Visualizing Wiki-Supported Knowledge Building: Co-Evolution of Individual and Collective Knowledge

Andreas Harrer
Catholic University Eichstätt-Ingolstadt
Ostenstr. 14
85072 Eichstätt, Germany
+49 8421 93-1163
andreas.harrer@kuei.de

Johannes Moskaliuk,
Joachim Kimmerle
University of Tuebingen
Konrad-Adenauer-Str. 40
72072 Tuebingen, Germany
+49 7071 979-322/-346
j.moskaliuk@iwk-kmrc.de
j.kimmerle@iwk-kmrc.de

Ulrike Cress
Knowledge Media Research Center
Konrad-Adenauer-Str. 40
72072 Tuebingen, Germany
+49 7071 979209
u.cress@iwk-kmrc.de

ABSTRACT
It is widely accepted that wikis are valuable tools for successful collaborative knowledge building. In this paper, we describe how processes of knowledge building with wikis may be visualized, taking Wikipedia as an example. The underlying theoretical basis of our framework is for collaborative knowledge building with wikis, as introduced by Cress and Kimmerle [2], [3], [4]. This model describes collaborative knowledge building as a co-evolution of individual and collective knowledge, or of cognitive and social systems respectively. These co-evolutionary processes may be visualized graphically, applying methods from social network analysis, especially those methods that take dynamic changes into account [5], [18]. For this purpose, we have undertaken to analyze, on the one hand, the temporal development of an article in the German version of Wikipedia and related articles that are linked to this core article. On the other hand, we analyzed the temporal development of those users who worked on these articles. The resulting graphics show an analogous process, both with regard to the articles that refer to the core article and to the users involved. These results provide empirical support for the co-evolution model. Some implications of our findings and the potential for future research on collaborative knowledge building with wikis and on the application of social network analysis are discussed at the end of the article.

Categories and Subject Descriptors
H.5.3 [Group and Organization Interfaces]: Collaborative computing, Web-based interaction
K.3.1 [Computer Uses in Education]: Collaborative learning

General Terms
Design, Human Factors, Theory

Keywords
Visualization, Knowledge building, Wikis, Co-evolution, Collective knowledge

1. INTRODUCTION
In the last few years, wikis have established themselves as excellent tools for learning and knowledge-building processes. Many research papers and practitioner reports refer to wikis as knowledge-building instruments [6], [12]. Wikis are undoubtedly convenient tools for collaborative writing and – consequently – for collaborative production of knowledge. In this paper, we present a technique for making these processes of collective knowledge building visible (for other examples of visualizing processes in wikis cf. [19], [20]). We will, first of all, introduce Scardamalia’s and Bereiter’s knowledge-building concept and present a model that develops this concept further. We will then explain in general terms the basic ideas of social network analysis (SNA), and present the SNA software which was used here. We will explain how knowledge building can be visualized applying social network analysis. In conclusion, our findings will be discussed in the light of oncoming research.

2. KNOWLEDGE BUILDING
Scardamalia and Bereiter [14], [15], [16] describe knowledge building as a socio-cultural process that takes place in a community. A knowledge-building community aims at creating new products such as ideas, explanations, or theories that support members of the community in understanding their environment. A major purpose of knowledge building is innovation. Knowledge-building communities aim at producing collective knowledge and at improving ideas constantly. Members of such a community may use software to contribute their ideas, theories, or examples. Then they will refer to each other, bring up their own ideas, discuss, revise or sometimes reject them. According to Scardamalia [8], [13], knowledge building is successful when as many members of a community as possible contribute to the progress of knowledge. In this regard, knowledge building should be a discourse-oriented process. Besides, community members should deal with real-life issues and consider knowledge as an improvement of ideas rather than a search for the perfect answer.

Taking all this for granted, there remains a certain vagueness as to what exactly happens during knowledge building. A theoretical model that seeks to describe and explain knowledge-building
processes in more detail was introduced by Cress and Kimmerle [2], [3] (for another model that describes knowledge building cf. [17]).

The approach of these authors is based on systems theory and they regard knowledge building as an interplay between cognitive systems and a social system. A cognitive system is characterized by cognitive processes which take place in an individual’s mind – such as thinking, memorizing, or learning. In contrast, a social system consists of social processes in a community. In the case of wikis, such a process which takes places in a community is manifest through the development of the wiki and its articles. Here we have a case of knowledge building which involves both the cognitive systems of individuals and the social system “wiki”, both developing with equal speed and in terms that can be compared. Figure 1 shows the processes of externalization and internalization between the wiki and the users’ cognitive systems (CS) and the co-evolution of the different systems.

Wikis are a first-class example of successful knowledge building supported by social software [4]. This is true for two reasons. Firstly, with respect to wikis, users have many opportunities to influence their content. Secondly, wiki communities are perfect for inducing socio-cognitive conflicts and – at the same time – providing a framework for solving these conflicts. According to Cress and Kimmerle, socio-cognitive conflicts and their solution are the key “incitement” factor of computer-supported collaborative knowledge building. An advancement of knowledge will take place when individuals perceive a discrepancy between their own knowledge and the information that is available in a wiki. Then, they deal with the wiki’s content, add new information to the wiki, and acquire new knowledge at the same time. In this way, there is, so to speak, a co-evolution of individual knowledge and of collective knowledge.

Recently, this co-evolution process has been outlined using a fictitious example [2], it was described referring to one particular Wikipedia article [3], and it could be found in an experimental laboratory study [11]. What is missing, however, is a description of the development of a real-life wiki community and the corresponding development of a bunch of articles that this community is dealing with. We will provide such data in the following sections, with the aim of supporting the co-evolution hypothesis.

For making co-evolution processes graphically visible, we applied the technique of SNA. This method will be explained in the following section. Then our data and illustrative material will be presented, and analyzed and discussed in detail.

Figure 1: Processes of internalization and externalization between users’ cognitive systems (CS) and a wiki.
3. SOCIAL NETWORK ANALYSIS
Social Network Analysis [21] is a research method that investigates social relationships and interactions between different actors, e.g. to detect specifically important actors, potential causes for communication breakdowns, (sub-)groups etc. Originally it was rooted in social psychology in the so called sociometry of Moreno [10] and involved actors in the real world and analyzed networks by means of questionnaires and direct observation. In recent years, SNA has also been used extensively in virtual, computer-mediated settings: On the one hand, the availability of communication through the Internet allows interaction with others across boundaries. On the other hand, computer-mediated interaction may be logged and archived easily, so rich data for research with SNA are produced more or less automatically. In addition, the use of computers for SNA computations allows analysis of larger networks with much less effort by the researcher, once appropriate algorithms have been developed. Faced with these larger networks and a highly dynamic evolution of virtual social networks, researchers need sophisticated visualization techniques to show overviews and also a more detailed perspective of specific details. Typical techniques used to show dynamics include “time slices” at specific points in time [5] and animations where actors can be traced over time [18]. In our own work with the Weaver software [7] we use a three-dimensional integrated visualization that adds time as a third dimension to the usual two-dimensional sociograms. We also use a filtering instrument and grouping tools to zoom in on the interesting subsets of the network. Specific properties of actors and relationships may be graphically represented according to the researchers' intentions by using variations of size, color, shape, and labeling actors to make the visual diagrams more expressive [9] even at a glimpse. This will be explained in more detail in the practical example of the next section.

When looking at wikis as a computer-mediated interaction space, all the features which were mentioned, such as the availability of archives (via the version history of a page) and automatic computation of SNA traits, are present. Yet, with the presence of actors and mediating artifacts, i.e. the wiki pages, we are dealing with two different kinds of network entities, which we may refer to as a two-mode network. If different versions of the same page are also taken into consideration, the resulting network structure may be one of quite high complexity: Artifact links may exist between pages (wiki links) or may be of temporal nature between different versions. This complex structure also enables deep discussions, as is shown by a study on controversy in wiki pages [1] that analyses and displays degrees of controversy in one wiki page. In the following section, we will bring together the co-evolution theory, SNA methods, and visualization to support our theory based on a practical wiki example, and to show the usefulness of diagrams for interpretation and research.

4. VISUALIZATION OF CO-EVOLUTION
In this section we will first introduce the method and material that were applied in this study. This will be followed by a description of the visualization of the artifact and, subsequently, the visualization of the authors. Finally, the co-evolution of the artifact and its authors will be discussed.

4.1 Method
SNA and the Weaver software allow perfect visualizations of knowledge-building processes in wiki communities. As an illustrative example we chose the article about “schizophrenia” in the German version of Wikipedia. This choice followed a previous study by Moskaliuk and colleagues [11], who had used the topic “causes of schizophrenia” for experimentally examining the co-evolution process and found this topic to be qualified to provoke significant cognitive conflicts. This is because there are originally two lines of research or general models of the causes of schizophrenia, and one theory that attempts to combine and merge these two models.

The first model or school emphasizes biological or genetical triggers for schizophrenia. The other one suggests social factors such as psychological stress to be the main reason for schizophrenia. And, eventually, the diathesis-stress model integrates the social and biological approach into one model by stating that it is external stress that may potentially uncover a person’s inborn vulnerability. There is also a psychoanalytical explanation for schizophrenia, but this approach tends to be regarded as an outsider position. So, on the whole, a presentation of the causes of schizophrenia is a rich field for potential socio-cognitive conflicts. Consequently, we took a closer look how the schizophrenia article (and related articles that were linked to it) developed over time, in order to observe knowledge-building processes.

For the data set analyzed here we used the immediate “neighborhood” of the schizophrenia page (i.e. the pages directly linked to it) and selected those pages that represent the two general models most accurately (according to an expert's advice). Then we used the export function of Wikipedia that allows creating an XML document which contains all the versions of a specific page, and then combined the pages to form our research corpus. This corpus allows a fine-grained analysis which includes all authors and article versions, but also a goal-oriented selection to reduce the data, e.g. by filtering out all minor versions and / or anonymous edits. This XML structure (raw size 500 MB because of the sheer size of the text content of the versions) was then parsed using an incremental approach or, in other words, we processed smaller parts of the document in order to make it possible to scale it on computers with little memory, and the result was stored as an internal network representation containing all versions, users, and relations. To constrain the data for analysis we did not use the full structure (some pages had up to 2000 versions), but decided to observe the data at six specific points in time, January 1st in each year from 2003 to 2008.

Due to the data richness that wikis provide, an analysis can look at wikis and their dynamics from several different angles. For the co-evolution theory two perspectives are particularly relevant. Firstly, the artifact network between wiki pages is important, because interlinkage between pages can indicate the development of different concepts. It can illustrate how concepts interact with
each other and it may potentially even describe the integration of concepts. In other words, visualizing the development of wiki pages and their interconnection yields a representation of the progress in and the dynamic of the social system. Additionally, it is also very interesting to analyze the authors of wiki pages and to observe developments and changes in these authors’ affinities over time. When one person is the author of a specific page that refers to one specific model (biological, social, or psychoanalytical respectively), this would indicate that this author belongs to a certain community of interest and, consequently, that she or he is endowed with respective cognitive requisites. When the same author also starts participating in another community, this may be taken as an indicator for changes in her or his cognitive system. Then, there seems to be an integration of different points of view in the author’s cognitive system.

4.2 Visualization of the artifact

In our first perspective on the data, we use the artifact network, i.e. the wiki pages and their interlinkage, at six points in time. We computed SNA traits, such as centrality and density [21]. In order to visibly represent the centrality of pages, the size of the graphical representation of a page is displayed proportionally to the number of references to that page. Figure 2 shows the overview representation containing the six dates of inquiry: the time axis goes from left (2003) to right (2008) and the three-dimensional representation has been tilted slightly so that each year can be seen as an oval slice.

![Figure 2: Overview of the whole page structure (slices Jan 1st 2003 - 2008)](image)

Each page is kept constant in its 2D-coordinates so that different versions of same page can be identified by following from left to right, highlighted also by a “history line” drawn from the version valid at time $t_i$ (e.g. 2004) to the version valid at the next point in time $t_{i+1}$ (2005 in the example; if there was no new version within that year, the line points to the next available year). Pages that did not have a valid version at a given point in time (the respective term may have had a page at one time during the observation, but not necessarily at the beginning) are represented by bright squares. If links to these squares are shown, they are so-called “red-links” – links to a page not filled with content yet.

The first-glimpse perception is that the size of the pages and the number of relationships has grown substantially over the years. In the last dates of inquiry – 2007 and 2008 – the density of links is much higher than in previous years. In the time slice for 2003 there are numerous bright squares with some red-links, and very few links between those pages that had already been defined.

Using filtering techniques, we took a closer look at each point in time and used a grouping algorithm using k-cycles to detect close connections among sets of pages: pages that are connected via a cycle of length 3 are assigned to the same group and are visualized as dark squares in a specific circle at the outer rim of the diagram. So called “floaters” or “boundary spanners” are pages that connect to more than one group and permit an exchange of information and are therefore particularly interesting for the co-evolution theory; these are represented as gray squares. The remaining pages, i.e. those that do already exist (in contrast to red-links) but are neither floaters nor part of a group, are represented by small triangles. Figures 3, 4, 5, and 6 show the artifact network in 2003, 2005, 2007, and 2008 respectively.

![Figure 3: Artifact network observed on Jan 1st 2003](image)

Figure 3 is mainly shown to highlight the contrast to the other dates of inquiry: In the year 2003 very few pages were filled with content, most of them only existed as red-links (bright squares). Only few links existed at that time, some between existing pages and some as red-links as an indication that the structure might expand over time with new pages that were still waiting to be written.

In the diagram for the years 2005, 2007, and 2008 we only labeled the “grouped pages” and the floaters (which are the most interesting pages by virtue of their network position) in order to have a better overview. In the illustration for the year 2005 (Figure 4) two clusters of pages can be seen that are linked through short cycles, indicating a strong contextual connection. These clusters can be roughly assigned to the psychoanalytical
model (top left) and to the model that suggests biological factors as the causes of schizophrenia (down right).

Figure 4: Artifact network observed on Jan 1st 2005

The psychoanalytical cluster is represented by pages such as “Sigmund Freud”, “Über-Ich” (super-ego), and „Ich“ (ego). Pages like “Nervenzelle” (neuron), “Gehirn” (brain), and “limbisches System” (limbic system) on the bottom right describe neuropsychological topics and can be assigned to the biological explanation. For 2005 there is no cluster on social aspects of schizophrenia.

Interestingly, in 2005 the schizophrenia page itself had the position of a boundary spanner, which indicates that the different scientific positions might influence (or have already influenced) that page. For the whole structure, the number of valid pages and of links has increased substantially, if compared to Figure 3; this can be checked using formal SNA traits (density, centralization) but may also be seen outright in our visualization (again, larger representations indicate a higher number of references / links to that page).

In the diagram for the year 2007 (Figure 5) the number of links to the pages have increased again, and many pages are densely connected (three clusters of considerable size as well as a small one; many of the remaining pages are floaters). While there is still a separation of the two clusters representing the psychoanalytical (bottom left) and biological (top right) point of view, there is also a new cluster at the top of the diagram. This cluster can clearly be assigned to the social model, as it deals with such topics as “Metakommunikation” (meta communication), “Paul Watzlawick”, or “Doppelbindungstheorie” (double-bind theory).

By comparing the diagrams for the years 2005 and 2007 we can describe an increase of links between the clusters. In 2005 the psychoanalytic cluster is grouped separately: Only one link connects this cluster with the other clusters. In 2007 there are much more links connecting the psychoanalytic cluster with the rest of the artifact network.

Figure 5: Artifact network observed on Jan 1st 2007
Additionally, there are new pages in 2007 which constitute one cluster but are also linked with other clusters. Examples for these pages are “Influenza”, “Viren” (viruses), or “Infektion” (infection) which clearly belong to the biological explanation.

In the time slice for 2008 (Figure 6), the biological and social clusters have merged into one cluster. The psychoanalytical cluster, however, is still separated in the 2008 graphic.

4.3 Visualization of authors
The second perspective of our analysis was the development of authors or users who deal with these articles. Our assumption was that an author who contributes to a wiki page that mainly belongs to a certain explanation model (biological, social, psychoanalytical, or diathesis-stress model) can be considered as belonging to a community of interest in that model. Even if we bear in mind that authoring a wiki page will not necessarily make a person a supporter of the position that is explained on that page, it can be regarded as an objective indicator for belonging to a community, which can be measured easily.

If the participation of certain authors in a community changes over time this will clearly indicate a development of those authors and their cognitive systems. According to the framework for collaborative knowledge building with wikis [2], these developments are results of socio-cognitive conflicts between individual knowledge on the one hand and the information contained in the wiki on the other hand.

To determine the existing communities, we enriched the original wiki data set with categories: experts decided for each page to which explanation models it belongs and assigned each page to a category. The categories used were “social”, “biological”, “psychoanalytical”, “integrated / diathesis-stress model”, and “indistinct”. For each creation of a version by an author the category of the respective page was checked and taken into account for assigning this author to a specific community. This was carried out using different parts of the edit history to see if the category with a particular author was mainly working on had changed over time. All anonymous users and those users who only made minor revisions (like correcting typing errors) were filtered out for the analysis. This procedure guaranteed that the assignment to a community was not simply the result of editing a spelling mistake or other minor correction, but indicated a significant contribution to a wiki article.
Figure 7 shows a diagram of the author development over time provided by the Weaver tool: Each user who edited at least once the schizophrenia page and at least one other page in the course of a year (making him or her a candidate for contributing to an integrated perspective on the schizophrenia page) is shown with her or his main category / community. But community memberships changed over time. The symbols in the midmost circle represent community membership between January 2nd 2007 and January 1st 2008, the symbols in the second midmost circle represent community membership between January 2nd 2006 and January 1st 2007, and so on. Each category is shown with a different shape and gray tone. Users who belong to the “social” community are represented by bright circles, users of the “biological” community are shown as black squares, and users with “psychoanalytical” orientation are represented by dark circles. Users who worked on pages that integrate the two models (diathesis-stress model) are shown as dark-gray triangles (in order to avoid an overload of information in this figure, we did not depict those authors who worked primarily on articles that were labeled “indistinct”).

The diagram shows that several authors changed their main community over time, which is visible through the change of symbols over time. A closer look at Figure 6 shows that there were authors who shifted from a biological to a social perspective and others who developed the opposite direction. In addition to that, we find that some authors shifted from a psychoanalytical perspective to an either social or biological point of view. In other words, these authors developed from a marginal to some form of mainstream explanation of the causes of schizophrenia.

The figure also suggests - and this is the most important aspect in this context - that many authors shifted from a biological or from a social perspective to an integrative point of view. Members of the psychoanalytical community, however, did not develop in this direction.

4.4 Co-evolution

An analysis of the data provided by sections 4.2 and 4.3 and a comparison of the development of the wiki on the one hand and the authors involved on the other hand provides clear evidence for the co-evolution hypothesis.
As far as the development of the artifact network is concerned, there is a convergence of the social and the biological position. In 2008, the biological and social pages were not arranged as two distinct categories any more but as one common cluster. As a result of a socio-cognitive conflict between users with different conceptual attitudes about the causes of schizophrenia, the different positions merged. The same development can be observed with regard to the authors who contributed to these articles. Here, there is also a trend from contributing either to social or biological articles to now contributing to both types of articles and to articles that represent an integrative point of view.

The progress of the wiki articles and the development of the authors and their cognitive systems are the two components of the co-evolution process. The parallel development of the wiki and its authors indicates a co-evolution of the social system and the individuals’ cognitive systems. Thus, what occurred here is obviously a situation in which authors take up information from the wiki and integrate new aspects into their individual cognitive system. And this, in turn, changes the social system wiki and so on.

This supports the co-evolution theory, particularly since a closer look shows that there is an obvious tendency towards the integrative diathesis-stress model (dark-grey triangles). At the same time those authors who contributed to integrative articles do not shift back to one of the original (biological or social) perspective; once they belong to the integrative community they seem to stick to it.

5. CONCLUSION

This paper and the data provided here show that SNA is a fertile approach for making wiki-supported knowledge-building processes visible. SNA is a method that is able to deal even with a huge amount of data. This allows the examination of large systems such as Wikipedia. Analyses like this one do, however, require laborious editing and processing of data in advance. Consequently, it is helpful if researchers know what they are looking for, or in other words, if they follow a theory-driven approach that leaves room for testing concrete hypotheses such as the co-evolution hypothesis in the present case. But SNA is also an interesting procedure for exploratory and merely descriptive investigation. An analysis of that type may, of course, also be used for generating future research hypotheses.

Although our study can be considered as successful, there are also some potential shortcomings that should be addressed: Certain methodological aspects could be handled differently in future studies. For example, the selection of wiki articles and the assignment of wiki articles to certain categories could have been conducted in a more systematic manner. The same is true of the selection of authors (excluding some authors because they have only done “minor editing” may be a very personal choice). Using a larger number of pages for categorizing users may lead to a higher degree of accuracy. This would also lead to a more inductive mindset and imply a higher validity of the findings. A new study could, for example, analyze the social network of all users inside the psychology portal of Wikipedia. Various language versions of this portal provide a large number of articles on various topics of psychology (and this term itself). But for this purpose, the problem of handling very large files and a high amount of information has to be resolved. It will be necessary to find some procedure for minimizing the research corpus.

Looking at Wikipedia, the user pages afford additional fascinating data which could be used for SNA, too. Users often write on their own user pages which articles they have started to write or revised intensively. It would be interesting to use these data for alternative visualizations of the respective authors’ communities.

In addition, it may also be useful to look more closely into a different domain of knowledge or different language version of Wikipedia to analyze the co-evolution of wiki articles and authors. This may turn out new evidence for the co-evolution model and would also stimulate the progress of the theory of knowledge building with wikis and other social software.

Of course, the method which is presented here will also work with the analysis of smaller wikis with smaller communities, such as wikis in a class at school or university, or wikis in industrial organizations.

As a next research step, we will use the SNA in the same way as was presented here in order to analyze data from a laboratory experiment. We will not only measure the authors’ community membership by looking at their edits, but also manipulate this community membership by giving authors different material as a basis for their work with the wiki. By examining smaller groups of users with different opinions and different information about a certain topic we hope to gain a deeper and more precise insight into the processes of collaborative knowledge building.

Another research goal, dealing with the visualization of authors, could be to gain deeper insights into the motivation of contributors to a specific wiki page, into their academic backgrounds and disciplines, and into their belonging to a community. Instead of only looking at artifacts, we may find some evidence for cognitive changes over time – and would confirm our findings on co-evolution between the users’ cognitive systems and the social system wiki –, if we asked those people who were involved directly in a qualitative interview or through a questionnaire. One problem could be that it is rather unlikely that users will remember their thoughts and motivations from three years ago (e.g. according to the artifact network that was observed on Jan 1st 2005) in a reliable way. So the best approach would probably be to conduct a study observing the co-evolution of a wiki in the form of a longitudinal analysis, and interviewing the authors several times during that study.

Our initial application of a two-mode network SNA for wiki research appears to be a productive procedure in that our findings have provided novel insights into how knowledge-building processes with wikis take place. At the same time, they have identified the strengths and weaknesses of this information-visualization technique.

6. ACKNOWLEDGMENTS

The Weaver application and partially the visualization methods have been developed during the first author's collaboration with his colleagues from the COLLIDE research group at the University of Duisburg-Essen (UDE).

The third author of this paper is indebted to the LANDESSTIFTUNG Baden-Württemberg for the financial support of this research project by the Eliteprogramme for Postdocs.
7. REFERENCES


