Corporate Wikis
Comparative Analysis of Structures and Dynamics

Klaus Stein
Laboratory for Semantic Information Technology, University of Bamberg
klaus.stein@uni-bamberg.de

Steffen Blaschke
Research Center for New Communication Media, University of Bamberg
steffen.blaschke@uni-bamberg.de

1 Introduction

Despite the popularity of public (e. g., Wikipedia) and corporate wikis (e. g., Bluepedia; cf. Due08), there is very little empirical research on the latter (for notable exceptions, see AB08). On the one hand, this is due to the fact that organizations are generally reluctant to foreclose unstructured data which may contain any information, from cafeteria opening hours to next-generation products. On the other hand, there are no established measures and methods for wiki evaluation which allow for a cross comparision of wikis data.

We have already introduced several quantitative and qualitative measures and methods for the corporate wiki analysis (BS08). In the following, we briefly discuss wikis as self-organizing collaboration networks and their application to organizational matters. We illustrate our contemplations with a comparative analysis of structures and dynamics of four wikis, and highlight the possibilities and limitations of visualization and data mining.[1]

2 Wikis—Self-Organizing Collaboration Networks

The first public wiki[2] already appears in 1995. Then and now, it is a site for the collaborative development of new ideas. Wikis are gaining more and more popularity ever since, just consider the unprecedented success of the free online encyclopedia Wikipedia. In contrast to the development of new ideas, however, Wikipedia is dedicated to the collaborative description of existing ideas.

More rigorously defined, 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 within a forms-capable Web browser client" (LC01,

[1] This research is supported by a grant from the Volkswagen Stiftung.
[2] Located at http://c2.com/cgi/wiki/

p. 14). Note that this definition refers to a wiki as both an application and a technology. For example, Wikipedia is an encyclopedia (application) running on the open-source software MediaWiki (technology). A wiki then offers some distinct features which sets it apart from other collaborative technologies such as document management systems:

Open Content. Any user can read all pages of a wiki. Moreover, any user can create pages, add information to, and correct errors on them. Although open content is the basis for efficient and effective collaboration, a wiki usually offers a dedicated user rights management to specifically control who can do what on which page. A wiki comprises a wide spectrum of bundled information, ranging from to-do lists and how-to articles on single pages to multi-page documents such as product manuals and project reports.

Revision History. A wiki tracks each and every change made to the content of a page. All of the changes are easily accessible through a page's revision history. Besides the actual changes, the revision history shows the times of changes and the users who made them. Thus, information may ascribed to particular users.

Dynamic Structure. A wiki reflects both a network of (hyper)linked documents and a network of social relations (i. e., coauthorship or, in simple words, who works with whom). The structure of each of these networks develops over time, dynamically adapting to the respective objectives of a wiki. Hyperlinks among pages and coauthorship relations among users appear and disappear, all in collaboration. In like manner, the pages on a wiki present their users with more of a work in progress than finished information. Both the (link) structure and the content of a wiki change dynamically.

Open content, revision history, and dynamic structure are the main features of a wiki. Together, they present a self-organizing collaboration network with a variety of applications, some of which we address next in the context of organizations.

3 Corporate Wikis

Wikis come at little initial cost, they are unstructured to begin with yet self-organizing in the course of action, and they require constant effort to maintain content. First and foremost, however, it is the success of public wikis that inspires more and more organizations to install wikis of their own. Some of these corporate wikis flourish at the grass roots, while others debut as projects spanning entire organizations. Either way, we find that wikis mostly spawn from functional departments, work groups, or project teams close to information technology (IT). Members of these departments, groups, or teams often enjoy a certain organizational slack when it comes to trying out new technological gadgets such as wikis, not the least because they may easily hide them from others. Notwithstanding, the success of corporate wikis at large hinges on all organizational members, not just IT personnel.

Wikipedia is the prototype of a wiki for many organizations. Yet there are large differences between public and corporate wikis, particularly in the ways the free online encyclopedia is used. The main purpose of Wikipedia is to create a collection of encyclopedic articles. In addition, it provides pages for discussion, category overviews, and personal use.

Wikipedia relies on a broad community of volunteers to write new and copy-edit old articles. Most, if not all, articles contain (hyper)links to others, in many ways similar to references within a classic encyclopedia. In order to look up information, people usually search Wikipedia by keywords, courtesy of the build-in search or any other generic search engine.

In contrast, the main purpose of a corporate wiki is to facilitate collaboration on all organizational matters. Thus, we find a wide range of different genres of organizational communication (YO92; OY94), anything from to-do lists, how-to articles, product manuals, project reports, meeting protocols, to encyclopedic articles (BS08). These pages are used for technical documentation, issues tracking, and reference information in software development as well as for status reports or meeting agendas in project management (for other findings from a survey on corporate wikis, cf. MWY06). Instead of using a keyword search to find information, users frequently follow links from one page to another (e. g., Start Page → Projects List → Project A → Meeting Protocol). Since the link structure of a corporate wiki is crucial to information retrieval, some organizations assign dedicated users (often interns) with the task of "gardening" this structure, though any user may always change the respective links.

A wiki is not an *un*structured information heap. The wiki technology (e. g., MediaWiki) supports structuring efforts within pages (automatic table of contents, templates, etc.) as well as across them (e. g., category systems). Despite the obvious advantages of a structure that dynamically adapts to the needs of its users, organizations—and particularly executive managers— often worry to provide their employees with such a flexible technology. Some organizations enforce measures and methods which are supposed to guide collaboration among employees. Unfortunately, quantitative goals such as "*x* number of pages per year" are detrimental, for the most part.

4 Analyzing Participation and Collaboration in Corporate Wikis

Organizations face the apparent need and, often, the plain desire to evaluate the success of their corporate wikis, not the least because they must legitimize their decisions. Rather than an in-depth case study of any one wiki, we put forth a more general view on collaboration in computer-mediated communication, thus providing organizations with a way to answer the question whether or not their corporate wikis are successful innovations. We illustrate our methodology on four corporate wikis which, in fact, are subject to our ongoing research. While we are not at liberty to disclose the names of the respective organizations, we briefly outline their business for contextual purpose and later discussion.

4.1 Case Studies of Corporate Wikis

The first organization is a start-up which provides one-stop solutions and services in mobile or proximity marketing (e. g., wireless hotspots). Its *startup wiki* supports all organi-

zational functions, ranging from research and development to marketing and sales.

The second organization is a public center of excellence which sponsors the innovation of information and communication technologies in small and medium enterprises. Its *facilitation wiki* features mostly research articles, project reports, and later publications thereof.

The third organization is a public news agency with a focus on online broadcasts. For the most part, its *news wiki* comprises production manuals (HOWTOs), frequently asked questions (FAQs), and official meeting protocols.

The fourth and last organization is a loosely-coupled community of students within a faculty of media science. Its *students wiki* centers around the self-assigned project to build a collection of encyclopedic articles on the topic of e-learning, knowledge management, and the like.

All four organizations are based in Germany, and all of them employ a standard installation of MediaWiki. Therefore, the technological environments are similar, whereas the organizational environments differ in management attitude towards the respective wiki. The *startup wiki* was installed because the incoming information needed to run the business had to be saved somewhere, and a wiki was a fast and simple ad-hoc solution. In effect, the wiki was the first and only content management system (CMS) available to all employees.

The second organization started its *facilitation wiki* without such urgent needs. Workflows were established and there was a CMS for quality assurance (QA) already up and running. A main reason for the installation was to learn about wikis themselves and thus support customers in their use of this technology. The wiki was launched and later maintained as a dedicated project, and employees were enforced to generate certain amounts of pages.

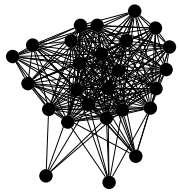
The third organization meant to complement its editorial CMS with an online help system, mainly for manuals. Due to high personnel turnover (many interns, mainly students) users started to add all kinds of HOWTOs and workflow descriptions to the *news wiki*. While every employee is invited to contribute, none is enforced to do so. Contrary to all others, this wiki allowed anonymous edits, so here we were not able to distinguish all single users but had to pool IP-ranges to prototypic user-groups.

The fourth and last *students wiki* is very different to the other wikis. Here, students use the wiki as collaboration platform for dedicated courses. At this, the body of encyclopedic articles grows with each semester of new authors arriving and old ones leaving.

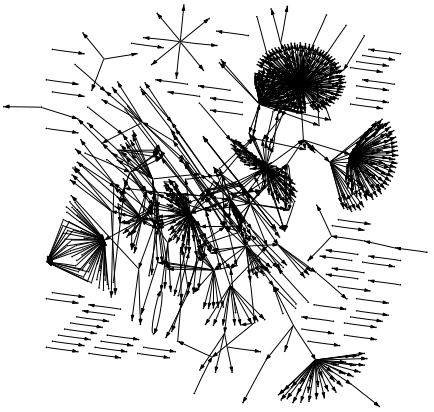
4.2 Analysis of Empirical Data

Based on our unrestricted access to these four wikis, we analyzed participation of and collaboration among users. Figures 1, 2, 3 and 4 give an overview of each one of the four wikis.[3] The network graphs at the top show (a) the coauthor relationship and (b) the

[3]The data is restricted to pages on namespace 0 (the MediaWiki main namespace) and excludes the system user, the administrator, and test users.

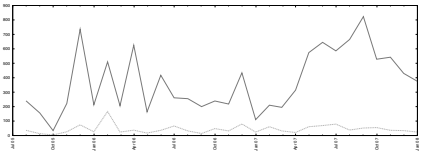


(a) coauthorship graph (26 users, 245 links, ⌀2)

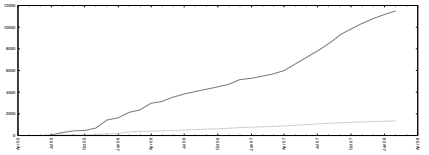


(b) pagegraph (1350 pages, 958 links, ⌀12)

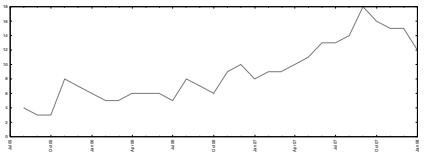
(c) histogram of coauthorship degrees



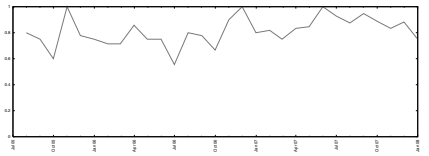
(d) revisions/pages per month



(e) cumulative revisions/pages per month



(f) author participation



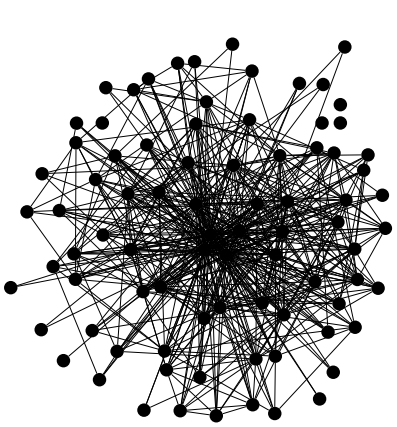
(g) relative author participation

Figure 1: Startup Wiki

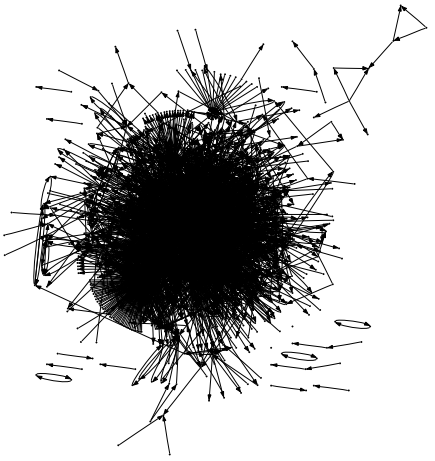
wiki pages link structure.[4] For each graph, the number of nodes, the number of edges, and the diameter of the largest component of the graph are given. The histogram (c) gives the degree distribution of the nodes of the coauthor graph (a). These three present a static snapshot of the wiki at the time of measurement. The following four graphs show the development of the wiki over time. (d) gives the number of revisions and the number of newly created pages for each month. (e) gives the same information in cumulative numbers (i. e., the number of revisions/pages up to this time[5]). (f) gives the author participation, i. e., the number of authors contributing to the wiki within one month and (g) the relative author

[4]The graph layout is based on a classic spring-embedder algorithm (KK89) as provided by `neato`.

[5]Pages imported by the system/admin user count to the number of pages, the corresponding revisions are suppressed as explained before.

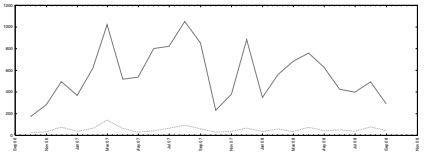


(a) coauthorship graph (83 users, 551 links, ⌀3)

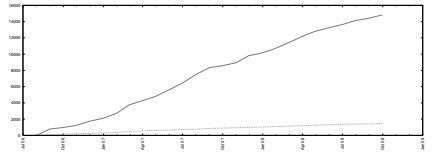


(b) pagegraph (1464 pages, 3819 links, ⌀10)

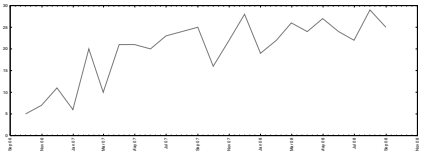
(c) histogram of coauthorship degrees



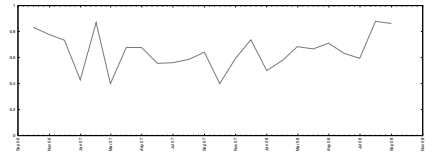
(d) revisions/pages per month



(e) cumulative revisions/pages per month



(f) author participation



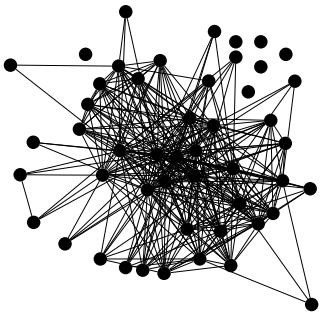
(g) relative author participation

Figure 2: Faciliation Wiki

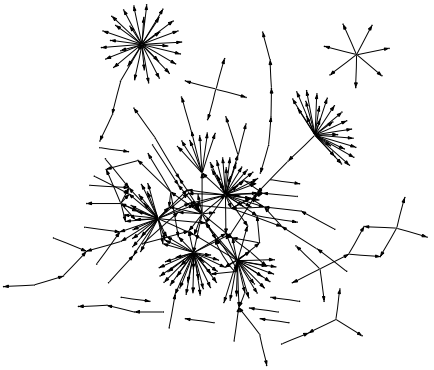
contribution, i. e., the fraction of (active[6]) authors contributing in this month.

In general, we are looking at wikis at a size of 26–140 users, approximately 400–1,500 pages, and 3,000–10,000 revisions. All of them display a steady growth in terms of cumulative users, pages, and revisions, although relative user participation as well as absolute numbers of pages and revisions fluctuate. Since our concern is not so much with a comprehensive analysis of a single wiki, we merely point out some of the most intriguing facts. Let us start with a general comparison of the graphs which reflects the different usages of wikis. Considering the coauthorship graphs, we observe that collaboration in the

[6] An author is counted as active in the time between the first and last edit

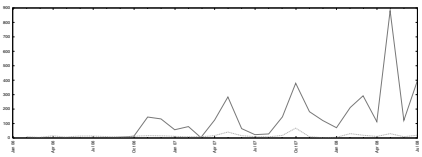


(a) coauthorship graph (49 users, 347 links, ⌀3)

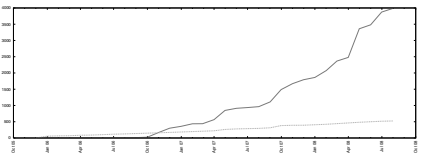


(b) pagegraph (524 pages, 307 links, ⌀11)

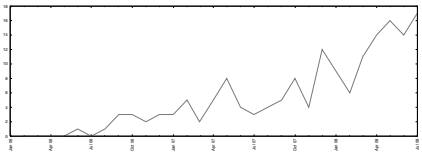
(c) histogram of coauthorship degrees



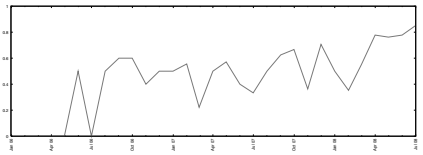
(d) revisions/pages per month



(e) cumulative revisions/pages per month



(f) author participation

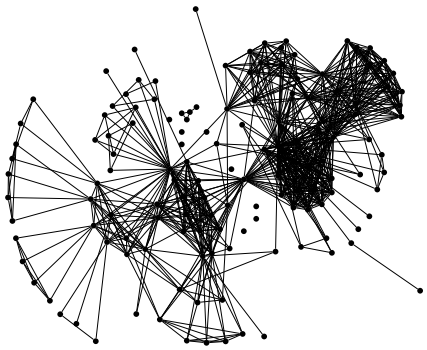


(g) relative author participation

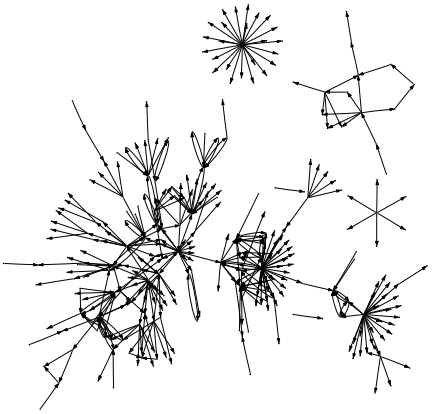
Figure 3: News Wiki

startup wiki is much more dense than in the *facilitation wiki* and *news wiki*. The respective histograms further highlight these findings, just compare the left-skewed 1c versus right-skewed 2c histograms. On the one hand we find a number of authors in close collaboration with others, while on the other hand most authors collaborate with just a few others, yet one author takes a central spot in collaboration with nearly everyone else (which causes the ⌀3). The *students wiki* has a a coauthorship distribution (4c) similar to the *facilitation wiki*, but most of the authors change with each semester (the avarage length of authorship of one user is 66 days). Unintuitively, this is *not* necessarily reflected in the structure of graph 4a, because an edge between two authors indicates an edit on the same page but not at the same time. Without closer investigation (supported by our knowledge that this *is* a students wiki) we would simply miss this fact.[7] In other words, the facts are present in the

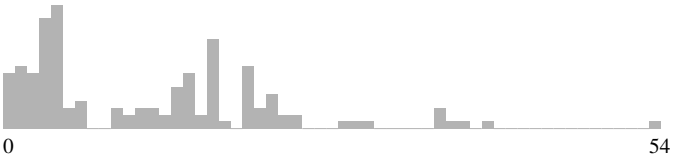
[7] Entry and exit of users are best seen in a dynamic presentation (see our flash movies at http://www.kinf.wiai.uni-bamberg.de/mwstat/Inhalt.html)



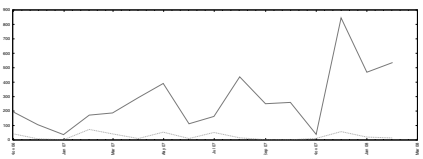
(a) coauthorship graph (140 users, 806 links, ⌀5)



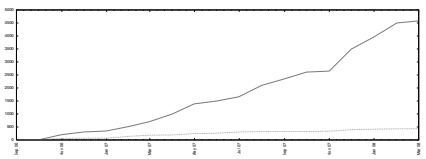
(b) pagegraph (432 pages, 382 links, ⌀13)



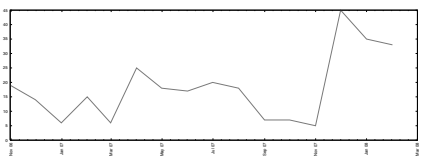
(c) histogram of coauthorship degrees



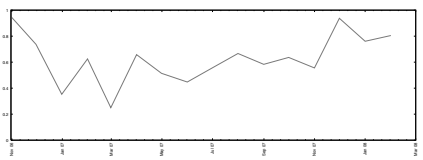
(d) revisions/pages per month



(e) cumulative revisions/pages per month



(f) author participation



(g) relative author participation

Figure 4: Students Wiki

quantitative data, but we need qualitative information to know where to look.

The next peculiarity can be found in page graph 2b. Obviously, it is much more dense than any other one, though it is smallest in diameter. The wiki has a an average of 2.6 links per page, and each page additionally belongs to at least one category, which effectively adds at least another link per page. The other wikis all have less than one link per page, making some of the pages only accessible by search engine or pages outside the main namespace (category pages, user pages, etc.). Why are page graphs so different in structure? Looking at the page graph visualization does not help much as all we see is an incomprehensible, tangled mass (truth be told, most pages are *not* directly linked). Again, we need additional

information: the *facilitation wiki* has a dedicated "gardener" whose job is to structure the wiki by adding links. Its structure is not a result of collaboration but the work of a single person, not *grown* but *installed*. Judging by density of the graphs alone, portal pages (i. e., collections of links to subpages), which are clearly visible in the other three graphs in star or hedgehog-like shape, are entirely hidden.

What do time series (d) to (g) tell us? The wikis are alive and growing, new pages are added, existing ones edited, and for the *news wiki* the growth even increases. For the *startup wiki* the relative user participation is close to 1, so any active author contributes continuously (remember, we defined an author as active in the time between her first and last edit). The other wikis are much harder to read: the *students wiki* has high fluctuation with short, more active sequences which should consequently lead to higher relative participation—but it does not. The participation rate of the *news wiki* seems to increase—but here we have anonymous edits (which we had to map manually to groups of users), so this is even harder to read.

If there is one thing to learn, then the presented data[8] (numbers, network graphs, and time series) helps us to explore the wikis in order to get ideas for the right questions to ask, especially if supported by additional information about the corresponding organization and the usage of its wiki. They are a starting point for exploration, for visual data mining: "these pages seem to be portals, take a look at them," "this user is central, what is her job in the wiki," and so on.

But neither visualization nor statistical values are enough on their own to proof complex statements (besides simple cases like: no one has been editing in over a year). We always need background information to know which data is worth looking at, what measures make sense computing, and finally draw the right conclusions (e. g., we would not know about the author turnover in the *students wiki* using only the data presented here). Most of the facts are present in the data (besides anonymization due to privacy purposes). The problem is to ask the right questions, as every statistical value, every graph is a low-dimensional cut through high-dimensional data. We have to find the right cut-in and backup our findings by additional information. Then, and only then, we may use visualization to *explain* our findings to others.

Wiki Explorator, the tool we used and developed for our research, offers numerous ways to analyze a wiki, including its evolution over time. We provide a web-based service for generating standard reports of arbitrary wikis[9]. Even further in-depth analysis of a wiki is possible with our Wiki Explorator library which will be released as open source soon.

5 Conclusion

Wiki analysis is not a one-pass task. Every wiki is different—in size, in usage, and in the function it supplies to the organization. These differences show up in the wiki structure. In this paper we did not focus on an in-depth analysis of one dedicated wiki including detailed

[8] Along with a myriad of other evaluations we did while preparing this paper

[9] See https://www.kinf.wiai.uni-bamberg.de/mwstat/ for the German version.

node-based measures (centrality, prestige etc.) as well as more sophisticated measures such as interlocking.[10] We focused on comparable static graph visualizations and time-series plots of values representing the wiki development.

We pointed out that differences do show up in the data and that visualizations do tell us where to look, but also that they are not sufficient to draw detailed, valid conclusions. Graphs and figures complement qualitative analysis to describe the circumstantial organizational surroundings of a wiki, but they can never replace them. In other words, qualitative information helps to discover structural idiosyncrasies in the quantitative data at hand, and the quantitative findings, in turn, require qualitative reasoning to yield the big picture.

A comparative analysis of the structures and dynamics of corporate wikis is hardly feasible for any single organization, mainly due to a lack of second-hand data. However, organizations may employ measures and methods similar to our approach to gain a first impression of their own empirical data. And if they are truely interested how their corporate wiki compares to others, they may simply use our Wiki Explorator for a report of their own.

References

[AB08] Paul Alpar and Steffen Blaschke, editors. *Web 2.0: Eine Empirische Bestandsaufnahme*. Vieweg+Teubner, Wiesbaden, 2008.

[BS08] Steffen Blaschke and Klaus Stein. Methods and Measures for the Analysis of Corporate Wikis: A Case Study. In *Proceedings of the 58th Annual Conference of the International Communication Association (ICA)*, Montreal, May 22–26 2008.

[Due08] Gunter Dueck. Bluepedia. *Informatik-Spektrum*, 31(3):262–269, 2008.

[KK89] Tomio Kamada and Saturo Kawai. An Algorithm for Drawing General Undirected Graphs. *Information Processing Letters*, 31(1):7–15, 1989.

[LC01] Bo Leuf and Ward Cunningham. *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley, Boston, MA, 2001.

[MWY06] Ann Majchrzak, Christian Wagner, and Dave Yates. Corporate Wiki Users: Results of a Survey. In *Proceedings of WikiSym '06*, pages 99–104, Odense, Denmark, 2006.

[OY94] Wanda J. Orlikowski and JoAnne Yates. Genre Repertoire: The Structuring of Communicative Practices in Organizations. *Administrative Science Quarterly*, 39(4):541–574, 1994.

[YO92] JoAnne Yates and Wanda J. Orlikowski. Genres of Organizational Communication: A Structural Approach to Studying Communication and Media. *Academy of Management Review*, 17(2):299–326, 1992.

[10] For a detailed study of one of the wikis presented here, see (BS08)