

Analyzing Wiki-based Networks to Improve Knowledge Processes in Organizations

Claudia Müller

(University of Potsdam, Germany
cmueller@wi.uni-potsdam.de)

Benedikt Meuthrath

(University of Potsdam, Germany
bmeuthrath@wi.uni-potsdam.de)

Abstract: Increasingly Wikis are used to support existing corporate knowledge processes. Wikis are an appropriate software solution to support these processes because of their specific design principles. However, it is not yet proven whether Wikis are an adequate knowledge management tool. This paper presents a new approach to analyze existing networks in Wikis. Four perspectives on Wiki networks are introduced to investigate all dynamic processes and their interrelationships in a Wiki information space. The Social Network Analysis (SNA) is used to uncover existing structures and temporal changes. Additionally, a data set of a corporate Wiki is used to present an analysis which has been carried out. Results of analyzes work can be utilized to improve the existing knowledge processes in a company.

Keywords: Social Software, Wiki, Knowledge Work, Network Analysis, Collaboration Network

Categories: A.1, H.0, H.4.3, H.4.m, J.3, K.4.2, K.4.3

1 Introduction

Formerly centralized hierarchical organizations are increasingly transforming to decentralized and network organizational constructs, which promote more openness. In these systems, which are subject to continuous change processes, people and available information are key resources [Tapscott, 93], [Barabási, 03]. Flat hierarchies and flexible structures, team orientation as well as a progressively growing dependency on knowledge are fundamental characteristics of companies today [Cross, 02]. These changes in organizational structures have also impacted corporate knowledge management. To ensure organizational learning, knowledge communities should be fostered [Romhardt, 02]. These knowledge networks within an organization, which are not organizationally authorized, were initially analyzed in the context of Communities of Practices (CoP) [Wenger, 98]. Information Technology (IT) is one important enabler for a successful knowledge community. In this context Wikis are increasingly used for decentralized, collaborative knowledge management of corporate knowledge communities [Klobas, 06], [Tapscott, 07]. The first Wiki, implemented by Ward Cunningham was designed based on various design principles [Leuf, 01]. These principles influence Wiki software, the user behavior and the application of knowledge management. A Wiki is “open”, implying that each

person (employee) can consume and edit every page. Information (externalized knowledge) in a Wiki is mergeable or dividable depending on existing knowledge requirements. Also, information context is arbitrarily adaptable by linking a specific article. When users externalize their knowledge they can link it to existing or non-existing articles (displayed in another link color). In the latter case, a Wiki enables users to formulate their existing knowledge needs. A Wiki information space has initially no predefined structure. The structuring of information objects (articles) is specified by a collaborative self-organized process controlled by the users. All article versions are saved in a history. A critical issue of knowledge work is solved in this way. Usually the final product of knowledge creation is visible but not the actual process of knowledge creation. With the page history the process from the first draft to the final version is traceable. Despite all known advantages it is still not proven whether Wikis are an adequate knowledge management tool to foster knowledge communities in organizations. This paper presents a novel approach to analyze existing networks in Wikis. The Social Network Analysis (SNA) is used to uncover existing structures and its temporal changes. An analysis of a corporate Wiki shows first results of this approach. Two hypotheses constitute the base for the presented research in progress:

- Social Network Analysis is an adequate method to analyze Wiki information spaces and
- The collaboration network shows the nature of cooperation in a Wiki and reveals special roles, like the Wiki-Champion.

This paper is organized as follows: the next section introduces social network analysis and possible application areas in knowledge management. Moreover, basic SNA metrics are explained shortly. In the following paragraph four perspectives of Wiki-specific networks and their characteristics are described. Finally one part of a conducted exploration of a collaboration network in a company is illustrated and results are evaluated. The third section gives an insight in future tasks.

2 Applying Social Network Analysis in Knowledge Management

Social network analysis (SNA) is an enabler to systematically describe and investigate network structures and their principles of order in organizations. Basically, a social network consists of a set of persons and relations defined in them [Wasserman, 97]. The social environment is described by a pattern or regularity among the interacting persons. Social relationships of persons constitute the social capital of an organization. The social capital metaphor is that people who do better are usually better connected. SNA is increasingly applied for analyzing and reorganizing distributed communication and collaboration structures in knowledge management (e.g. [Cross, 02]). Existing (cross-functional) relations are determined in their relation strength and intensity. Expertise within knowledge networks can be located.

The required data for network analysis can be gathered electronically or via interviews. A common approach utilizes user interaction with computer systems (e.g. Email, Blogs, Wikis) as applied in this contribution. Network analysts use matrices and the graph theory to represent information about patterns of ties among social actors. A network can be abstracted as a graph where nodes are described as vertices

and relations as edges. This graph is a model of a network with unidirectional dichotomic relations [Wasserman, 97]. The network analysis offers a range of metrics to analyze social structures in a systematic manner. Measures to investigate the structure of the network are for instance *density*, *average distance*, and *average degree*. These metrics are used here to describe the network structure on a macroscopic level.

Density of a graph is the ratio of relations over number of maximum possible relations [Wasserman, 97]. The highest possible number of edges is reached if each vertex of the graph is connected with each other vertex in the graph. The (geodesic) distance is defined as the length of a geodesic between two vertices. The average distance (average shortest path between all vertices) is calculated for the whole network. This metric is equal to the degree of separation which connects all pairs of nodes in a group. Short distances transmit information accurately and in a timely way. A long distance degree implies slower even distorted information transmission [Cross, 02]. The degree of a node is the number of edges that are connected to this node (adjacent nodes). This measure helps to evaluate the “attractiveness” of a node. The average degree is calculated for all nodes in the network. On a microscopic level, the position of a person in a network can be investigated with the metric *degree centrality*. This metric helps to identify the “most important” actor, meaning those that are extensively involved in relationships with other actors in the social network. A vertex is central in the sense of degree centrality, if a vertex has many relations to adjacent vertices.

All these measures can be utilized to investigate dynamic characteristics of networks as well. The dynamic perspective of network analysis deals with the progression of state changes over time and enables the observation of events and its analysis. To determine changes of any form in the network it is necessary to compare at least two different network states at different times. Dynamic network analysis (DNA) enables analyzing network states, the vertices and their edges as well as changes in structure and configuration of the network [Carley, 04]. The cumulative analysis enables the aggregation of a long time period of the network, including all changes and elements, which are or were part of the network during the period. The calculated measures and their changes are visualized in a diagram.

2.1 Specifying Wiki Networks

A Wiki with all related activities forms a specific information space in a company. This information space consists of different networks. Based on a classification made by Carley [Carley, 04], Wiki-specific networks can be arranged in four categories (Figure 1): social perspective (who knows who), knowledge perspective (who knows what), information perspective (what refers to what), and temporal perspective (what was done before). Relationships in one network usually imply relationships in another [Carley, 04]. Therefore, to understand all knowledge processes in a Wiki an analysis of these different network types is necessary.

The first group contains collaboration networks and discussion networks. Both types are social networks (cp. Figure 1). A *collaboration network* is used to investigate the nature and extend of collaboration between authors of Wiki articles and it enables the analysis of information exchange between communities. Collaboration is based on mutually referred asynchronous generated contributions in a

knowledge process. In a collaboration network vertices are authors, whereas the edges are constructed on common edits of an article. Whenever two or multiple authors worked on the same article in a Wiki, they are connected to each other. The central assumption is that changes of different authors on an article imply collaboration. A strong tie exists if two authors have contributed comparatively more than other authors (in terms of content) or they work collaboratively very often on one or more articles (in terms of frequency). By analyzing collaboration networks the time component has a high impact because cooperation is possible over a long period. Consequently, the effect of aging on network structures has to be considered [Zhu, 03]. With the second network in this perspective - the *discussion network* - a topic specific communication process on a Wiki talk page can be studied. Vertices are authors of specific postings and edges are related postings. The hierarchical order of these postings is defined by the time of edits. An activity level of discussions in a Wiki information space is measurable. The last network in this perspective is the *message exchange network*. Each registered user can exchange messages. This exchange is carried out by an edit on a personal talk page (like an entry in a guest book). Users get informed about new messages on their personal talk pages for instance via email. Nodes are users and relations exist based on exchanged messages in this network. This network supports analyzing and evaluating the amount and characteristics of communication processes in Wikis.

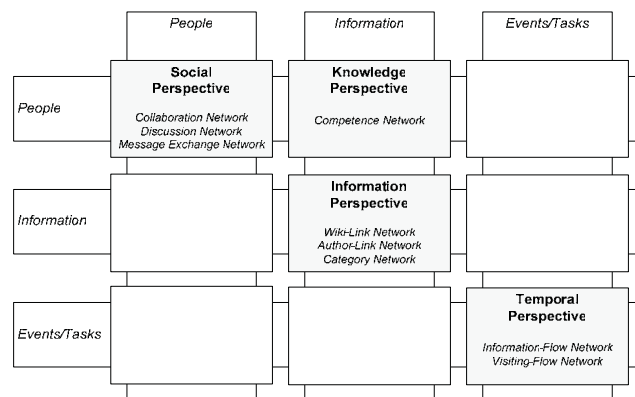


Figure 1: Specification of Wiki networks depending on different perspectives

The second perspective comprises of knowledge networks (cp. Figure 1). Wiki specific realizations are *competence networks*. This type of network shows existing thematic competences of authors. It is a two-mode network which means there are two different types of nodes – authors and articles. The edge between two vertices (article, author) represents the editing activity of a certain author for a given article. An investigation of a competence network gives an impression which author works on what article very often and/or has contributed a lot to one article. This is an index of existing knowledge in a specific topic area. A main assumption in this network is that if authors edit articles they externalize their knowledge as well. Minor changes, e.g. typographical errors, are left out of these considerations. An interpretation of

competence networks is due to their character more complicated than in the networks already introduced.

The information perspective defines information networks (cp. Figure 1). In the Wiki context there are wiki-linked networks, author-linked networks, and category networks. A *wiki-linked network* consists of articles as vertices and wiki-links as edges. It is a directed network since wiki-links are not reciprocal. The temporal analysis of wiki-linked networks enables the investigation of structural developments of themes in Wikis. The evolutionary process of emerging and “dying” topics is visible in this way. A structural analysis based on degree centrality shows highly connected areas in this network. Identifying topic clusters is another structural approach. Existing theme affiliations can be detected by text analysis. The second network in this perspective is the *author-linked network*. It focuses on articles with a high modification rate. The vertices are articles and edges represent changes on both articles by a certain author. The weight of such a relation increases with the number of authors which work on both articles. Opposite to the wiki-linked network, these network centers emerge based on thematic relations of authors and not of articles. Furthermore, this network describes the “attractiveness” of a single article in the information space. One disadvantage of this network is, that authors, who are active with one article only are neglected. Network usefulness decreases, if the number of such authors is high in the investigated information space. The third network is a simple hierarchical network – the *category network*. In Wiki information spaces content is often structured in categories. Besides a hierarchical structure of categorization there are also cross-links between the various category entries. One single category consists of thematic similar topics. The objective of this network is to describe these categories and their utilization (the more articles a category has the higher is the diameter of nodes). The category network comprises category pages as vertices and main-category/sub-category relationship as directed edges. Based on a temporal analysis the development of articles in specific categories can be studied.

The fourth perspective comprises information-flow and visiting-flow networks. Especially in specific topic areas, an analysis of the temporal development of articles and their interdependencies is interesting. The *information-flow network* sequences Wiki articles. The nodes are articles and the relations are defined based on the article history and/or the day of creation. Therefore, these relations show the temporal ordering of article creation or change. The measuring of an information flow from one article to the other is not trivial and time is not a sufficient parameter. Rather the content of articles has to be considered. This means a relation should be defined by transferred text from one article to the other. In the Wiki context it may be interesting to gather information about how the Wiki is used not only by its contributors but also by its readers. The *visiting-flow network* sequences Wiki pages by its temporal order by means of the referrer web server log. A referrer is the page a reader visits directly before the actual page. This information is sent by most browsers and is logged by web and proxy servers in default configuration. Assuming that there are two starting points (external and no referrer) the referrer information is sufficient to generate a directed network with Wiki pages as nodes and their referrer relationship as edges, weighted by its frequency. To clean up results editing and resource access (style sheets, java scripts etc.) should be filtered.

2.2 Analyzing and Evaluating Wiki Networks to Improve Knowledge Work

The predefined Wiki networks are applied to enhance existing network analysis approaches. Using these networks provides a comprehensive insight into existing knowledge processes in a Wiki in terms of people interaction, topic emergence, communication activities and their dynamic changes. Here, the collaboration network is visualized and analyzed to emphasize this new approach. The data set used is based on a Wiki that was launched in the summer of 2005. It serves as a knowledge exchange platform for about 1,000 employees. In April 2007 the corporate Wiki had 185 users and about 8,200 articles. To conduct these analyses a special tool was implemented. The SONIVIS:Tool provides statistical analyses of Social Software. It is based on the eclipse Rich Client Platform (RCP). There are predefined graphical analyses which offer a quick overview on recent developments. Additionally, various SNA metrics are provided on a microscopic and macroscopic level.

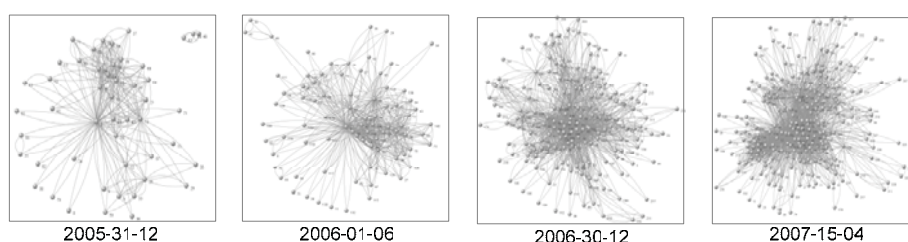


Figure 2: Temporal development of a collaboration network

Figure 2 yields a first insight into the development of the collaboration network. At the beginning of the investigated timeframe (2005-31-12) there is one specific area (sub-network) in which specific vertices are closer than the others. There is also one component with no connection to the main network. This picture changes over the investigated period (2006-01-06). The network is growing. More and more authors join the network and existing collaborations get stronger in certain areas. In the third snapshot (2006-30-12) there is one specific sub-network showing strong relations, but this group has grown compared to the year before. There are still a number of lightly connected nodes in the periphery of the collaboration network. There are maybe only a limited number of authors or they work only on specific articles in the Wiki. The last visualization (2007-15-04) shows three subgroups. A conducted analysis reveals that these subgroups have been formed around specific projects. Therefore, it will be interesting to examine how the Wiki structure is changing because of external, organizational changes.

A pure visualization of a network is not sufficient for understanding the whole evolutionary process. Hence, specific metrics should be used to analyze the network structure on a macroscopic and microscopic level. Figure 3 shows on the left the temporal development of the network's density and on the right side the temporal development of the average distance. Both measures allow an evaluation of the network on a macroscopic level. The density declines continuously during the considered period. However, Figure 2 shows that there are specific zones in the network with a density of almost 1 (so called cliques). The average distance increases

during the evaluated period. During the 2005-10-09 and 2006-10-04 period the network has two components (unconnected sub-graphs). Therefore, the average distance is infinite, in the diagram equal to zero. Overall, the average distance increases which is a result of the growing number of authors and a decrease of the average degree (number of realized connections). Consequently, the spread of information in this network needs more time. A correlation analysis was constitutively conducted based on these results to reveal existing relations between the collected measures.

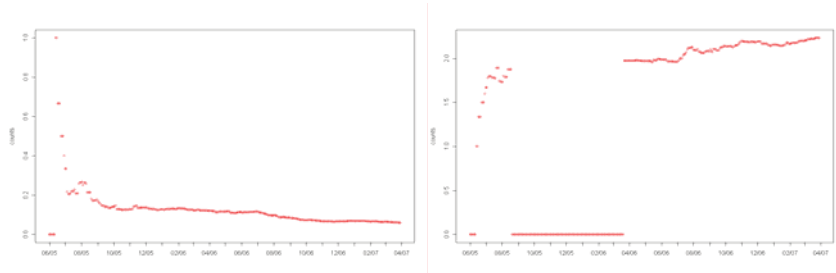


Figure 3: Density (left side) and average distance (right side)

A positive correlation between article count and average degree (0.85) and a negative correlation between density and average distance (-0.34) was determined. The first result is attributed to a minor collaboration in the whole network but a high collaboration in a subgroup. This is in accordance with the results of visualization (cp. Figure 2). A negative correlation implies that the average distance increased due to emerging subgroups. Simultaneously, the density decreases, but not at the same rate.

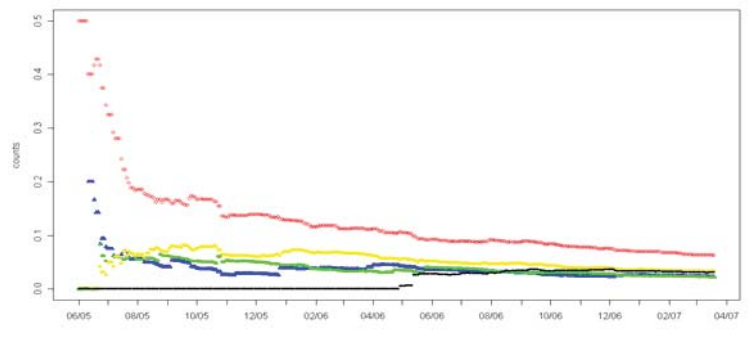


Figure 4: Temporal change of degree centrality on a microscopic level

This result is an indication for existing hubs in the network. Hubs are persons with a high level of activity because they are connected to a lot of persons through different articles. In early stages of a Wiki so called Wiki-Champions have a critical importance for the development of a Wiki information space. They are recognized as early adopters who understand how to use a Wiki very well, encourage others, and get

people involved by informally training them and being available for ongoing support. They serve as a model in using Wikis (<http://www.wikipatterns.com>). In this case one extremely active person can be identified in the beginning. As time moves on this person loses its special position gradually. But there is no reason for concern because the activity level in the Wiki is increasing overall. Also at the same time other persons gain importance in the network (cp. in Figure 4 the top curve). More and more people join the Wiki-Champion and engage more people in the Wiki. Based on these selected results the SNA enables the analysis of existing collaboration in a Wiki and reveals specific Wiki roles such as Wiki Champions.

3 Outlook

This paper shows possibilities how to analyze knowledge activities in terms of collaboration in a Wiki. The network analysis can be used to enable a measurement of an existing network. Due to the complexity of such systems four groups of networks are defined to analyze existing interdependencies. This work was done by a research group - the SONIVIS:Team (<http://www.sonivis.de>). In the future, all defined networks will be implemented to enhance analysis work and to improve our understanding of social processes in a Wiki information space.

References

- [Barabási, 03] A.-L. Barabási: “Linked”; Plume Printing, New York (2003)
- [Carley, 04] K. M. Carley: “Dynamic Network Analysis”; In: R. Brelger K. Carley, P. Pattison, (eds.): “Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers” National Academy Press, (2004), 133-145.
- [Cross, 02] R. L. Cross, A. Parker: “The Hidden Power of Social Networks: Understanding how work really gets done in organizations” Harvard Business School Press, Cambridge (2004)
- [Klobas, 06] J. Klobas: “Wikis: Tools for Information Work and Collaboration“ Chandos Publishing, Oxford (2006)
- [Leuf, 01] B. Leuf, W. Cunningham: “The Wiki Way - Quick Collaboration on the Web“ Addison-Wesley, New York (2001)
- [Romhardt, 02] K. Romhardt „Wissensgemeinschaften. Orte des lebendigen Wissensmanagements“ (in German), Versus Verlag Zürich (2002)
- [Tapscott, 07] D. Tapscott, A. D. Williams: “Wikinomics“ Penguin Books Ltd., London (2007)
- [Tapscott, 93] D. Tapscott, A. Caston: “Paradigm Shift: The new promise of information technology“ McGraw-Hill Companies New York (1993)
- [Wasserman, 97] S. Wasserman, K. Faust: “Social network analysis: methods and applications“ Cambridge Univ. Press, Cambridge (1997)
- [Wenger, 98] E. Wenger: “Communities of Practice: Learning, Meaning, and Identity“ Cambridge University Press, New York (1998)
- [Zhu, 03] H. Zhu; X. Wang, J. Zhu: “Effect of aging on network structure“ Phys. Rev. E, American Physical Society, 68:056121 (2003)