Latent Barriers in Wiki-based Collaborative Writing (Extended Abstract)

Alexander Mehler, Christian Stegbauer and Rüdiger Gleim

1   Research Questions

Following the rise of Wikipedia as an online encyclopedia, wikis became famous as a medium of collaborative writing in many different areas of communication. This includes special wikis for consumer communication (e.g., `wiki.siemens-enterprise.com`), leisure communication (e.g., `en.memory-alpha.org/wiki`), news communication (e.g., `de.wikinews.org`) or regional communication (e.g., `www.arborwiki.org`) – to name just a few examples that go beyond knowledge communication, which is Wikipedia's primary domain. In all these areas, wikis promise to enable web communities to collaboratively write on a certain topic – be it a knowledge domain (e.g., `www.glottopedia.de`), a community (e.g., `semanticweb.org`), a collaborative project (e.g., `wiki.openstreetmap.org`), a city (e.g., `dresden.stadtwiki.de`) or a programming language (e.g., `wiki.python.org`). Apparently, wikis are an outstanding example of the web as a kind of laboratory in which the emergence and disappearance of novel webgenres can be observed in any communication area (Santini et al., 2009). A prominent example of this process is the *GuttenPlag Wiki* (`de.guttenplag.wikia.com`), which in 2011 was the first wiki used for documenting plagiarism in scientific communication. In any of these cases, wikis are the primary means of collaborative writing as they promise to overcome barriers of participation that may affect agents of divergent educational or social provenance. Today, we know that this has to be treated with caution (Stegbauer and Rausch, 1999; Stegbauer, 2001, 2009). This is due to the unequal distribution of participation in media like Wikipedia, which has been said to manifest the so called *90/9/1 pattern* (Fuster Morell, 2009). In terms of this model, 90% of the participants are so called lurkers or audience[1], 9% are given by so called weak contributors while 1% includes committed participants (Fuster Morell, 2009). For related findings see Ortega et al. (2008), Schroer (2008), Mehler and Sutter (2008), Stegbauer (2009), and Reagle (2010).

   In this paper, we present a study of collaborative writing in special wikis of four different communication areas (see Table 1) in comparison to Wikipedia. More specifically, we analyze 25 wikis to get insights into the specifics of collaborative writing in Wikipedia in comparison to its relatives. We ask for the degree of collaboration in such wikis in terms of the distributions of collaboration of wiki authors as well as in terms of the transitivity of their social relations in the underlying collaboration network. In models of these networks, vertices denote authors, while edges denote social relations among them. In the case of collaboration networks, a social relation between two agents is established whenever they work together in writing at least one text (e.g. wiki article) (Newman, 2004a,b). As we are interested in cooperation of a higher quality, that is, in collaborative writing, we look for temporal constraints of this sort of cooperation (cf. Stegbauer and Rausch, 2006). This is done by raising the following research questions:

   Q1 *Is participation in collaborative writing scale-free such that a couple of highly active authors (center) is complemented by an overall majority of more or less inactive followers (periphery)?*

---

[1] In mailing lists, we observed lurkers in the range of 56-81% – (Stegbauer and Rausch, 2002)

Q2 *Is this scale-freeness a characteristic of the wiki medium or of the communication area in which the wiki is used?*

Q3 *Does the way collaborative writing is performed by wiki communities hint at a barrier for the participation of large groups of agents?*

By means of our comparative analysis, we shed light on the degree to which different areas of communication are characterized by different patterns of collaborative writing. We do that by putting emphasis on the question to which degree wikis of different communication areas are distinguishable in terms of the collaboration patterns they manifest. This is done with respect to degree distributions and transitivity analyses. Starting from the findings of Stegbauer (2009), who pointed at latent barriers of participation in collaborative writing, we hypothesize that a similar process also affects participation in special purpose wikis.

2   Approach

As a resource of our analysis we use a corpus of 25 special purpose wikis together with the German Wikipedia. This is done to provide a comparative study of collaborative writing across different wiki-based webgenres. We apply complex network analysis to perform our analysis (Newman, 2010). We show that a remarkably simple apparatus of network analysis suffices to detect a pattern of collaborative writing that contradicts the notion of wikis as high-degree participatory media. In order to do that, we apply a time-window-based approach to detecting collaboration relations that varies window size as a time-related parameter of collaboration measurement (Mehler et al., 2012). This approach is in line with recent advancements of social network analysis based on Wikipedia (Tang et al., 2008; Brandes et al., 2009; Liu and Ram, 2009; Fong and Biuk-Aghai, 2010; Laniado and Tasso, 2011). We complement these approaches in that we comparatively study wikis of several communication areas to get a *tertium comparationis* for assessing the peculiarity of Wikipedia as the prime example of collaborative writing.

Table 1 shows the corpus of special purpose wikis analyzed in this study. In parentheses we show the order (number of vertices) and size (number of edges) of the collaboration networks that result from aggregating the underlying time-window-specific collaboration networks. In these cases, we induced collaboration networks for time windows of a length of 30 days that were shifted day by day (see Mehler et al. 2012 for the details of the underlying method of network induction). Numbers to the left to the parentheses show the order and size of the corresponding collaboration networks that were induced irrespective of time constraint. The full paper of this extended abstract will provide all numerical results of our analysis for the different wikis under consideration.

3   What the Final Paper Will Provide

Regarding the three research questions raised in Section 1, we provide evidence in support of the following interpretations:

1. We show that collaborative writing is scale-free in terms of a center of highly active authors that is surrounded by a large periphery of more or less inactive authors. However, other than classical approaches to scale-free networks we show that this pattern is affected by the size of the time windows (in which collaborations are observed) in a non-trivial manner. As a consequence of qualifying short- and long-term collaborations differently, we get access to hidden patterns of collaboration that tend to be overlooked from the point of view of time-indifferent analyzes. This

Special Purpose Wiki	Order	Size	Communication Area
berlin.wikia.com	840(614)	8,869(1,356)	regional communication
de.anime.wikia.com	510(305)	3,743(613)	leisure communication
de.avatar.wikia.com	983(835)	23,394(3,109)	leisure communication
de.beyblade.wikia.com	621(591)	10,001(3,673)	leisure communication
de.digimon.wikia.com	823(742)	21,425(4,036)	leisure communication
de.dofuswiki.wikia.com	1,960(1,363)	25,292(2,843)	leisure communication
de.dragonica.wikia.com	537(395)	3,837(887)	leisure communication
de.fallout.wikia.com	1,207(984)	12,158(2,406)	leisure communication
de.flyff.wikia.com	2,240(1,736)	48,287(5,356)	leisure communication
de.guttenplag.wikia.com	16,675(11,809)	157,308(137,740)	scientific communication
de.harry-potter.wikia.com	8,667(7,747)	246,494(38,588)	leisure communication
de.kingdomhearts.wikia.com	969(884)	22,797(3,438)	leisure communication
de.lotr.wikia.com	558(435)	6,432(977)	leisure communication
de.mario.wikia.com	1,005(947)	29,603(5,203)	leisure communication
de.rom.wikia.com	818(643)	10,110(1,447)	leisure communication
de.scrubs.wikia.com	3568(3,279)	311,377(27,255)	leisure communication
de.software.wikia.com	9,192(8,514)	683,891(45,214)	technical communication
de.vroniplag.wikia.com	1,827(1,791)	49,916(33,845)	scientific communication
diealdor.wikia.com	5,667(5,084)	136,734(26,813)	leisure communication
forscherliga.wikia.com	1,461(1,281)	14,169(4,191)	leisure communication
supertux.lethargik.org	1,506(1,379)	67,124(9,412)	leisure communication
wiki.muenster.org	1,395(1,123)	36,371(30,703)	regional communication
www.c64-wiki.de	2124(1,725)	36,703(6,763)	technical communication
www.hamburgwiki.de	752(678)	9,015(2,926)	regional communication
www.pfenz.de	3,720(3,105)	49,643(10,698)	regional communication

Table 1: The corpus of 25 special purpose wikis.

is exemplified by the GuttenPlag Wiki that exhibits a stable transitivity pattern among authors only for short time windows. Thus, we complement a recent study on Wikipedia as a scale-free collaboration network (Stegbauer and Mehler, 2011) from the point of view of time series analyses.

2. As we ask whether the variation of patterns of collaboration as observed in our corpus of wikis is due to their membership in different communication areas (leisure communication, scientific communication etc.), we additionally provide an experiment in automatic classification that learns differences of collaboration patterns as a function of the topological structures they induce. This is done by means of *Quantitative Network Analysis* (QNA) (Mehler, 2008), which implements semi-supervised learning on vectors of graph invariants in order to solve this classification task. In this context, we observe a strong divide between extracts of Wikipedia (even if being reduced to single articles and their one-link neighborhoods) on the one hand and special purpose wikis on the other. As a result, we see the need for a typological study of wiki-based patterns of collaboration that starts from a large corpus of special wikis.

3. Regarding Question Q3, we do not yet provide enough evidence to decide it appropriately. However, we can point at recent findings about positions of Wikipedians. In Stegbauer and Mehler (2011), we show that deviations from the ideal of a power-law-like distribution of node connec-

tivity in collaboration networks can be explained by decomposing this distribution according to the positions of the corresponding authors. The position "vandal hunter", for example, shows a specific collaboration pattern in that hunters work on many, thematically diversified articles. Hunters do not need to have detailed knowledge about the topics of the articles they work on. As a result, hunters are positioned in a way that manifests many collaborations with many different authors. In contrast to this, article writers know the respective field of knowledge of the articles they work on. Accordingly, they tend to collaborate with a few other, likewise specialized authors. Following this line of reasoning, one can identify several positions (i.e., types of authors), that allow for decomposing degree distributions of collaboration networks such that the resulting constituent distributions manifest separate power laws. Stegbauer (2009) shows that membership to positions is restricted in Wikipedia in terms of quantity. While there is, for example, a stable number of administrators and only a couple of hunters, there are many specialized authors. Restrictions of this sort result, for example, from expertise, trust or competition.

From the point of view of our findings, we conclude that, in Wikipedia, collaboration structure reflects barriers to participation on an equal footing. Since we find similar patterns of collaboration in special purpose wikis, we hypothesize that barriers of the same sort also affect the participation in these wikis. Of course, we do not expect that positions are the same for every kind of wiki. Irrespective of such variations, we assume that position-dependent collaboration structures tend to impose latent barriers of participation. In a future study we plan to investigate this hypothesis in more detail. As mentioned above, this will be done by means of a large-scale corpus of special purpose wikis.

References

Ulrik Brandes, Patrick Kenis, Jürgen Lerner, and Denise van Raaij. Network analysis of collaboration structure in wikipedia. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 731–740, New York, NY, USA, 2009. ACM.

Peter Kin-Fong Fong and Robert P. Biuk-Aghai. What did they do? deriving high-level edit histories in wikis. In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*, WikiSym '10, pages 2:1–2:10, New York, NY, USA, 2010. ACM.

Mayo Fuster Morell. Online creation communities for the building of digital commons: Paricipation as an eco-sytem? In *Contribution to the panel on Organizational principles and political implications of the International forum on free culture, Barcelona, October 30*, 2009. URL \url{http://openfsm.net/projects/freecultureforum/preparation-discussion-and-working-materials/MayoFusterMorell_PasticipationAsAnEco_Barcelona.pdf}.

David Laniado and Riccardo Tasso. Co-authorship 2.0: patterns of collaboration in Wikipedia. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*, HT '11, pages 201–210, New York, NY, USA, 2011. ACM.

Jun Liu and Sudha Ram. Who does what: Collaboration patterns in the Wikipedia and their impact on data quality. In *Proceedings of 19th Annual Workshop on Information Technologies and Systems (WITS 2009)*, Phoenix, Arizona, USA, 2009.

Alexander Mehler. Structural similarities of complex networks: A computational model by example of wiki graphs. *Applied Artificial Intelligence*, 22(7&8):619–683, 2008.

Alexander Mehler and Tilmann Sutter. Interaktive Textproduktion in Wiki-basierten Kommunikationssystemen. In Ansgar Zerfaß, Martin Welker, and Jan Schmidt, editors, *Kommunikation, Partizipation und Wirkungen im Social Web – Weblogs, Wikis, Podcasts und Communities aus interdisziplinärer Sicht*, pages 267–300. Herbert von Halem, Köln, 2008.

Alexander Mehler, Christian Stegbauer, and Rüdiger Gleim. Zur Struktur und Dynamik der kollaborativen Plagiatsdokumentation am Beispiel des GuttenPlag Wiki: eine Vorstudie. In Barbara Frank-Job, Alexander Mehler, and Tilman Sutter, editors, *Die Dynamik sozialer und sprachlicher Netzwerke. Konzepte, Methoden und empirische Untersuchungen am Beispiel des WWW*. VS Verlag, Wiesbaden, 2012.

Mark E. J. Newman. Who is the best connected scientist? A study of scientific coauthorship networks. In Eli Ben-Naim, Hans Frauenfelder, and Zoltan Toroczkai, editors, *Complex Networks*, pages 337–370. Springer, Berlin, 2004a.

Mark E. J. Newman. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101:5200–5205, 2004b.

Mark E. J. Newman. *Networks: An Introduction*. Oxford University Press, Oxford, 2010.

Felipe Ortega, Jesus M. Gonzalez-Barahona, and Gregorio Robles. Passive participants in mailing list. In *Proceedings of the Hawaii International Conference on System Sciences*, 2008. `http://doi.ieeecomputersociety.org/10.1109/HICSS.2008.333`.

J. M. Reagle. *Good faith collaboration: The culture of Wikipedia*. History and foundations of information science. MIT Press, Cambridge, MA, 2010.

Marina Santini, Alexander Mehler, and Serge Sharoff. Riding the rough waves of genre on the web: Concepts and research questions. In Alexander Mehler, Serge Sharoff, and Marina Santini, editors, *Genres on the Web: Computational Models and Empirical Studies*, pages 3–32. Springer, Dordrecht, 2009.

Joachim Schroer. *Wikipedia: auslösende und aufrechterhaltende Faktoren der freiwilligen Mitarbeit an einem Web-2.0-Projekt*. Logos, Berlin, 2008.

Christian Stegbauer. *Grenzen virtueller Gemeinschaft. Strukturen internetbasierter Kommunikationsforen*. Westdeutscher Verlag, Wiesbaden, 2001.

Christian Stegbauer. *Wikipedia: Das Rätsel der Kooperation*. VS, Wiesbaden, 2009.

Christian Stegbauer and Alexander Mehler. Positionssensitive Dekomposition von Potenzgesetzen am Beispiel von Wikipedia-basierten Kollaborationsnetzwerken. In *Proceedings of the 4th Workshop Digital Social Networks at INFORMATIK 2011: Informatik schafft Communities, Oct 4-7, 2011, Berlin*, 2011.

Christian Stegbauer and Alexander Rausch. Ungleichheit in virtuellen gemeinschaften. *Soziale Welt*, 50(1):93–110, 1999.

Christian Stegbauer and Alexander Rausch. Passive participants in mailing list. In B. Batinic, U-D. Reips, M. Bosnjak, and A. Werner, editors, *Online Social Sciences*. Hogrefe & Huber Publishers, New York, 2002.

Christian Stegbauer and Alexander Rausch. "Moving Structure" als Analyseverfahren für Verlaufsdaten am Beispiel von Mailinglisten. *Sozialwissenschaftlicher Fachinformationsdienst soFid (Methoden und Instrumente der Sozialwissenschaften)*, 1:11–30, 2006.

Libby Veng-Sam Tang, Robert P. Biuk-Aghai, and Simon Fong. A method for measuring co-authorship relationships in mediawiki. In *Proceedings of the 4th International Symposium on Wikis*, WikiSym '08, pages 16:1–16:10, New York, NY, USA, 2008. ACM.