Graph Complexity in visual recommender systems for scientific literature

Stephan Abels[3], Christoph Greven[1], André Calero Valdez[2], Ulrik Schroeder[1], Martina Ziefle[2]

Learning Group, RWTH Aachen University[1]
Human-Computer Interaction Center, RWTH Aachen University[2]
RWTH Aachen University[3]

Abstract

Digital libraries are becoming larger, while suffering from inefficient interfaces and search patterns. Recommender Systems are a sensible and important service for users of digital libraries. The aim of recommender systems is to reduce cognitive effort, simplify search and to embed results in a larger context. In this article we compare to recommender systems – the Action Science Explorer and Papercube. Both systems are used to recommend scientific literature and use graph-based approaches. From user studies we derive the need for research to understand complexity of graphs.

1 Big Science, Big Data and the Flood of Information

Visual recommender systems help to evaluate, filter and structure large amounts of digital information. They are helpful for finding relevant objects from a larger set of objects.

One area of use of recommender systems is in digital libraries. According to the National Science Foundation, the rate of publishing in scientific literature increased about 2.3% from 1995 to 2005, collaborations across institutional border from 40% to 61% (Bergrström & Atkinson 2009). Therefore it is becoming increasingly important to process information in a sensible fashion and present it in interfaces that are easy to understand and that have good usability. The aim is to suggest literature from related field and to structure literature in a meaningful context.

Veröffentlicht durch die Gesellschaft für Informatik e.V. 2016 in
B. Weyers, A. Dittmar (Hrsg.):
Mensch und Computer 2016 – Workshopbeiträge, 4. - 7. September 2016, Aachen.

http://dx.doi.org/10.18420/muc2016-ws11-0005

1.1 Scientific Recommender Systems

It is important to classify literature in a meaningful context. Finding literature is easy but it is hard to process the enormous amount of information. You do not only require recommender systems, but visual recommender systems. You need interfaces with good usability to browse metadata like keywords, authors, citation networks, collaborations or related topics. Without these interfaces, the user has to cope with a huge cognitive load to find their way in the ever growing digital libraries. Hereinafter two approaches to visualize scientific literature are introduced.

2 Visualization of scientific literature

Although there are several different approaches to visualize scientific literature, we limit our elaboration to the projects PaperCube and Action Science Explorer (ASE) because they use different methods. Besides those tools there are also CiteScape (Liu et al. 2015), MyMediaLite (Gantner et al. 2011) or Sci2 Tool (Solomon 2015). Their focus is less the visual visual representation, but more on the technical aspects of creating and weighting recommendations and are therefore not considered in the following.

2.2 PaperCube

The goal of PaperCube – a web application – is the navigation of bibliographic metadata without losing the paper out of focus. By means of the citation network the paper is integrated in a bigger context. Through spatial visualization of the metadata the user can get new insights into related literature. The main goals of PaperCube can be summarized as *focus & context* (Bergström & Atkinson 2010*).* PaperCube offers two basic visualizations, visualization of literature and of authors. Figure 1 shows the basic interface[1].

[1]Für weitere Informationen und Darstellungen http://papercube.peterbergstrom.com

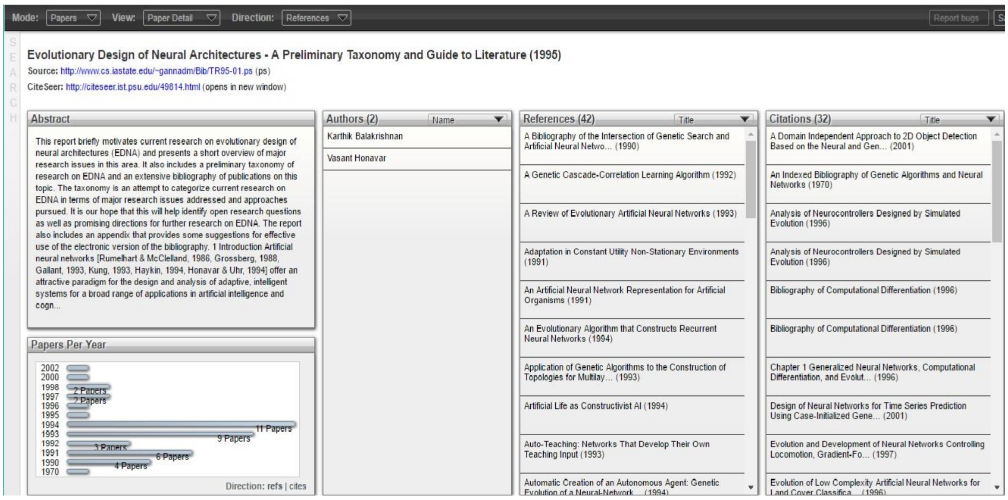

Figure 1: Details of an article after a search request

2.2.1 Visualization of literature

It is possible to visualize and explore the respective citation network in PaperCube in three different ways. By use of the direct citation network, it is possible to find directly referenced literature which has a strong connection to the specific paper. To find weaker connections, PaperCube uses a hierarchic tree structure. This structure allows to advance further into the citation network to find new and related literature which was the basis for the respective work. The third possible approach is the chronological visualization of the citation network (Bergström & Atkinson 2010).

2.2.2 Visualization of authors

The goal is to reveal collaborations and connections between authors. It allows finding articles and papers of co-authors and likewise shows the references among the authors. (Bergström & Atkinson 2010).

2.2.3 Different kinds of visualization

PaperCube offers different kinds of visualization for literature and authors. First of all, the detail view as shown in Figure 1 which is the standard view after the users searches. Futhermore PaperCube offers circle view, tree view (Figure 2 left), articles per year and PaperGraph, an alternative view to the tree structure. Besides the standard detail view of the authors, PaperCube offers several alternatives like for instance the possibility to show collaboration network of the authors via weighted graphs (Figure 2 right). Every view has the option to change focus to another author or publication.

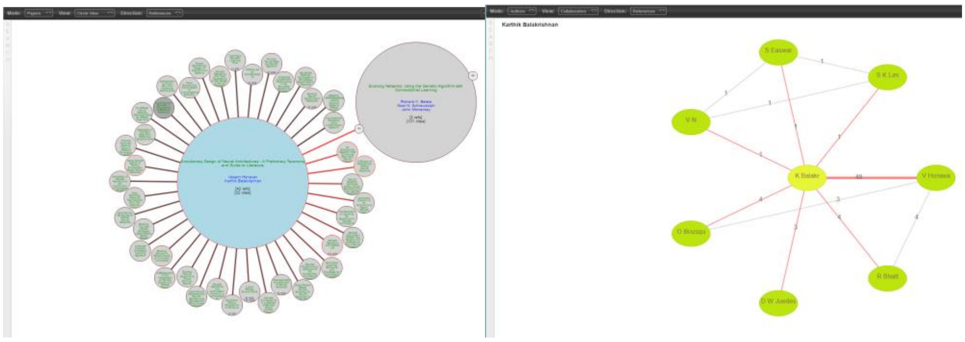

Figure 1: left tree View of a search request; right collaborations network of an author in PaperCube

2.2.4 User Study

A user study with students, researchers and professors was conducted to examine which visualization is the preferred one (Table 1) and to get general feedback on the features and interface. 85% stated that PaperCube is a useful extension to the usual web interfaces and 100% would use PaperCube in future search queries. 87% preferred the literature view over the author view, because it is more intuitive. The users prefer slim design with less information shown on display. It is important for the participants to discover new connections as fast and easy as possible and they do not want to be overloaded with information (Bergström & Atkinson 2009).

Rank	Paper View	Pct.	Author View	Pct.
1	Circle View	32%	Details	40%
2	Details	29%	Collaborators	37%
3	Papers per Year	24%	Author Cities	17%
4	Tree View	9%	Papers	7%
5	Paper Graph	6%	-	-

Table 1: Preferred views in PaperCube (vgl. Bergström & Atkinson 2010)

The participants reported that PaperCube lowers cognitive load and simplifies the discovery of new or related literature. The easy approach was commended as well. Exemplary remarks were for instance: *“reference searches [to be] much faster and complete.“[PaperCube is a] ”fantastic way to navigate and explore to find serendipitous connected papers / studies, and would make a particular library’s catalog that much more valuable.“* (Bergström & Atkinson 2010).

2.3 Action Science Explorer

Another project to visualize scientific literature is the Action Science Explorer(ASE). ASE also wants to present a quick overview about a new topic and tries to satisfy different target audiences, like students or experts (Dunne et al. 2012). ASE also uses graph based visualization.

2.3.1 Features von ASE

The interface of ASE (Figure 3) offers lots of features and function but is not as easily understood as PaperCube. One of its disadvantages is the necessity of a large screen and high resolution because of the amount of different windows and the information overload. Furthermore, several windows on the main screen can change with a single click which means a high cognitive load for the user (Gove et al. 2011).

The following features are implemented in ASE:

- Paper Ranking: Ranking and filtering of the literature by the citation network.
- Search: Searching any bibliographic data like title, author, keywords or abstract.
- Sorting: Sorting of the bibliographic entries.
- In-Cite Text: Shows sentences of related literature which quote the respective paper, including a link to the quoting publication.
- In-Cite Summary: Automatically generated summary of all sentences that cite the respective paper.
- Out-Cite Text: Full text of the source paper with highlights and links to the referenced literature.

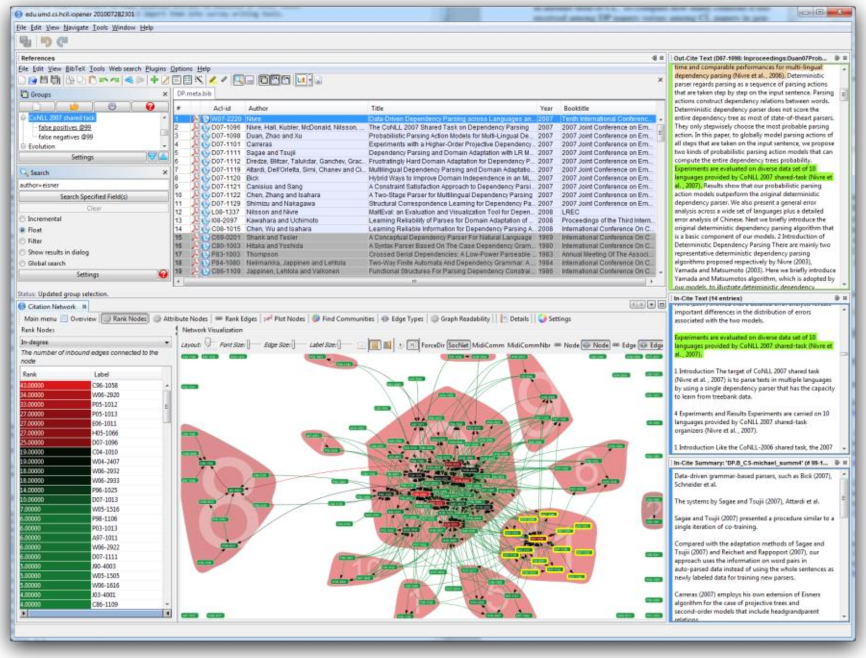

Figure 3: Interface of ASE

2.3.2 User Study

Three user studies (n=4-5) were conducted to improve usability and to research how effective user used ASE to discover new and related literature (Dunne et al. 2012). The study shows that users could not use all features after a 30-minute introduction. The most used feature was the Paper Ranking. It allows the user to filter less important publications at a glance and to focus his attention to relevant literature.

Three out of four participants rated ASE as useful for searching literature. One participant stated problems concerning the interface because it is too overloaded. Furthermore, the participants had problems using and understanding all feature ASE offers, like the Out-Cite text for example. Another hint for the problematic usability is the fact, that a 30-minute introduction was not enough to use and understand all features (Dunne et al. 2012).

3 Discussion

PaperCube and ASE differ in several important functions. PaperCube offers a simple design and good usability. Its main goal is to lower cognitive load and put literature in a bigger context without losing focus. Another advantage is the choice between several visualizations (detail view, tree view, papers per year) to satisfy different needs the user might have.

In contrast, ASE offers lots of useful features, but they are not understood straight away. It adds a lot of value with the linking of referenced publications and publications which reference the respective paper. Being able to see referenced publication in its context is useful in understanding the quotation. Maybe ASE is more useful for a different (more professional) target audience that PaperCube. Moreover, the Out-Cite feature in ASE might be useful in the study of plagiarism.

A hybrid approach from both projects might be very effective. The most popular function from ASE, the Paper Ranking, could be integrated in PaperCube. This way the user does not have to manually browse the citation network which is the least popular feature in PaperCube. Another advantage from ASE are the quotations and references which are directly displayed in the program and in its contexts. This saves time and tasks to embed literature in its bigger, meaningful context.

4 Research Agenda

The presented approaches help the user to visualize literature in its context. Both approaches take workload of the user and help to decrease their cognitive load, although they take different paths. Both utilize graph based visualizations with various results. The nature of the data leads to the necessity to process complex network structures in a visual way. Improved usability leads to the reduction of the displayed amount of data. The question arises how visual complexity in publication graphs should be captured (vgl. Calero Valdez et al. 2015) and how to handle the trade-off between the wealth of information and reduction of information. Graph complexity could correlate in a U-shaped way with their usefulness.

Hence, one aim of human factors research should be to develop perception metrics for graphs and correlate those with the resulting cognitive load. It is obvious to assume that domain knowledge (about specialist field and topic), spatial sense, expertise and more individual characteristics have an impact on users' system perception. Direct metrics for graphs can possibly be used to value the perceived complexity of these (e.g. graph entropy, topology, see Holzinger et al. 2013). The examination of which metrics can be applied is topic of future research.

Acknowledgements

We would like to thank the anonymous reviewers for their constructive comments on an earlier version of this manuscript. The authors thank the German Research Council DFG for the friendly support of the research in the excellence cluster „Integrative Production Technology in High Wage Countries“.

References

Bergström, P. & Atkinson, D. C. (2009). *Augmenting the exploration of digital libraries with web-based visualizations*. pages 63–69.

Bergström, P. & Atkinson, D. C. (2010) *Augmenting digital libraries with web-based visualizations. Journal of Digital Information Management*, 8(6):377–386.

Calero Valdez, A., Brauner, P., Schaar, A. K., Holzinger, A., & Ziefle, M. (2015). Reducing Complexity with simplicity-Usability Methods for Industry 4.0. In *Proceedings 19th Triennial Congress of the IEA* (Vol. 9, p. 14).

Dunne, C., Shneiderman, B., Gove, R., Klavans, J. & Dorr, B. (2012). *Rapid understanding of scientific paper collections: Integrating statistics, text analytics, and visualization. Journal of the American Society for Information Science and Technology*, 63(12):2351–2369.

Gantner, Z., Rendle, S., Freudenthaler, C.& Schmidt-Thieme, L. (2011). *Mymedialite. A free recommender system library. In Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys '11*, pages 305–308, New York, NY, USA.

Gove, R., Dunne, C., Shneiderman, B., Klavans, J. & Dorr, B. (2011). *Evaluating visual and statistical exploration of scientific literature networks*. In VL/HCC, volume 11, pages 217–224.

Holzinger, A., Ofner, B., Stocker, C., Calero Valdez, A., Schaar, A. K., Ziefle, M., & Dehmer, M. (2013). On graph entropy measures for knowledge discovery from publication network data. In *Availability, reliability, and security in information systems and HCI* (pp. 354-362). Springer Berlin Heidelberg.

Liu, Z., Yin, Y., Liu, W. & Dunford, M. (2015). *Visualizing the intellectual structure and evolution of innovation systems research: a bibliometric analysis.* Scientometrics, 103(1):135–158.

Solomon, D. (2015). *A different view: An inquiry into visualization of bibliometric data. volume 122nd ASEE Annual Conference and Exposition: Making Value for Society.*

Tejeda-Lorente, A., Porcel, C., Peis, E., Sanz, R. & Herrera-Viedma, E. (2015). *A quality based recommender system to disseminate information in a university digital library. Information Sciences*, 261:52–69.

Autoren

Abels, Stephan

Stephan Abels is a student at RWTH Aachen University, studying both Communication Science and Computer Science in the B.S.c. Technical Communication since 2015.

Greven, Christoph

Christoph Greven completed his master degree in computer science and works as a researcher at the Learning Technologies Research Group of the RWTH Aachen University. His scientific focus is knowledge management and mobile learning as well as related aspects with the potential for further improvement of these, including Recommender Systems. Currently, he is mainly engaged in the award-winning joint research project Professional Reflective Mobile Personal Learning Environments (PRiME).

Calero Valdez, André

André Calero Valdez has studied computer science at the RWTH Aachen University and holds a PhD in Psychology also from RWTH Aachen University. He a senior researcher at the Human-Computer Interaction Center of the RWTH Aachen University and visiting professor with the HCI-KDD group in Graz, Austria. His thesis dealt with the topic of user-centered design of small screen devices for diabetes patients. He currently conducts research in the topics of knowledge management, social media, and decision support by visualizations. The aim is to manage complexity of information by applying human-computer interaction principles.

Ziefle, Martina

Martina Ziefle holds the chair of communication science and is founding member of the Human-Computer Interaction Center of the RWTH Aachen University. Her research addresses the communication between human-human and human-machine with the research focus on technology acceptance for various technologies with respect to user diversity.

Schroeder, Ulrik

Ulrik Schroeder is Professor in Computer Science at RWTH Aachen University. He heads the research group Learning Technologies, and the Center for innovative Learning Technology of RWTH Aachen University (CiL: http://www.cil.rwth-aachen.de/), which is responsible for the universities eLearning infrastructure. Furthermore, he is director of the extracurricular learning laboratory for computer science for K-12 school students (Infosphere). He is responsible for the Computer Science Teacher Education Programm at RWTH Aachen. He is vice chairman of the technical committee for Informatics and Education / Didactics of Informatics as well as the Special Interest Group E-Learning within Gesellschaft für Informatik. His research interests are assessment and intelligent feedback with focus on learning processes, learning analytics, social software in education, mobile Internet and learning, gender mainstreaming in education, as well as computer science didactics and teachers' education.