Visualizing Dynamics in Virtual Information Spaces

Claudia Mueller
(University of Stuttgart, Germany
cmueller@iits.uni-stuttgart.de)

Stefan Kroeger
(University of Potsdam, Germany
kroegers@uni-potsdam.de)

Abstract: In this contribution Wikis are interpreted as social information spaces. These information spaces can be decomposed in different networks. Here, one network is introduced – the collaboration network. This network type exemplifies how dynamics in social information spaces can be analyzed. For this, different approaches of visualizing networks are explained. The chosen approach is applied in an descriptive study. The open community project Wikiversity is examined to introduce one possible analysis in SONIVIS:Tool – an open source network mining software.

Key Words: Social Web, graph theory, network visualization, Wikis
Category: A.1, H.0, H.4.3, H.4.m, J.3, K.4.2, K.4.3

1 Emerging Sociability in the Web

During a conference brainstorming session, Tim O'Reilly and others have coined the term "Web 2.0". They wanted to characterize new techniques and services in the World Wide Web (WWW) [O'Reilly 2005]. But soon, this term was also applicable to a changing behavior of WWW users. More and more, the Web becomes a mirror image of the real world. With the dissolving cyber world the real social identity of an individual is transferred to a virtual equivalent. Meanwhile, the term "Social Web" is often used to describe this development. Here, Social Software plays an important role. Various applications are counted among it. One of them are Wikis. A Wiki is a *"freely expandable collection of interlinked Web pages, a hypertext system for storing and modifying information a database, where each page is easily editable by any user"* [Leuf and Cunningham 2001]. Wikis can be seen as social information spaces, which are formed by activities of authors and their content contributions. Existing dynamics in such information spaces are already investigated in terms of a temporal analysis of the Wikigraph [Buriol et al. 2006] or spatiotemporal knowledge visualization [Klamma et al. 2007]. Here, an new approach is introduced. This paper is organized as follows: first of all, a short introduction in the graph theory is given to describe collaboration networks in Wikis. Then, different approaches of visualizing dynamics in networks are explained. After an overview about the functionality of SONIVIS:Tool, an descriptive study is carried out.

Important developments in the Wikiversity project are illustrated and results of a time frame based network visualization are shown. This contribution closes with some conclusions and remarks about future work.

1.1 Applying graph theory

In Wiki-based information spaces, three domains can be defined: persons, knowledge/information, and tasks. Different perspectives can be specified based on these domains [Müller and Meuthrath 2007]. Every perspective contains at least one network and this networks supplies the context of interpretation. All networks in these perspective are interdependent [Carley 2003]. Here, one network definition is introduced more precisely – the collaboration network. Collaboration networks describe collaboration of authors on Wiki-pages. For example, author A_1 changes page N_1 at a particular time. After that another author A_2 changes this page N_1 at another time too. Therefore these two authors collaborate asynchronously. Collaboration networks can be formally described by applying graph theory. A collaboration network G_{CN} is an undirected graph. It consists of a set of authors (vertices) V_A. An author A is connected to an author B, when both authors revised the same Wiki-page E_C. In addition the collaboration network is a weighted graph. Each edge has a specific weight based on the number of collaborations W_{REV} of its connected authors. The combination of these three elements leads to the following description:

$$G_{CN}(V_A, E_C, W_{REV}), \{A_1, A_2\} \subseteq V_A \text{ of two actors } A_1, A_2 \text{ with } A_1 \neq A_2.$$

The objective of analyzing collaboration networks is to investigate the nature and extend of collaboration between persons in social information spaces.

2 Analysis of dynamics in collaboration networks

Basically, static and dynamic network visualization approaches can be differentiated. Static analysis deals with data which are accumulated over a specific time period. The network is visualized in a specific time frame (cp. [Gloor et al. 2003], [Moody et al. 2005], [Trier et al. 2007]). This time frame is part of a major time interval. Single vertices and edges are visualized, when the time frame is located in the interval of activity. The dynamic perspective of network analysis deals with the progression of state changes over time and enables the observation of events and its analysis. Dynamic network analysis allows for analyzing network states, vertices and their edges as well as changes in structure and configuration of the network [Carley 2003]. There are two kinds of visualizing dynamics in networks – cumulative analysis and sliding-window based analysis [Moody et al. 2005]. Their differ in the way of recording the network respective network parameters.

Cumulative analysis includes the aggregation of all changes and elements in a network over a long time period. The recorded strip can be interpreted as a complete image of the network. The network is growing over the period of time until it reaches the current state of the original network. All events in the considered interval are aggregated in one graph. The graph has an infinite memory of all active vertices and edges in the past and grows up to its final size which is equal to the static network's size at this time. In opposite to static network analysis, the network is investigated at different moments. Changes in the network structure are visible, but the sequence and existing interdependencies in network processes are not identifiable [Moody et al. 2005]. It is not sufficient to collect network information at different moments and then to subsume them [Doreian and Stokman 1997]. Events are not deterministic – they are stochastic. If two or three measuring points are selected randomly, then no or only unimportant events might occur. A evenly distribution of measuring points is possible, but it is without any sense, if all events happen at the same time. Therefore, another approach is applied here – the sliding-window. In the sliding-window based approach, a small period of time of the network or a sub-network is analyzed. A visualization of network data is carried out by a continuous time interval. The network is completely detached and based on single events. A sequence of these events can be evaluated based on time of occurrence and rhythm. Events can be specific attributes or positional changes. The size of the time frame depends on the number of occurred events. Only events in a specific time frame are aggregated to a network. The main objective of such investigation is to identify the impact of previous events on following events.

2.1 Analyzing information spaces using SONIVIS:Tool

Besides the specification of a visualization approach, a tool – SONIVIS:Tool – is developed[1]. SONIVIS:Tool is a Java-based open source software, which is based on Eclipse Rich Client Platform (RCP). Network visualization is implemented using the Prefuse Visualization Toolkit and metric calculation is done by GNU R, an open source software environment for statistical computing and graphics. The SONIVIS:Tool graphical user interface is based on completely configurable workspace layouts. At the moment, three main perspectives are predefined: *Analysis* (as standard perspective), *Manipulation*, and *Statistics*. The *Analysis* perspective contains different views, which provide users a quick overview about certain developments in Wiki information spaces. The *Manipulation* perspective allows for an enhanced visual manipulation of networks. There are three different layout algorithms implemented: Kamada-Kawai [Kamada and Kawai 1989], Fruchterman-Reingold [Fruchterman and Reingold 1991] and Circle Layout. A

[1] For more information have a look at http://www.sonivis.org.

visual filtering of the network is possible by minimal node degree and the minimal edge weight. Selected node metrics can be mapped on node size and color. Additional functions, like the network clustering, are offered.

3 Dynamic network visualization of Wikiversity

Wikiversity [Wikimedia Foundation, Inc. 2003] is a project of the Wikimedia Foundation. The objective of Wikiversity is *"to further the discovery and distribution of knowledge in a very natural way, by helping each other to learn."* [Wikimedia Foundation, Inc. 2003]. Important events of the German Wikiversity project are summarized in table 1.

Date	Event
2003-07	Idea of Wikiversity was published and discussed.
2005-02	First activities in the German Wikiversity.
2005-08	German Wikiversity went online.
2005-09-15 to 2005-11-01	Official poll about the autonomy of Wikiversity as a project ended.
2006-08-04	Jimbo Wales announced Wikiversity on the Wikimania conference.
2006-08-15	Launch of beta phase of English Wikiversity (6 month duration) was announced.
2006-08-24	Start of beta phase.
2006-08-25	Relaunch of the German Wikiversity.

Table 1: Important events of the Wikiversity project

In this explanatory scenario, the following question is answered: How do people work continuously in the Wikiversity project? Firstly, a copy of the database was download from the Wikimedia server [Wikimedia Foundation, Inc. 2008]. The first revision in the database was made on 2005-02-21 and the last revision 2008-02-05. In this period 70,929 revision where made and 29,322 links on 7,676 pages were created. The database size is about one gigabyte, whereas 13% are used by articles and 30% are used by discussion pages. 3,388 authors could be determined, whereof 1,212 are registered authors (35.8%). The Wikiversity project is a Wikimedia project, the platform is based on the MediaWiki software, which means that the data is stored in a MySQL database. In SONIVIS:Tool *DynamicCollaboration* is used as network type, the analysis period starts at 2005-02-20 and ends at 2008-02-06. Fruchterman-Reingold algorithm is used as visualization algorithm. The analysis period is divided into 36 time frames with an overlap of 5% of the frame size. The parameters form the configuration for the measurement and visualization with SONIVIS:Tool. In earlier analysis a number of 36 was evaluated as a good compromise between number of events per time

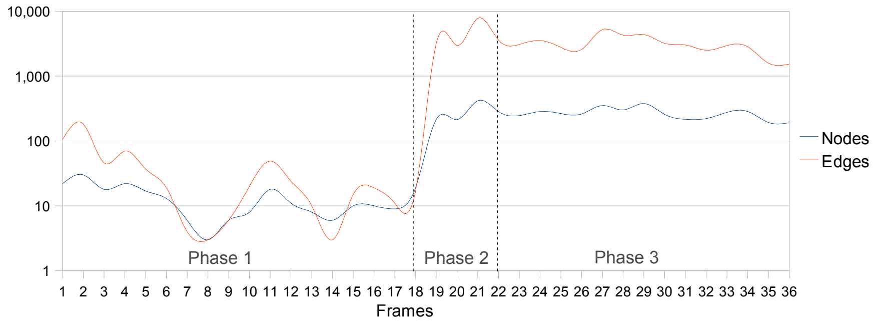
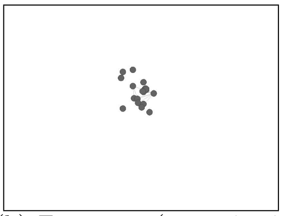


Figure 1: Development of number of vertices and edges in the investigated period (2005-02-21 to 2008-02-06)

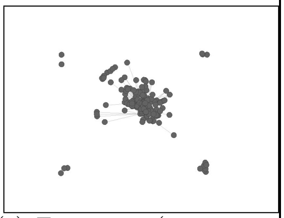
frame and rate of changes over time [Moody et al. 2005]. The average size of time frames is about one month. The overlap is used to avoid information loss at the time frame borders [Bender-deMoll and McFarland 2006].

Figure 1 shows the number of authors for every time frame. A visualized author (vertex) denotes that an author created or edited at least one article in the specific time frame. A strong increase of active authors can be recognized at time frame 18 (marked by left orthogonal dashed line) after a long time of less or no increase (time frames 1 to 17). The number continuously rises up from frame 19 until time frame 21 (marked by a second right orthogonal dashed line). After time frame 22 the number of authors stays on a high, but nearly constant level. Figure 1 gives a first impression of the evolution of network structure. Three phases of network evolution are visible. Also from the network visualization using SONIVIS:Tool changes of the network structure and number of authors becomes visible. The kind of visualization improves the recognition of changes in network topology. Figure 2 shows the visualization of time frame 17, 18 and 19. Especially the high increase of node count and edge count between time frames 18 and 19 suggests a transition in the network. It indicates a point in the evolution of this information space. Comparing these measuring results with information about the Wikiversity project (cp. table 1), we find a specific event which might cause these changes. The announcement of JIMBO WALES on Wikimania conference and the launch of the beta phase of Wikiversity in August 2006 have increased the popularity of the project. This assumption is supported by the activity measurement of the most active registered authors in figure 3. All authors which are active in more then four time frames are shown in this visualization. These are 103 of all registered users. In phase 1, only ten of them edited or created articles. From phase 2 (after time frame 19) on the majority contributes content to Wikiversity.

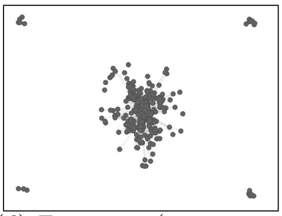
(a) Frame 17 (2006-06-14 to 2006-07-17): 9 nodes, 11 edges



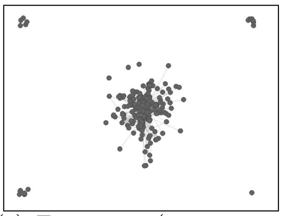
(b) Frame 18 (2006-07-14 to 2006-08-16): 19 nodes, 17 edges



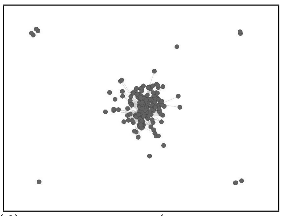
(c) Frame 19 (2006-08-13 to 2006-09-15): 216 nodes, 3,207 edges



(d) Frame 21 (2006-10-12 to 2006-11-14): 417 nodes, 7,782 edges



(e) Frame 22 (2006-11-11 to 2006-12-14): 284 nodes, 3,589 edges



(f) Frame 23 (2006-12-11 to 2007-01-13): 247 nodes, 3,082 edges

Figure 2: Screenshots of time frames 17, 18, 19 (above) and 21, 22, 23 (below) from dynamic network visualization of German Wikiversity

A second phase transition is visible between time frame 21 and 22, where the count of authors decreases slightly, but from this moment on stays on a high level. Later, changes are still minimal and an established amount of authors continuously work on the platform. Figure 2 shows the network visualizations for time frames 21, 22 and 23, where the number of authors decreased from over 400 (frame 21) to about 250 authors (frames 22 and 23). The comparison between the average node count of the phases substantiates this result. The average count of authors in phase 1 is 12, in phase two 257, and in phase three 246. In the beginning, the sliding-window based visualization of collaboration networks and the measured data show only few active authors. After a strong increase of active authors, the activity levels off at a continuous rate. Three development phases could be identified. Each phase outlines differences in the amount and intensity of authors work in the Wiki. Dynamic network visualization enables the identification of phase transitions. Additionally, user activities and their behavior in the social information space can be examined.

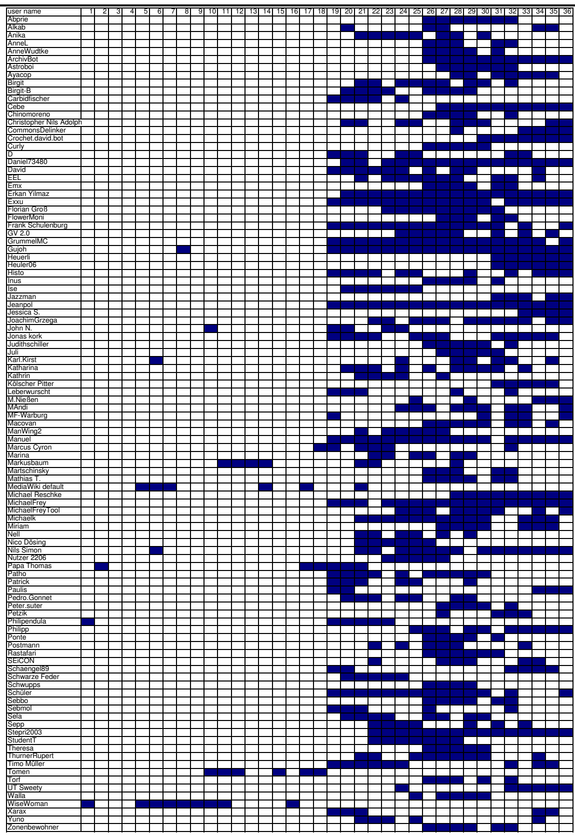


Figure 3: Activity matrix of the authors which work in more than four time frames (columns 1 to 36 represent the time frames)

4 Conclusions and Future Work

The implementation of dynamic network visualization extends the range of applications of SONIVIS:Tool. Using SONIVIS:Tool enables the investigation of dynamics in Wiki-based collaboration networks. The presented work provides a promising approach for analyzing network dynamics and network evolution of Wikis. It supports the identification and analysis of macroscopic changes of networks, like shown on the German Wikiversity platform. Phases of network evolution become reproducible and understandable. The analysis relates the external influence on the network to the resulting changes of structure and the behavior of the network element (authors).

As a next step, microscopic analysis of collaboration networks should be improved. Especially the influence of one author to another is an interesting question for further research. Dynamic visualization of the networks can be enriched by calculating and visualizing network metrics (e.g. centrality measures) and node metrics. In this way, more detailed information about reasons of network changes and influences on network structure and evolution can be uncovered.

References

[Bender-deMoll and McFarland 2006] McFarland, D. and Bender-deMoll, S.: "The Art and Science of Dynamic Network Visualization"; JoSS: Journal of Social Structure 2, 7, 2006, http://www.cmu.edu/joss/content/articles/volume7/deMollMcFarland.

[Buriol et al. 2006] Buriol, L. and Castillo, C. and Donato, D. and Leonardi, S. and Millozzi, S.: "Temporal Analysis of the Wikigraph"; Proceedings of the Web Intelligence Conference (WI 2006), Los Alamitos, CA, USA 2006, IEEE Computer Society http://www.chato.cl/papers/buriol_2006_temporal_analysis_wikigraph.pdf.

[Carley 2003] Carley, K. M.: "Dynamic Network Analysis"; Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, Committee on Human Factors, National Research Council 2003.

[Doreian and Stokman 1997] Doreian, P. and Stokman, F. N.: "Evolution of Social Network"; Routledge 1997.

[Fruchterman and Reingold 1991] Fruchterman, T. M. J. and Reingold, E. M.: "Graph Drawing by Force-directed Placement"; Software - Practice and Experience 11,21, 1991, 1129-1164.

[Gloor et al. 2003] Gloor, P. and Laubacher, R. and Zhao, Y. and Dyne, S. B.C.: "Temporal Visualization and Analysis of Social Network"; MIT, Center for Coordination Science, Cambridge, MA 2003.

[Kamada and Kawai 1989] Kamada T. and Kawai, S.: "An algorithm for drawing general undirected graph"; Inf. Process. Lett. 31,2, Amsterdam, The Netherlands 1989, 7-15.

[Klamma et al. 2007] Klamma, R. and Cao, Y. and Spaniol, M. and Leng, Y.: "Spatiotemporal Knowledge Visualization and Discovery in Dynamic Social Networks"; Proceedings of I-KNOW '07, 7th International Conference on Knowledge Management, Graz, Austria, Sept. 5-7, 2007.

[Leuf and Cunningham 2001] Leuf, B. and Cunningham, W.: "The Wiki Way - Quick Collaboration on the Web"; Addison-Wesley, New York 2001.

[Moody et al. 2005] Moody, J. and McFarland, D. and Bender-deMoll, S.: "Dynamic Network Visualization"; AJS Volume 110 Number 4, 2005, 1206-1241.

[Müller and Meuthrath 2007] Müller, C. and Meuthrath, B.: "Analyzing Wiki-based Networks to Improve Knowledge Processes in Organizations"; Proceedings of 7th International Conference on Knowledge Management, Graz, Austria 2007.

[O'Reilly 2005] O'Reilly, T.: "What is the Web 2.0? Design Patterns and Business Models for the Next Generation of Software."; 2005, http://www.oreilly.de/artikel/web20.html.

[Trier et al. 2007] Trier, M. and Bobrik, A. and Bartels, T.: "Towards Understanding the Dynamics of Communication Networks"; eOrganisation : Service-, Prozess-, Market-Engineering : 8. Internationale Tagung Wirtschaftsinformatik - Band 2, Karlsruhe, Germany 2007, 593-610.

[Wikimedia Foundation, Inc. 2003] Wikimedia Foundation Inc.: "Wikiversity (english)"; 2003, http://en.wikiversity.org.

[Wikimedia Foundation, Inc. 2008] Wikimedia Foundation Inc.: "Wikimedia Downloads"; 2008, http://download.wikimedia.org.