roth, S. Obiedkov, and D. Kourie. Towards concise representation for taxonomies of epistemic communities. In S. Ben Yahia, E. Mephu Nguifo, and R. Bělohlávek, editors, *Concept Lattices and Their Applications*, volume 4923 of *LNCS*, pages 240–255. Springer, Berlin/Heidelberg, 2008.
53. E. Salvat and M.-L. Mugnier. Sound and complete forward and backward chainings of graph rules. In P. Eklund, G. Ellis, and G. Mann, editors, *Conceptual Structures: Knowledge Representation as Interlingua*, volume 1115 of *LNCS*, pages 248–262. Springer, Berlin/Heidelberg, 1996.
54. J. F. Sowa. *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley, Reading, 1984.
55. J. F. Sowa. Conceptual graphs summary. In P. Eklund, T. Nagle, J. Nagle, and L. Gerholz, editors, *Conceptual structures: current research and practice*, pages 3–51. Ellis Horwood, 1992.
56. J. F. Sowa. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks/Cole Publishing, Pacific Grove, 2000.
57. T. Tilley, R. Cole, P. Becker, and P. Eklund. A survey of formal concept analysis support for software engineering activities. In B. Ganter, G. Stumme, and R. Wille, editors, *Formal Concept Analysis: Foundations and Applications*, volume 3626 of *LNCS*, pages 250–271. Springer, Berlin/Heidelberg, 2005.
58. P. Valtchev, R. Missaoui, and R. Godin. Formal concept analysis for knowledge discovery and data mining: The new challenges. In P. Eklund, editor, *Concept Lattices*, volume 2961 of *LNCS*, pages 352–371. Springer, Berlin/Heidelberg, 2004.
59. D. van der Merwe, S. Obiedkov, and D. Kourie. Addintent: A new incremental algorithm for constructing concept lattices. In P. Eklund, editor, *Concept Lattices*, volume 2961 of *LNCS*, pages 205–206. Springer, Berlin/Heidelberg, 2004.
60. R. Wille. Restructuring lattice theory: an approach based on hierarchies of concepts. In I. Rival, editor, *Ordered Sets*, pages 445–470, Dordrecht/Boston, 1982. Reidel.
61. R. Wille. Conceptual graphs and formal concept analysis. In D. Lukose, H. Delugach, M. Keeler, L. Searle, and J. Sowa, editors, *Conceptual Structures: Fulfilling Peirce's Dream*, volume 1257 of *LNCS*, pages 290–303. Springer, Berlin/Heidelberg, 1997.
62. R. Wille. Boolean concept logic. In B. Ganter and G. Mineau, editors, *Conceptual Structures: Logical, Linguistic, and Computational Issues*, volume 1867 of *LNCS*, pages 317–331. Springer, Berlin/Heidelberg, 2000.