Visualisation of the Dynamics of Computer-mediated Community Networks

Andreas Harrer, Sam Zeini, Sabrina Ziebarth and Daniel Münter
Department of Computer Science and Applied Cognitive Science
Universität Duisburg-Essen, Germany
{harrer, zeini, ziebarth, muenter}@collide.info

In this paper we will demonstrate the potential of processing and visualising the dynamics of computer-mediated communities by means of Social Network Analysis. According to the fact that computer-mediated community systems are manifested also as structured data, we use data structures like e-mail, discussion boards, and bibliography sources for an automatic transformation into social network data formats. Currently our developed converter DMD (Data Multiplexer Demultiplexer) supports GraphML, UCINET, and Pajek formats besides our own data formats which are used for real-time analysis of CSCL (Computer Supported Collaborative Learning) activities. In the case of communication data our converters utilize conversation graphs reflecting aspects of speech act and conversational theory to produce directed graphs in the cases where one-mode person networks are desired.

The paper will demonstrate a 3-dimensional visualisation of an author community based on Bibtex bibliography data converted into GraphML. Based on this dataset we visualise publications network with a tool called Weaver, which is developed in our research group. According to Lothar Krempel's algorithm, Weaver uses the first two dimensions to embed the network structure within a common solution space. The third dimension is used for representing the time axis and thus the dynamics of co-authorship relations.

Concluding we aim to discuss potential issues and problems of our approach and the possibilities especially concerning the appropriate visualisation and segmentation of long term communications, such as mailing lists.

Introduction

Social Network Analysis (SNA; for an overview please see: Wasserman & Faust 1994) is becoming more and more interesting to the domains of computer science related to communities. Besides the body of work of computer scientists in the graph drawing community[1] (where we were inspired by the work of Ulrik Brandes and Dorothea Wagner), we believe that SNA will once be established on par with statistical analysis as well as with qualitative approaches in the interdisciplinary fields of "Computer Supported Cooperative Work" (CSCW) and "Computer Supported Collaborative Learning" (CSCL), where empirical grounding is broadly common to the community. As an example of current development for CSCW one can mention the workshop on SNA at the international CSCW conference in the year 2004[2]. In the field of CSCL research using SNA was presented to a broad audience at the international conference on the learning sciences by Palonen and Hakkarainen (2000) and the international CSCL conference by Reffay and Chanier (2003) and is now also published in international journals within the community (e. g. Martinéz-Mones et al. 2006 or Harrer et al. 2006). We identify two major reasons for attractiveness of the SNA approach to these two research domains, which are related to each other. The first reason is that wherever groups or communities are mediated through computer systems, a representation of the community is also coexistent in the computer systems as structured data. This "natural data" often can be transformed easily into social network data, once we assume the networks being defined as closed networks by the fact of user management implicated by the systems. Such systems can be well known basic technical support approaches like mailing lists or discussion forums. They also can be advanced Cooperation tools like the BSCW system[3] or advanced collaborative learning environments like shared workspace systems in combination with

[1] www.graphdrawing.org
[2] http://projects.ischool.washington.edu/mcdonald/cscw04/
[3] http://www.bscw.de/english/index.html

learning object repositories (e. g. Hoppe et al. 2005). The basic concepts on which we build the data transformations will be a part of this paper.
The second reason which is related to the operational availability of data is related to the possibilities of visualisation provided by concepts and techniques of Social Network Analysis. According to Krempel (2005) advanced visualisation techniques should be able to impart complex knowledge in a very efficient way. In the area of CSCL the potential of awareness through visualisation of collaborative and social aspects is discussed for several target groups:

- Researchers
- Teachers
- Students

Researchers can use these visualisations as well as the related SNA indices as means to support other methods of analysis (Martinéz-Mones et al. 2006; Harrer et al. 2006) since triangulation research design is common ground within the interdisciplinary CSCL field. Teachers are enabled to understand the group structures in their computer supported classes and courses (e. g. for school classes using discussion forums or blogs and university courses with blended scenarios) and potentially can use this information to guide and advice the students, e. g. when participation of specific students is extraordinarily high or low. Finally, students could use the visual feedback for self reflection and self regulation in reaction to the information transported (cp. Sun &Vassileva 2006). Other potential scenarios related to work in the CSCL are situations in further education, where social networks are critical to the aspects of development of competencies on the job. Here such awareness support systems can provide orientation within networks, which are often complex.
A third aspect which emerges from both reasons described above is that the data captured from the collaborative systems contains additional information. In our cases the timestamps captured by these systems are used to include the dynamics of the networks in the representation. For these reasons we developed a toolset of transformation concepts and tools as well as techniques for visualisations to support the mentioned user groups and that are tailorable for their specific needs.

Approach for Data Collection and Processing

The data used in our network analyses is – in the classification of Wassermann & Faust (1994) – mainly of the category "archival records", i.e. compiled in formal and structured documents that can be parsed with the help of computer programs. Our initial research interest (Harrer 2004) was the support of asynchronous collaboration in thread-based discussion forums. The data that is created there provides well-defined connections between the contributions written by the network actors. Formally, these networks of postings are connected with each other in a "refers-to" relation and each posting is created by a network actor. These networks can be represented graph-theoretically as two-mode networks. Depending on the complexity of the discussion forum, the posting entities in the network might have several additional attributes interesting for analyses, such as creation time, categorical information (e.g. speech act categories (Austin 1962) that have been used by us in the SPREKON system (Harrer 2004), intended readers, detailed text etc.
Building up on first analyses of these communication archives, we transferred our data formats and tools to other archival collaboration data, such as third-party web portals, mailing lists, and bibliographical data.
All these data sources are mapped to graph-theoretical data structures, that provide a uniform XML-binding, and that can be used for a variety of data conversions and analysis techniques. Among these are typical operations, such as reductions from two-mode to one-mode networks (Wassermann & Faust 1994), conversions to third party formats such as UCINET, Pajek, or GraphML, and our own analysis and visualisation techniques, as the one we describe in this

paper. An overview of the formal aspects of conversions between different network types can be seen in Figure 1.

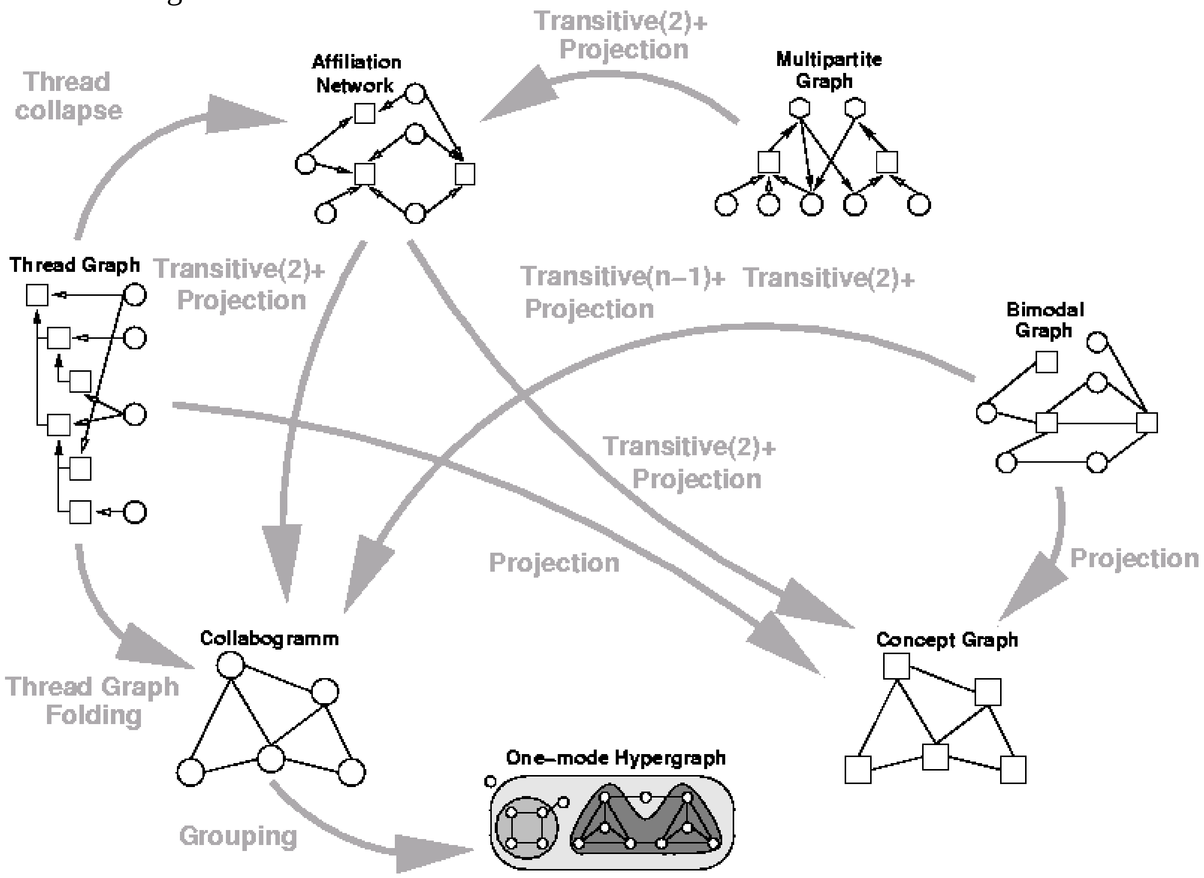

Figure 1: Schema for formal network transformations

For the example case in this paper we will focus on bibliographic networks, which consist of authors, publications, and potentially the conferences / journals / book series it is published in. This is a typical case of a multi-partite graph, that can be processed further to a bipartite author-publication network (comparable to an affiliation network) and finally to an author network (comparable to a sociogram or – as we call it in CSCL contexts – a collabogram). Other typical graph-theoretic transformations we support are based on thread graphs that occur naturally in discussion forums (see above) and that can be transformed to either strict person-topic networks (bi-partite type) or one-mode networks (either focussing on persons or on topics, similarly to the duality of affiliation networks). Grouping and partitioning of actors into subsets with similar properties or shared relations (this can be represented by a hyperedge or enriching the actor attributes) will also be discussed in this paper and the proposed visualisation approach to improve the user in interpreting the network structures. More technical aspects of the supported data formats will be described in the Implementation section of this paper.

Approach for Network Visualisation

In addition to powerful methods for analysing social networks, it is often very useful to have a visualisation of the results which is as simple as possible. A very common visualisation of a social network is a sociogram, which contains actors as nodes layouted on a circle and their relationships as edges connecting the nodes. While this kind of visualisation seems to be self-explanatory it might also lead to misinterpretation. Although the nodes are randomly assigned to the circle and the lengths of the connecting lines have no meaning, users tend to interpret visual adjacencies as social adjacencies. Thus, the layout of the nodes is a central problem of network visualisation. Krempel (2005) suggests an algorithm for arranging the nodes in simple solution spaces. A solution space defines a set of feasible positions for the nodes, e.g. coordinates on a circle. To optimise the layout of the network graph, edges should cross as few as possible and thus highly connected nodes should be placed close to the nodes they are connected with. Unfortunately, the calculation of the optimal positions of the nodes would require to test all of their possible positions, which has high computational efforts. To avoid testing all permutations, Krempel (ibid.) arranges the nodes with the highest structural importance first and accepts that nodes which are structurally less important are arranged less optimal. Structural importance is defined by the centrality measures of the nodes, thus central nodes, which have either many direct or many short indirect links to other nodes, are placed with priority. An advantage of the use of simple solutions spaces is that nodes can be placed according to their properties in predefined regions of the visualisation. For example, each set of nodes belonging to one time interval could be arranged on one circle. Our Weaver application uses Krempel's algorithm (ibid.) for arranging the nodes according to their properties and the perspective the user is interested in, e.g. there are special views for group visualisation or multi-mode networks. Furthermore, Weaver allows the definition of the appearance of the nodes according to their properties in the form of rules, which can be saved and exchanged as rule sets related to specific type of networks. For example, in a two-mode network actors could be represented as filled circles and events as squares, and the nodes could have a size proportional to their importance and colour by defined other properties, such as formal rules or hierarchy in a group. From the user view this means, that he or she is able to perceive the properties of a node immediately and intuitively.

Using temporal dimension

The consideration of the dynamics of networks over time has been discussed early in the SNA literature, as the well known example of the Sampson monastery data (cp. de Nooy et al. 2005) shows. Yet, the capturing of discrete snapshots of a social network for each interesting moment and the isolated diagrammatic representation does not properly support the reception and interpretation of network dynamics easily. Some related work relies on animated transitions through the snapshots to integrate the temporal dimension (e. g. the commetrix system[4], cp. Trier 2005). In our approach, we make use of the third dimension for the representation of dynamics: all interactions in our archival data contain timestamp information, that represents creation date, publication data etc. Thus it is possible to augment the typically two-dimensional sociogram representations with temporal information along a third dimension.

One possibility we explored for visualisation using a temporal dimension is the radial rendering with a “starting point” (i.e. the timestamp of the first interaction) and subsequent timestamps on concentric circles surrounding the previous ones. While this creates an intuitive grasp of the degree of activity at each time slice (see Figure 2), the optimized graphical rendering of actors and their close collaborators is a difficult challenge with this approach. Even when each actor is represented by an own radius line in the diagram, the visual representation inherently would imply that relations between the same actors at later time stamps are more distant / less intense than at earlier time stamps. This might obscure or inhibit the intuitive interpretation of the diagram type.

[4] www.commetrix.de

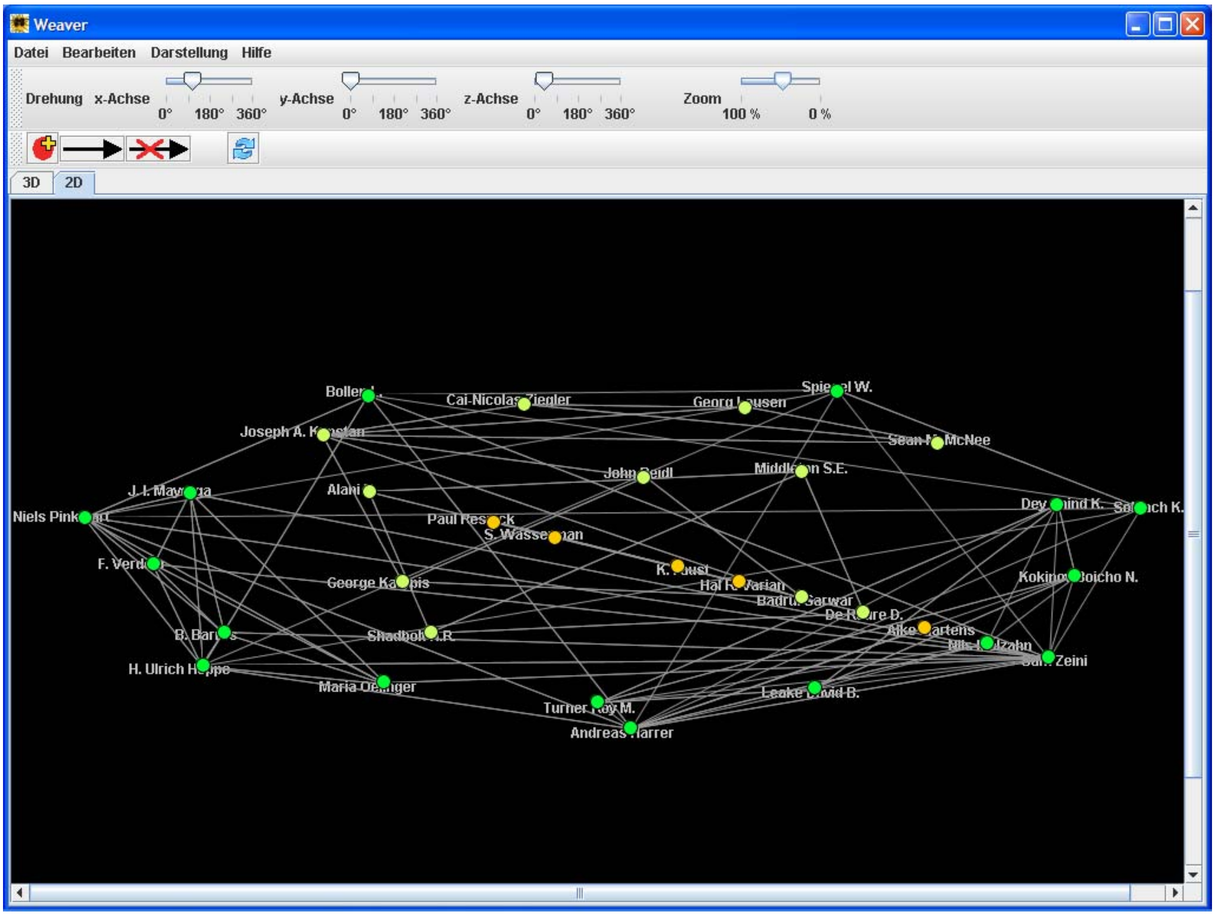

Figure 2: Radial diagram for temporal dynamics

In our proposed visualisation method, the integration of temporal dynamics is conducted as follows: Each network actor is represented with a timeline showing the timestamps her interactions took place at. These timelines are oriented for all actors along the third (Z) dimension. To optimize a graphical rendering of a "full community sociogram", all the actors within one timeslice are taken into account to define the layout in two dimensions (X-Y) according to the metrics of embedding into solution spaces (Krempel 2005). Thus, when the whole community is of interest for analysis a flat X-Y view is chosen. For a conventional snapshot representation of the network community at a given moment, the filtering mechanism of our visualisation tool can be used using the timestamp of interest as the filter. For the integrated perception of the dynamics of the network, the "flat view" is rotated along the Z-axis, revealing both the individual actors' timelines and the different slices representing network interaction at one given moment. A detailed inspection of interesting aspects is further supported by the zooming mechanism and a filtering on specific actors.

Implementation

The data processing for our visualisation approach follows along a tool chain we developed in recent years: Archival data, for our scenario in this paper formal bibliographic data in the Bibtex-format, is parsed using a configurable parser generator provided by the Data-Multiplexer-Demultiplexer (DMD) application. This flexible approach allows us to handle different input sources such as mailinglists and newsgroups, discussion forums (cp. Harrer et al. 2006) and photo community web galleries, as well as our own formats SPREKON (Harrer 2004) and CoNaVi (used by our Community Navigation Visualiser: Malzahn et al. 2005). The parsed entries are mapped to our internal SPREKON data structure representing the publications, the authors and potentially the editors. This network can be exported as a multi-mode network or can be reduced to an author network in different formats, such as Pajek, UCINET, or for this purpose a specific GraphML dialect (http://graphml.graphdrawing.org/). Here it is also possible to export networks which contain the thread graphs representing the relations between the topics. The DMD application also allows the user to choose directed or undirected graph exports. For us, one of the important aspects of the internal handling of the thread graphs is the possibility to derive directed one-mode person networks from the multimodal communication networks. DMD uses an internal object structure which consists of a set of postings based on the SPREKON format and provides us the flexibility to extend DMD for new input and output formats.

One of the main problems we faced during converting large sets of electronic communication data is the possible redundancy of users e.g. a user is known by different email accounts. To solve this problem we introduced a user merge function (see Figure 3) in DMD which provides several filtering concepts (similarity, Levenshtein, etc.) to support the researcher while cleaning the noise of the source data. Last but not least the DMD application handles also the temporal information in the data transformation process. Here the user is able to choose between the exact timestamps or a slicing into discrete time points representing an interval of days, months, or years. Currently we support timeline information in the output of Pajek and GraphML files.

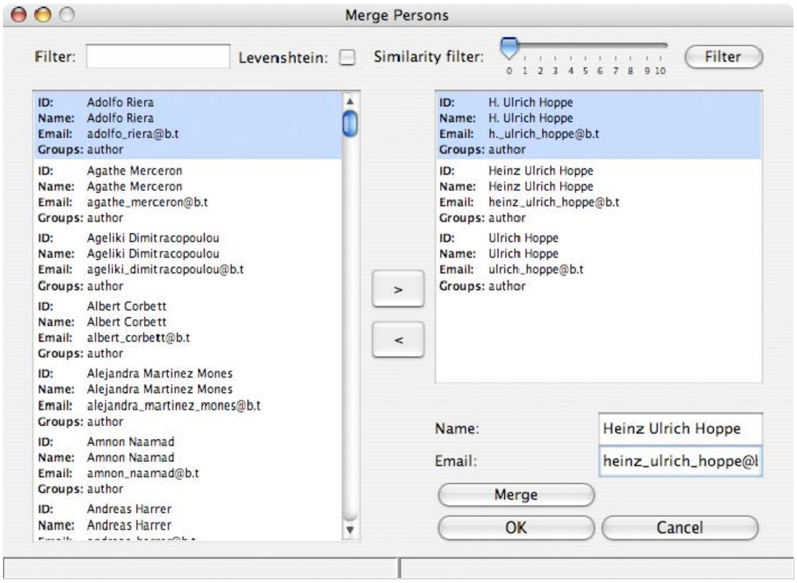

Figure 3: DMD's user merge window

That GraphML file can be imported into our Weaver application and is best viewed then with our "temporal dynamics" view and an appropriate rendering style that gives specific visual

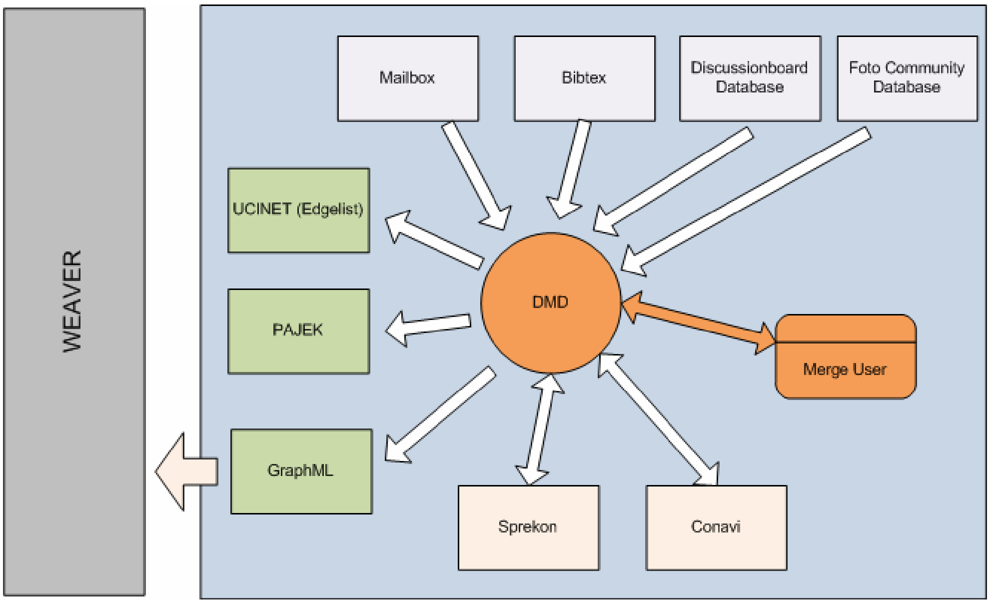

Figure 4: Overview of the toolchain

attributes to actors based on the actor properties (e.g. actor size proportional to actor's centrality). The schema for the tool chain with data sources (i.e. the raw communication data sets on top of the figure), sinks (i.e. output formats that can be produced in the DMD, but not be used for further transformations; shown on the left side of the figure) and interoperable tools that can analyse the data, enrich it and feed it back to the DMD (such as SPREKON and CoNaVi shown on the bottom that are used by our own implemented SNA tools) is sketched in Figure 4. According to (Krempel 2005) techniques exist to integrate structural properties of networks in the display. For this purpose we created the Weaver application, a 3D visualiser for social networks, which arranges and draws the nodes according to properties such as degree, centrality, or externally defined properties within a simple solution space. From the user view this means, that he or she is able to perceive the properties of a node immediately and intuitively (e.g. what is the most central topic in the network).

Example Case for the Visualisation Method

To demonstrate the features of our visualisation method and our tool chain, we analysed a specific set of bibliographic information, that was taken from a citation network of CSCL literature with special focus on computer-supported collaboration, interaction analysis, and communication networks. The raw data was represented in the Bibtex-format, as structured format containing information about authors, publication titles, type of publication, editorial information, publication date etc. Frequently bibliographic data contains multiple writings for the same persons: Sometimes the publisher's formats demand full names, sometimes only the "initial" of the first name, and some authors even vary their names across publications using or not using "middle initials" or double family names. To identify persons with different writing across the publications we used the "user merge" feature of DMD and thus cleaned the data for further analyses. While we have features to visualise and analyse multi-mode networks, we will concentrate in this example on one-mode co-authorship networks created from this data.
This data from a *data source* (cf. Figure 4 on top) was converted by the DMD tool to the GraphML format and – for comparison reasons – to the Pajek .net format. Figure 5 shows an isolated diagram produced by the Pajek application (Batagelj & Mrvar 2003) with our time

slice export. By navigating through the single slices the user has to interpret the dynamics of the network on his own.

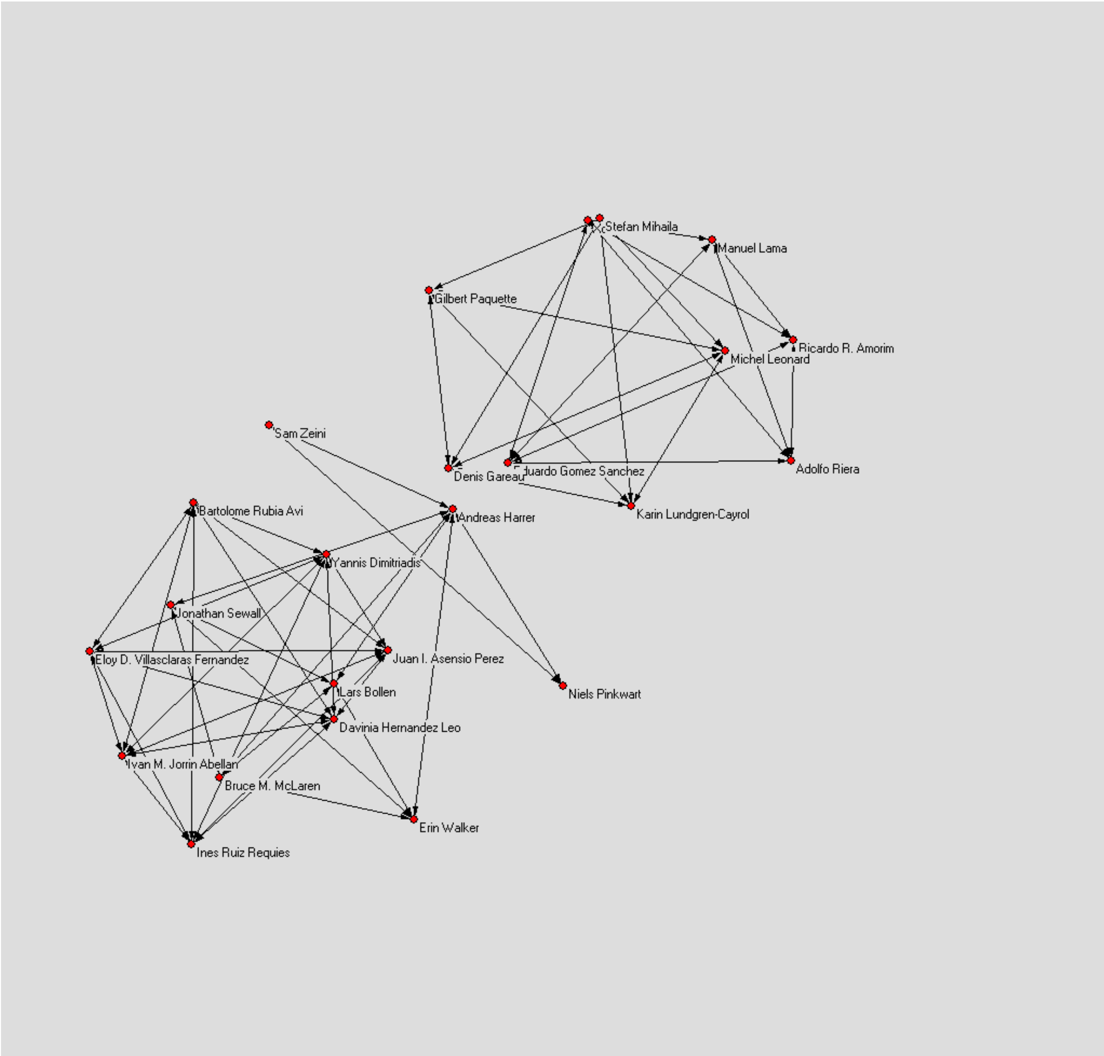

Figure 5: Isolated time slice diagram of the author network

The visualisation method we propose in this paper uses the GraphML output produced by the DMD. Using the temporal information in the third dimension, we get a representation that can be manipulated interactively by the user through changing the perspective in the 3-D coordinate system.

Figure 6 shows a flat view on the 3-D visualisation that shows the temporal axis from right to left. Each vertical set of authors represents the authoring network in a specific time interval. Actors that published at various intervals are represented using colour-highlighted “lifelines” symbolising their publication history. Actors with numerous co-authors are rendered in different colours than authors that publish with few co-authors, using the SNA properties and algorithms of k-core (cp. Batagelj & Mrvar 2003). These augmentations of the network by visual differentiation are supported flexibly in Weaver using node styles and edge styles that define rule sets for visual rendering.

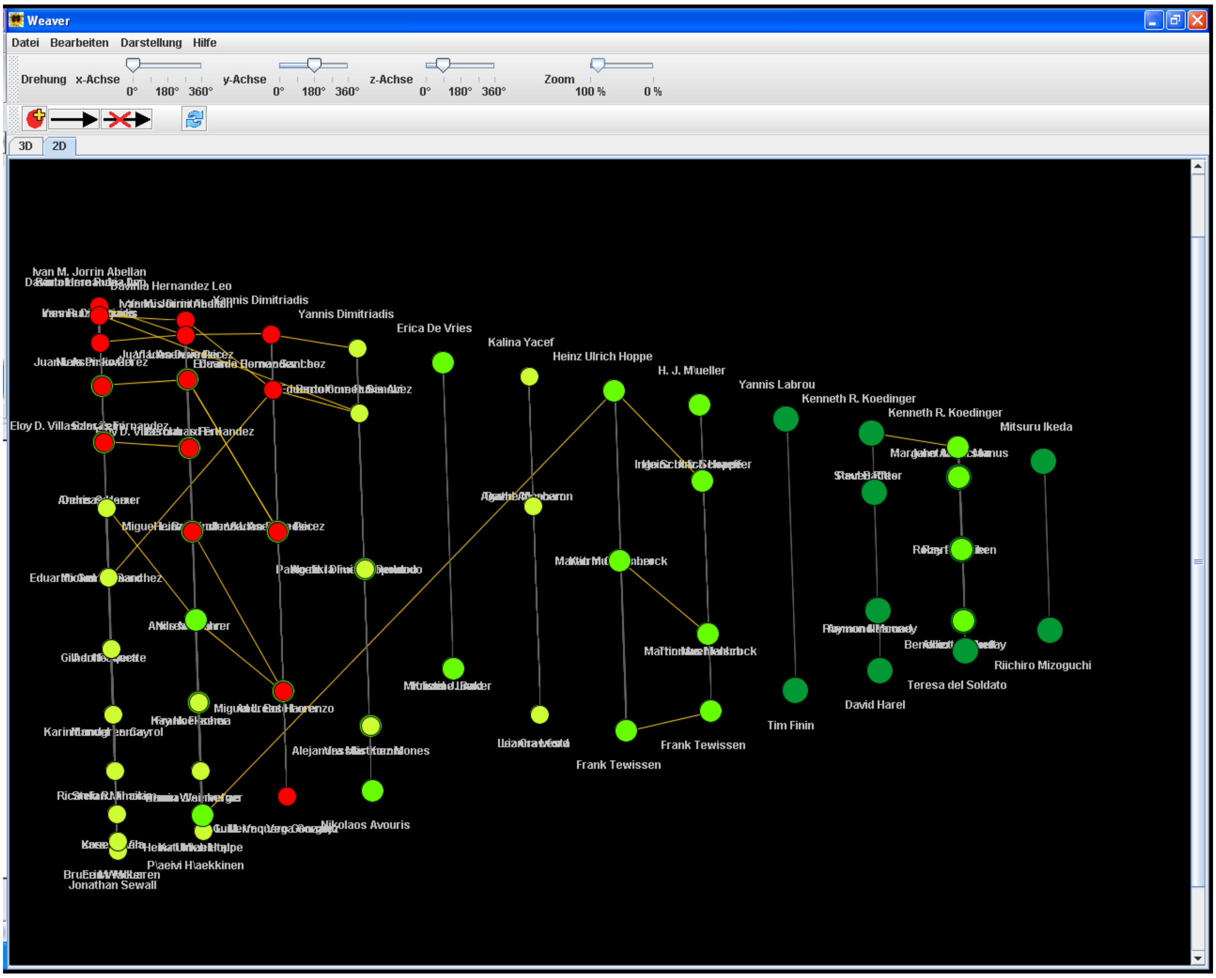

Figure 6: View on the time dimension (left to right) and part of the authoring network

The use of all three dimensions simultaneously can be seen in Figure 7 where the 3-D model is shifted slightly towards the previously hidden axis. While each time slice is still discernable, the information about the author network in the slice is additionally visible. Because the authors' history (lifeline) is also visible, this perspective is well suited to get a first-glance impression of the network including the temporal dynamics.

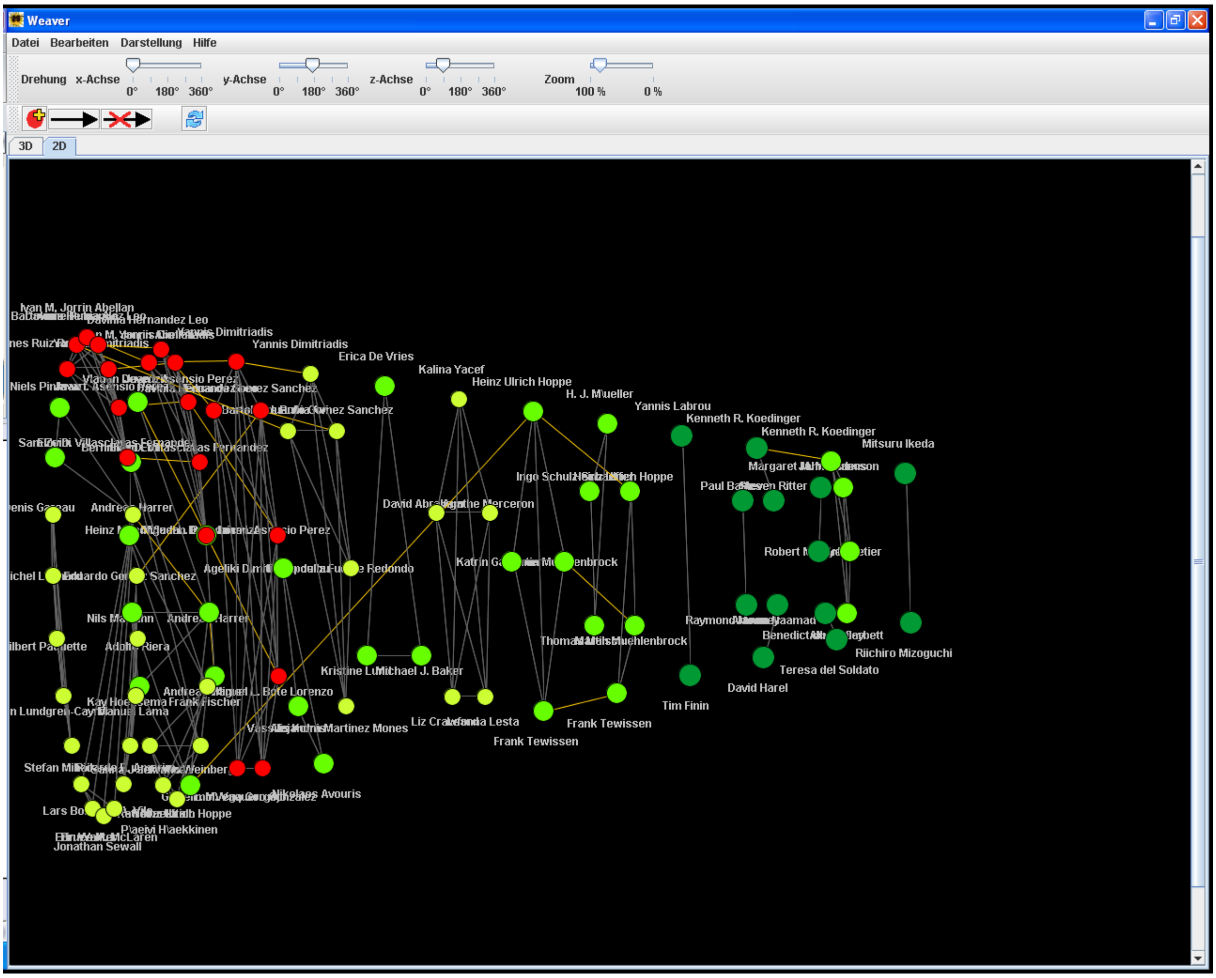

Figure 7: Using all three network dimensions simultaneously

When the user is interested especially in zooming on specific periods of time, another feature of the Weaver application, a flexible filter mechanism can be used. A filter on the year 2005 produces the timeslice represented in Figure 8, which is well comparable with the network view in Figure 5.

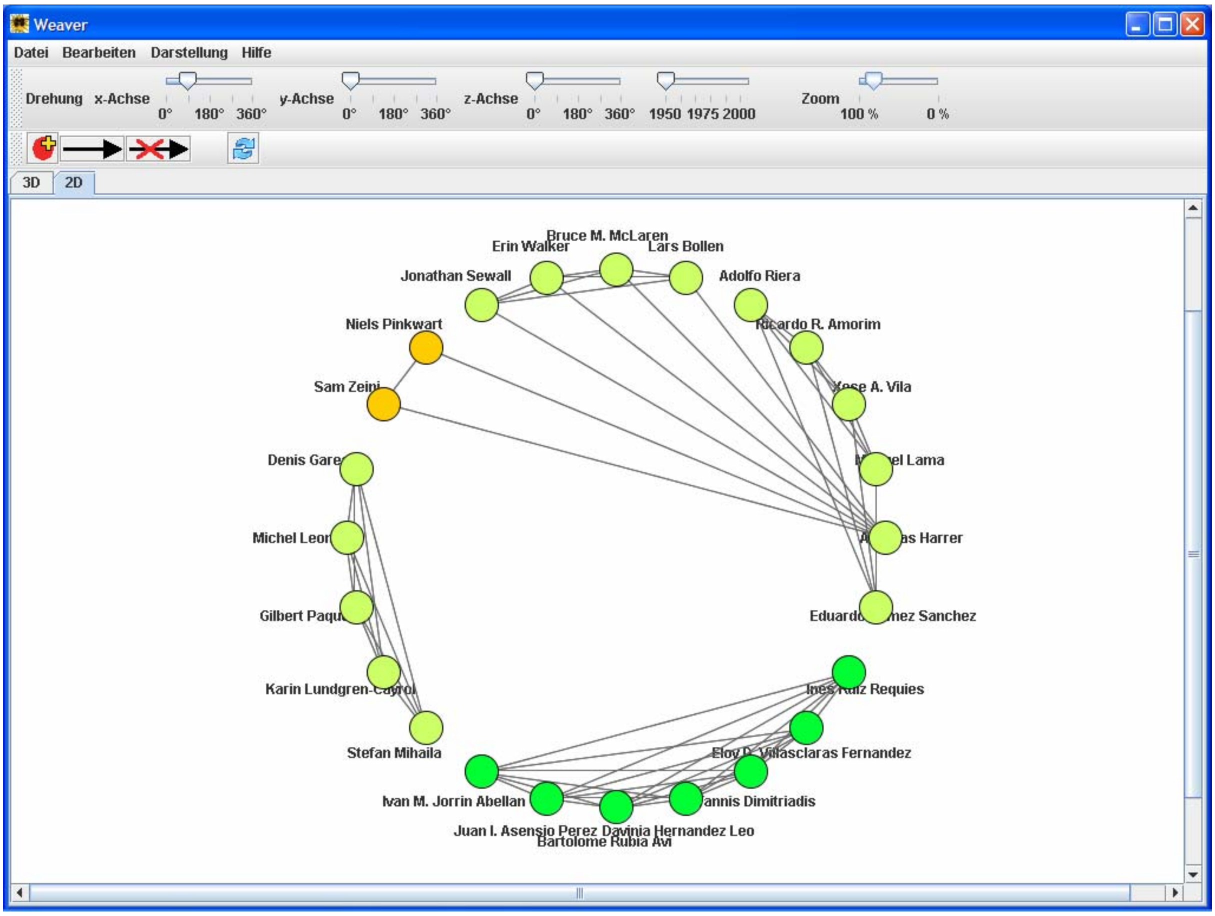

Figure 8: Projection into one timeslice using a filter

Using combinations of filters, finer selections on arbitrary subsets of the whole dataset can be visualised flexibly in the same manner as in the previous figures (e.g. concentrating on the latest 5 years of publication in the shifted 3-D view). Filters can also be used on specific persons to visualise their publication histories in an isolated view or only with their close co-authors.

Conclusion and Perspectives

In this paper we presented a visualisation method to augment sociograms representing network communities with information about the temporal aspects and dynamics of the community. For this end we created a three-dimensional representation that can be manipulated interactively to

- Give an overview about the full actor community
- Allow a temporal snapshot of the community at a given moment
- Show the activity of any single actor in a dedicated timeline
- Integrate all the aspects in one configurable diagram.

In addition to the concept of the visualisation method, we described the data processing method, and showed our approach exemplarily with the bibliography data of the Bibtex representation of selected CSCL literature.

Besides the practical testing of the visualisation method in different contexts, such as online discussion forums, project mailing lists, and scientific publication/citation networks, we also want to use our advanced configuration features for the visual rendering via styles to highlight temporal changes of SNA traits: e.g. an actor that gains centrality over time (computed in each time slice and compared between these), should be represented with a special code, such as a bright colour showing the temporal specific of a "rising star".

Acknowledgements

Main parts of this work were done in the VIP-NET Project (01HU0128) funded by the German ministry of education and science (BMBF) and in the European research team CAViCoLA of the Kaleidoscope network of excellence supported by the European Community (IST Contract No. 507838) under the Information Society Technologies priority of the 6th Framework Programme. We also thank Oliver Lemm for the implementation of the first versions of the DMD application.

References

Austin, J. L. (1962). How to Do Things with Words, Cambridge, Mass.: Harvard University Press.

Batagelj, V. & Mrvar, A. (2003). Pajek - Analysis and Visualization of Large Networks. In Jünger, M. & Mutzel, P. (Eds.), Graph Drawing Software. Springer, Berlin 2003. p. 77-103

Harrer, A. (2004). Rechnergestützte Soziale Netzwerkanalyse in virtuellen Lerngemeinschaften. In: Harrer, A. & Martens, A. (Eds.), Proceedings of the Workshop on Teaching and Learning Systems - The Role of Artificial Intelligence in Past, Present and Future, German Conference on Artificial Intelligence KI-2004, Ulm.

Harrer, A., Zeini, S., Pinkwart, N. (2006). Evaluation of communication in web-supported learning communities – an analysis with triangulation research design. In: International Journal of Web Based Communities 2006, Vol. 2, No.4 pp. 428 - 446, Inderscience.

Hoppe, H. U., Pinkwart, N., Oelinger, M., Zeini, S., Verdejo, F., Barros, B., Mayorga, J.I. (2005). Building Bridges within Learning Communities through Ontologies and "Thematic Objects". In: Proceedings of the International Conference on Computer Supported Collaborative Learning (CSCL2005), Taiwan, June 2005.

Krempel, L (2005). Visualisierung komplexer Strukturen. Grundlagen der Darstellung mehrdimensionaler Netzwerke. Frankfurt a. M.: Campus.

Malzahn, N., Zeini, S. & Harrer, A. (2005). Ontology Facilitated Community Navigation - Who Is Interesting for What I Am Interested in? In: Dey, A. K., Kokinov, B. N., Leake, D. B. & Turner, Roy M. (Eds.): Modeling and Using Context, 5th International and Interdisciplinary Conference, CONTEXT 2005, Paris, France, July 5-8, 2005, Proceedings. Lecture Notes in Computer Science, Volume 3554, Springer, Jul 2005, Pages 292 - 303

Martínez, A., Dimitriadis, Y., Gómez-Sánchez, E., Rubia-Avi, B., Jorrín-Abellán, I., Marcos, J. A. (2006). Studying participation networks in collaboration using mixed methods in three case studies. In: International Journal of Computer-Supported Collaborative Learning, vol. 1, Issue 3, march 2006. Springer.

de Nooy, W., Mrvar, A. & Batagelj, V. (2005). Exploratory Social Network Analysis with Pajek, Cambridge University Press.

Palonen, T. & Hakkarainen, K. (2000). Patterns of interaction in computer-supported learning: A social network analysis. In: Fishman, B. & O'Conner-Divelbiss, S. (Eds.), Proc. of the Fourth International Conference on the Learning Sciences. Mahwah, NJ. Lawrence Erlbaum.

Reffay, C., & Chanier, T. (2003). How Social Network Analysis can hep to measure cohesion in Collaborative Distance Learning. In Wasson, B. Ludvigsen, S & Hoppe, H. U. (Eds.): Proceedings of the International Conference on Computer Support for Collaborative Learning. Dordrecht: Kluwer Academic Publishers, pp. 343-352.

Sun, L & Vassileva, J. (2006). Social Visualization Encouraging Participation in Online Communities. CRIWG 2006: 349-363

Trier, M. (2005). IT-supported Visualization of Knowledge Community Structures. In: Proceedings of 38th IEEE Hawaii International Conference of Systems Sciences HICCS38, Hawaii, USA, Jan 2005.

Wassermann, S., & Faust, K. (1994). Social Network Analysis: Methods and Application, Cambridge: University Press.